

# Great apes infer others' goals based on context

David Buttelmann · Sebastian Schütte ·  
Malinda Carpenter · Josep Call · Michael Tomasello

Received: 16 August 2011 / Revised: 31 May 2012 / Accepted: 18 June 2012 / Published online: 30 June 2012  
© Springer-Verlag 2012

**Abstract** In previous studies claiming to demonstrate that great apes understand the goals of others, the apes could potentially have been using subtle behavioral cues present during the test to succeed. In the current studies, we ruled out the use of such cues by making the behavior of the experimenter identical in the test phase of both the experimental and control conditions; the only difference was the preceding “context.” In the first study, apes interpreted a human’s ambiguous action as having the underlying goal of opening a box, or not, based on that human’s previous actions with similar boxes. In the second study, chimpanzees learned that when a human stood up she was going to go get food for them, but when a novel, unexpected event happened, they changed their expectation—presumably based on their understanding that this new event led the human to change her goal. These studies suggest that great apes do not need concurrent behavioral cues to infer others’ goals, but can do so from a variety of different types of cues—even cues displaced in time.

**Keywords** Intentional action · Goal understanding · Nonhuman primates · Chimpanzees

---

Declarations: The experiments comply with the current laws of Germany.

---

D. Buttelmann (✉)  
Research Group “Kleinkindforschung in Thueringen”,  
University of Erfurt, Nordhaeuser Strasse 63,  
99089 Erfurt, Germany  
e-mail: david.buttelmann@uni-erfurt.de

D. Buttelmann · S. Schütte · M. Carpenter · J. Call ·  
M. Tomasello  
Max Planck Institute for Evolutionary Anthropology,  
Leipzig, Germany

## Introduction

Organisms can interact with others in much more effective ways if they understand what those others are doing in terms of their underlying goals, that is, in terms of the changes of state in the environment they are trying to bring about. Understanding actions in terms of goals enables an observer to predict an actor’s actions not just in recurrent situations, when the same situational and behavioral cues are again present, but also more flexibly in many novel situations (Tomasello and Call 1997).

There are currently four sets of experiments claiming that chimpanzees understand the actions of others in terms of their underlying goals. All of these studies test great apes’ understanding of accidents and failed attempts—since these are the cases in which what the actor does overtly in behavior does not match his underlying goal (the question being whether subjects then respond to the behavior or to the goal). First, in an object choice paradigm, Call and Tomasello (1998) taught chimpanzees and orangutans that a marker on top of one of three containers indicated that hidden food was inside that container. Then, in the test, two containers were marked, one accidentally and the other on purpose. When apes were then allowed to choose between the containers, they more often chose the one marked intentionally than accidentally and they did this from the earliest trials (see also a study by Povinelli et al. 1998, in which chimpanzees chose who to receive juice from based on the actor’s past intentional or accidental actions).

Second, using a helping paradigm, Warneken and Tomasello (2006) had a human experimenter drop an object accidentally in the presence of each of three juvenile chimpanzees, and then strain and reach toward it (with several different objects in several different situations).

The chimpanzees fetched it for him. Importantly, they did not fetch it for him in various control conditions in which he threw the object away or otherwise indicated a lack of interest. The chimpanzees' different behavior in the experimental and control conditions could be interpreted as indicating an understanding of the experimenter's different goals in the two situations (see also Woodruff and Premack's 1979, study with a single chimpanzee). Warneken et al. (2007) set up another situation in which one chimpanzee might help another. In this study, one chimpanzee was attempting to get into an adjoining room, often shaking the door in his attempt. Other chimpanzees then quite often, from their advantageous location, pulled a chain that unlocked the door, so that the first chimpanzee could have access to the room he wanted to get into. They unlocked that door more than in a control condition in which the first chimpanzee was trying to get out of another door. These studies of instrumental helping suggest that chimpanzees can tell when someone needs help achieving his goal.

Third, Tomasello and Carpenter (2005) used two imitation tests (from research with human infants) with the same three juvenile chimpanzees tested in the Warneken and Tomasello (2006) study. In one test, based on a study by Meltzoff (1995), a human tried but failed to perform various actions on objects. The chimpanzees showed that they successfully discerned the actions the human was attempting to perform by performing those actions themselves (as often as when they had seen her perform them successfully and more often than when she had just manipulated the object). In the other test, based on a study by Carpenter et al. (1998), a human performed two actions on a series of apparatuses, one intentional and one accidental, before it was the chimpanzees' turn. Chimpanzees copied the intentional action more often than the accidental action. In addition, in a different imitation study based on that of Gergely et al. (2002), Buttelmann et al. (2007) demonstrated an unusual action for enculturated chimpanzees in two different situations: one in which some constraint forced the human to perform that action if he wanted to attain the goal and the other in which no such constraint was present. The chimpanzees differentiated these two situations, imitating more often when the human's action was freely chosen intentionally.

Finally, in a natural interaction paradigm, Call et al. (2004) had a human experimenter give chimpanzees food repeatedly through a glass panel. Then, on some trials, the experimenter did not give the food. The experimental manipulation was that sometimes he did not give it because he was *unwilling* to in various ways, whereas other times he did not give it because he was *unable* to in various ways. For example, in the unwilling condition, the experimenter teased the chimpanzee with the food (pulling it back when the chimpanzee reached for it) and in the

unable condition, he dropped the food clumsily while trying to give it to the chimpanzee. The basic finding was that chimpanzees reacted similarly to the different unwilling actions by expressing frustration and impatience and they reacted similarly to the different unable actions by being patient (because he was "trying"). This similarity of reaction across the different instantiations of the two experimental conditions suggests that the chimpanzees understood the different goals of the experimenter in the different conditions, regardless of how they were expressed behaviorally.

Together, these studies provide very suggestive evidence that chimpanzees (and perhaps other great apes) understand the goals of others. However, in all of the studies at test, the human actor behaved slightly differently in the experimental and control conditions; indeed, the slightly different behavior in the test phase was the cue to the different underlying goals involved. This means that these results could be explained in terms of behavioral differences of the experimenter at test. In our view, these explanations are unlikely because of the novelty and diversity of behaviors used as both cues and responses in the different studies. Nevertheless, there is still in all cases a behavioral difference between conditions in the test phase that could serve as a discriminative cue.

In the current experiments, we investigated whether great apes could pass a goal understanding test when they were not able to rely on concurrent behavioral cues. By doing this, we went one step further than all the studies cited above: we made the immediate use of such a behavioral cue impossible by making the experimenter's actions during the test phase of the study identical in the experimental and control conditions. What differed was only the context leading up to those actions which, for organisms capable of understanding goals, would lead to a different interpretation of what the actor was doing in the two cases (but not for individuals that can base their response only on concurrent behavioral cues alone).

### Study 1a

In this study, we assessed the reactions of all four species of great apes to an ambiguous action to see whether, based on the previous context alone, they perceived it as an attempt to open a box containing food. Note that this is the first direct test of goal understanding in two of these species, gorillas and bonobos.

During the test phase, the experimenter—identically in both the experimental and control conditions—turned a piece of metal on top of a box which the ape knew contained food, an action that in some contexts could be seen as trying to open the box. The difference between

conditions was that in the experimental condition apes had previously observed the experimenter manipulating different locks and latches on the top of other boxes and so opening them (and giving apes the food from inside). Thus, apes in this condition could assume that she would be trying to open the final test box as well (see Carpenter et al. 2002, for a similar procedure with 2-year-old children). In contrast, in the control condition, apes had previously observed the experimenter simply manipulating the locks and latches on the top of the other boxes (but not opening them), and then giving apes food from her pocket. Thus, apes in this condition should have no assumption that she would be trying to open the final test box.

If our subjects paid attention only to concurrent behavioral cues, they should behave identically in the two conditions since the experimenter's action on the latch of the target box was identical in both conditions. If, in contrast, great apes were able to include the context leading up to this action into their assessment of the experimenter's goal, this would lead to a different interpretation in the two conditions: in the experimental condition, they should wait more patiently because they perceive the experimenter to be attempting to open the box, whereas in the control condition, they should be more likely to leave because they perceive that the experimenter does not want to extract the grape from the box.

## Methods

### Participants

Thirty great apes participated in this study. There were 15 chimpanzees (*Pan troglodytes*), 5 bonobos (*Pan paniscus*), 5 gorillas (*Gorilla gorilla*), and 5 orangutans (*Pongo abelii*). There were 19 females and 11 males who ranged in age from 4 to 32 years (mean age = 14.7 years), see Table 1. The apes were housed socially in groups of at least five individuals (separated by species) at the Wolfgang Köhler Primate Research Center in the Leipzig Zoo, Germany. Each species had access to an indoor area (230–430 m<sup>2</sup>) and an outdoor area (1,680–4,000 m<sup>2</sup>) furnished with various climbing structures, shelter, and natural vegetation. At night, the apes sleep in several series of cages (40–50 m<sup>2</sup>). In addition to experiments, the animals are provided with an enrichment program, including different kinds of tools and foraging containers. Several times per day, the apes are fed a diet consisting primarily of vegetables, fruits, and cereals with regular additions of eggs and meat for the chimpanzees. Test sessions took place in a familiar enclosure (approximately 15 m<sup>2</sup>). The subjects were used to being separated in adjacent enclosures from their group members for testing. They were not food deprived for testing, and water was available

**Table 1** Overview of sex (*F* female and *M* male) and age (in years) of subjects that participated in each study

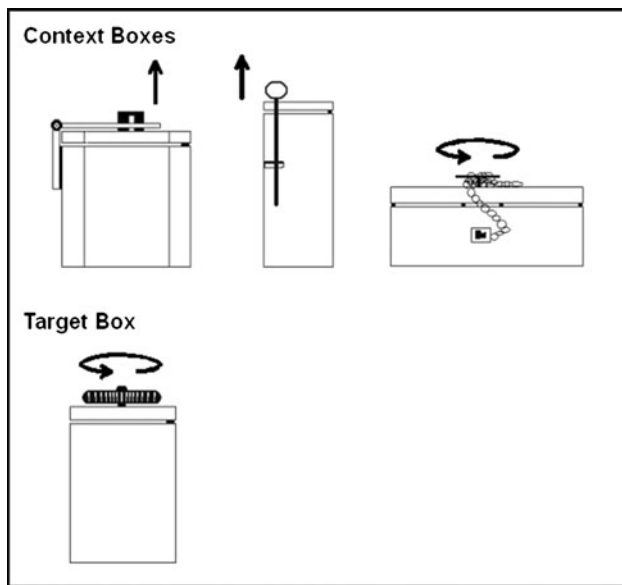
Species	Name	Sex	Age	Study 1a	Study 1b	Study 2
Chimpanzees	Pia	F	5	x	x	x
	Alexandra	F	6	x	x	
	Anett	F	6	x	x	
	Swela	F	10			(x) <sup>a</sup>
	Fifi	F	12	x	x	x
	Gertruida	F	12	x	x	x
	Jahaga	F	12	x	x	x
	Sandra	F	12	x	x	x
	Natascha	F	25		x	x
	Dorien	F	25		x	x
	Riet	F	27	x	x	x
	Ulla	F	28		x	
	Corrie	F	28		x	x
	Fraukje	F	29	x	x	x
	Alex	M	4	x	x	
	Lome	M	4	x	x	(x) <sup>a</sup>
	Patrick	M	8	x	x	x
Unyoro	M	8	x		x	
Frodo	M	11	x	x	x	
Robert	M	29	x	x	x	
Bonobos	Yasa	F	8	x	x	
	Ulindi	F	11	x	x	
	Kuno	M	8	x	x	
	Limbuko	M	9	x	x	
	Joey	M	22	x	x	
Gorillas	Viringika	F	10	x		
	Effi	F	13	x		
	Bebe	F	26	x		
	N'Diki	F	28	x		
	Gorgo	M	25	x		
Orangutans	Padana	F	7	x		
	Dokana	F	16	x		
	Pini	F	18	x		
	Dunja	F	32	x		
	Bimbo	M	24	x		

<sup>a</sup> These subjects had to be dropped because of inattentiveness

throughout all testing times. They were not distressed and were free to stop participating at any time.

### Materials and design

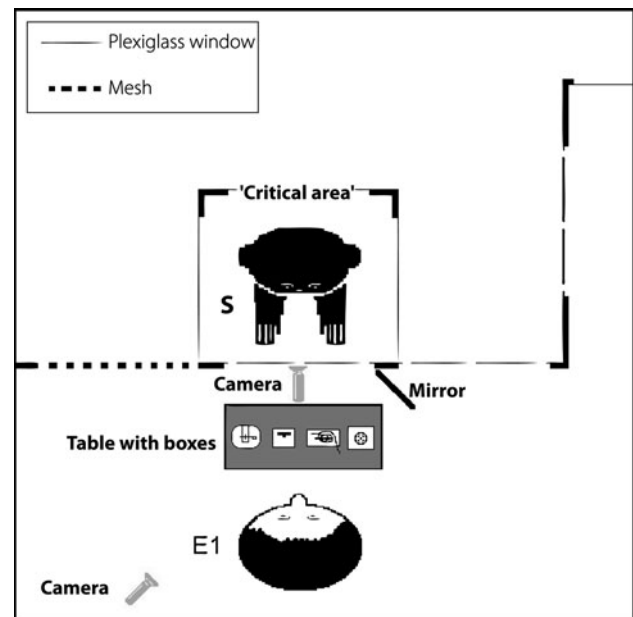
Testing materials consisted of a plastic table (80 × 40 cm), onto which three differently shaped gray plastic boxes were permanently fixed; these were the “context boxes.” A fourth box, the “test box,” could be attached to an empty space on the right side of the table from the experimenter's



**Fig. 1** Sketch of the closing devices of the four boxes used in Study 1

perspective. The four boxes were of a similar size (1–1.5 l volume). During sessions, the plastic table was attached to the outside of a transparent plexiglass window (75 × 50 cm) in apes' enclosure. At the bottom of this window were three holes (5 cm in diameter), the middle one of which served to provide the food reward to the apes. An opaque rectangular piece of plastic (the “occluder,” 50 × 100 cm) was used to conceal the boxes when necessary (see below). All boxes were closed by a lid on top, with each lid featuring a distinctive closing mechanism (see Fig. 1). On the lid of the test box, there was a metal wheel which could be turned, making an audible, creaking sound. Two camcorders recorded the sessions. See Fig. 2 for a sketch of the general setup.

Each ape participated in both the experimental and the control condition, within subjects, participating in each condition in a different session. The two sessions took place 4 to 6 weeks apart (with one exception where the delay was 8 weeks). A session consisted of three consecutive testing days: In the first 2 days and part of the third day (“context phase”), the relevant context (opening vs. simply manipulating the boxes) was established. During that phase, only the three context boxes were attached to the table. On each of the first 2 days, an experimenter (E1) acted on all three context boxes in three consecutive rounds. On the third day, after E1 had again acted on the context boxes for two rounds, a fourth box, the test box, which was unfamiliar to the apes, was added to the table. E1 then acted on the context boxes for one more round and then the test box (“testing phase”). All together, each ape thus witnessed nine actions on each of the context boxes (three times per day × 3 days) and one action on the test box in each condition. Each condition was administered by



**Fig. 2** Sketch of the experimental setup used in Study 1

a different experimenter to reduce carryover effects. On each day, testing lasted about 15 min per subject. Order of conditions was counterbalanced, with gender and age matched as closely as possible.

### Procedure

During the context phase of the experiment, apes were provided information about the context in which an experimenter (E1) was operating. The information provided varied according to condition both in terms of (1) the causal relatedness of E1's action on the context boxes and (2) the retrieval of the grapes. In the experimental condition, it was necessary to manipulate the boxes to retrieve the grapes because they were inside the boxes, whereas in the control condition, it was not necessary because the grapes came out of E1's chest pocket.

After baiting the pocket or the context boxes out of the ape's view and setting up the test equipment, E1 sat down on a stool in front of the testing table, called the ape's attention and began to manipulate the context boxes one by one, from left to right, differently according to the experimental condition:

**Experimental condition** E1 slowly opened the latch of the first context box for 5–7 s, then lowered her hands to her lap (for approximately 2 s), then opened the lid of that box, retrieved a grape from out of the box, closed the lid, and handed the grape to the ape through the middle hole in the window. She then moved to the next box in the sequence and opened it following this procedure.

**Control condition** E1 acted on the latch of the first context box in the same way, for the same amount of time as in the experimental condition, lowered her hands for 2 s, and then without opening the box retrieved a grape from the chest pocket in her overalls and handed it to the ape through the middle hole in the window. She then repeated this behavior with the next box in the sequence.

On the first 2 days, she performed this procedure three times, and on the third day twice. For repetitions, after having manipulated all three boxes, E1 blocked the ape's view of the boxes with the occluder and restored the boxes to their original state with latches closed. In the control condition, in which E1 did not open boxes, she nevertheless occluded them and made some audible noise with the latches so as to match the other condition.

The testing phase, which consisted of a single trial, took place at the end of the third day. After E1 had acted on the context boxes as described above for two rounds, she put up the occluder and left the testing room. A second experimenter (E2) entered the testing room and attached the unfamiliar test box with the lid open to the right end of the table. E2 removed the occluder, called the ape to ensure that it attended, and then showed the ape a grape, placed it in the test box, closed the lid, and pressed on the lid to demonstrate that it was closed tight. E2 then handed the ape a grape and left to a distant corner of the testing room. E1 returned and started to manipulate the boxes from left to right. After she had given the ape the third grape, she called the ape's attention again and started to act on the test box by spinning the wheel on top of its lid. At first, E1 acted on the test box in much the same way as she had acted on the context boxes but then she did not stop spinning the wheel for a total of 120 s. She looked only at the box during this time (as she had done in the context phase). During these 2 min, the ape's behavior was coded (see below). After the 2 min had passed, the test was over. E2 signaled to E1, who stood up and left the room. E2 fed the ape another grape and retrieved the closed test box to store it away for the next ape.

To eliminate superficial cues, special care was taken in both conditions to display only a completely neutral facial expression and to look only at the boxes. Note that the two conditions thus varied only in E1's action during the context phase *prior* to the testing phase. E1's action on the test box during the testing phase was identical in both conditions.

### Coding and analyses

Two main types of response measures were scored from videotape (frame by frame) by the second author during the test phase, the 120 s that E1 acted on the test box:

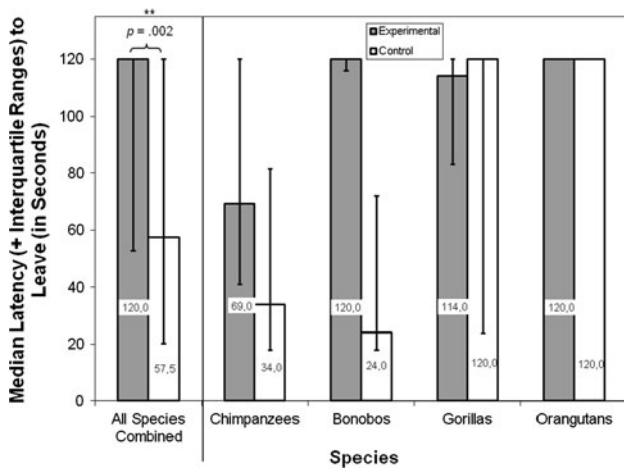
“participation” and “behavioral rate” (as was done in Call et al. 2004). *Participation* was the amount of time apes remained at the testing site. The corners of a large square (1 m<sup>2</sup>) were marked with paint on the floor in front of the plexiglass window inside the enclosure (see Fig. 2). An ape leaving this critical area was coded as leaving the testing site. Two different participation measures were scored: “latency” and “attendance.” *Latency* was the time in seconds from the moment E1 began manipulating the test box until the moment the apes first left (i.e., moved their whole body out of) the square. *Attendance* was the overall time in seconds accumulated in the square during the 2-minute response period. For example, an ape who left the testing site after 15 s, came back 20 s later, and stayed for another 35 s before leaving for the rest of the trial would accumulate an overall attendance of 50 s.

*Behavioral rate* was the frequency with which apes showed communicative behaviors to E1 while they were in the critical square during the test. For example, if a subject stayed at the testing site overall for 100 s and produced a behavior five times, the corresponding behavioral rate would be  $5/100 = 0.05$ . Following Call et al. (2004), we specifically looked for behaviors that express impatience and/or requests for food. These behaviors were (1) *bang-ing*: making an audible noise by hitting the plexiglass window or any part of the enclosure with any body part (two or more events with less than a second in between them were scored as one event—one bout) and (2) *poking*: sticking one or more fingers through one of the three holes at the bottom of the plexiglass window (each time one or more fingertips protruded through the plexiglass window it was scored as one event). No other behaviors were coded.

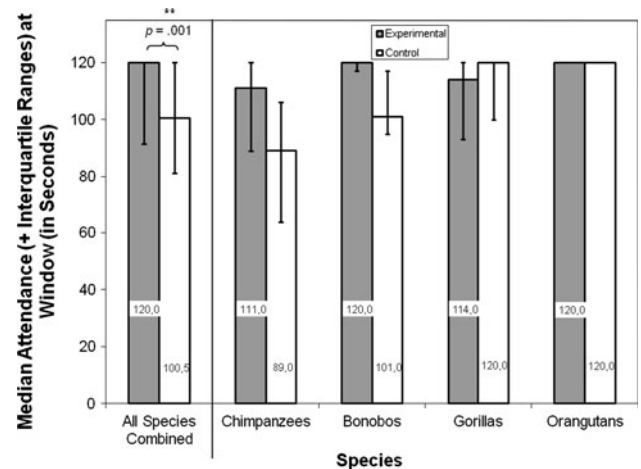
For inter-observer reliability, a second coder who was naïve to the hypotheses of the study and blind to condition scored a randomly chosen 20 % of the trials from all apes. Agreement was excellent: Spearman correlations (all  $p$ 's two-tailed) for attendance:  $r_s = 1.00$ ,  $N = 12$ ,  $p < 0.001$ , for latency:  $r_s = 0.998$ ,  $N = 12$ ,  $p < 0.001$ , and for behavioral rate:  $r_s = 0.975$ ,  $N = 12$ ,  $p < 0.001$ . The coders' judgments did not differ on any of the measures: Wilcoxon tests, all  $p$ 's  $\geq 0.161$ .

Our data severely violated the assumption of normality. Therefore, we used nonparametric statistics. For analyses across all apes (group level analyses), we conducted Wilcoxon exact tests. Further, Kruskal–Wallis  $H$  tests were used to analyze for differences between species. In particular, to assess the effect of species on the dependent variables, we calculated the difference between the experimental and the control condition for each individual and then compared the species as described. We conducted Wilcoxon exact tests for analyses on each species separately (species level analyses). Given our clear predictions regarding the direction of the expected effects, we report





**Fig. 3** Medians (+interquartile ranges) for latency to leave for all species combined and for each species separately in Study 1a



**Fig. 4** Medians (+interquartile ranges) for the number of seconds spent in front of the window for all species combined and for each species separately in Study 1a

one-tailed *p* values throughout Study 1 when comparing performance in the two conditions. We used Fisher’s Omnibus test (Haccou and Meelis 1994) to correct for multiple testing.

**Results**

*Participation*

*Latency* There were significant differences in latency between conditions at the group level: apes left the testing site significantly faster in the control than in the experimental condition,  $T^+ = 194.50$ ,  $N = 21$  (9 ties),  $p = 0.002$ ,  $r = 0.501$ . Although this effect was mainly caused by the chimpanzees’ and bonobos’ performances (see Fig. 3, Table 2), there were no significant differences in latency across species,  $H = 1.99$ ,  $N = 30$ ,  $p = 0.574$ .

*Attendance* Apes also spent significantly more time at the testing site in the experimental than the control condition,  $T^+ = 183.00$ ,  $N = 20$  (10 ties),  $p = 0.001$ ,  $r = 0.532$ . Again, although this effect was mainly caused by the

chimpanzees and bonobos (see Fig. 4, Table 2), there were no significant differences across species,  $H = 4.32$ ,  $N = 30$ ,  $p = 0.229$ .

*Behavioral rate*

There were no significant differences between conditions, with apes showing the behaviors we measured in an average of 7–8 % of the time in both conditions,  $T^+ = 196.50$ ,  $N = 26$  (4 ties),  $p = 0.432$ ,  $r = 0.033$ . There were also no significant differences in behavioral rate across species,  $H = 0.726$ ,  $N = 30$ ,  $p = 0.867$  (see Table 2).

There were no significant effects of gender (all  $p$ ’s  $\geq 0.174$ ) or testing order (all  $p$ ’s  $\geq 0.566$ ) for any of the measures, Mann–Whitney *U* tests.

**Discussion**

In this study, the apes’ performance as a group on two of the three measures supports the hypothesis that they took

**Table 2** Means (and SEM) for all three measures in Study 1a

	Measure					
	Latency (in sec.)		Attendance (in sec.)		Behavioral rate	
	Experimental condition	Control condition	Experimental condition	Control condition	Experimental condition	Control condition
All species	89.97 (7.24)	64.67 (8.32)	107.07 (3.28)	95.03 (4.98)	0.082 (0.014)	0.089 (0.015)
Chimpanzees	74.33 (11.01)	48.60 (10.12)	102.47 (4.83)	85.33 (6.82)	0.129 (0.022)	0.140 (0.021)
Bonobos	100.20 (18.82)	49.80 (20.38)	109.00 (10.27)	105.00 (5.72)	0.032 (0.014)	0.047 (0.024)
Gorillas	96.60 (14.39)	80.60 (24.14)	106.00 (7.60)	97.40 (18.02)	0.037 (0.018)	0.034 (0.017)
Orangutans	120.00 (0.00)	111.80 (8.20)	120.00 (0.00)	111.80 (8.20)	0.037 (0.002)	0.031 (0.009)

the experimenter's goal into account during the test: they waited longer before leaving the critical area and stayed in this area for a longer duration in the experimental condition in which it was natural to see the first experimenter's manipulation of the top of the box as an attempt to open it. They did this even though during the entire test phase the experimenter's behavior was identical in the experimental and control conditions. Although we found no significant differences between species, an inspection of the data indicated that the results were driven by the responses of the chimpanzees and bonobos. In contrast, the gorillas and orangutans had similar latency and attendance results in both conditions. This is perhaps especially surprising for orangutans, given that they do succeed on some tests of goal and intention understanding (see Buttelmann et al. 2008; Call and Tomasello 1998). One possibility is simply that gorillas and orangutans generally move less or more slowly than chimpanzees and bonobos, and that our response period was too short to reveal any differences between conditions.

It is important to be clear about the logic of this experiment in terms of what caused the difference in apes' behavior in the two conditions. One possible interpretation is the following. During the context phase, apes in both conditions repeatedly experienced moving from box to box, giving them food in association with each box with precisely the same timing (every 5–7 s). Then, for the test box, apes in both conditions were expecting food at the end of that 5- to 7-second interval: apes in the experimental condition from the box and apes in the control condition from the first experimenter's pocket. But in neither condition did the experimenter hand over any food at this point; instead he kept on turning the wheel on top of this box. So how long should the ape wait after that in the hopes of still receiving food? If apes were operating on pure temporal associations, they should have waited equally in the two conditions: in both conditions, in the context phase they had experienced repeatedly the experimenter manipulating the gadget on top of the box and then giving them food (either from the box or her pocket). The temporal associative strength between manipulating the top of the box and receiving food should thus have been identical in the two cases. However, we found that the apes in the experimental condition stayed longer. It is thus likely that apes waited longer in the experimental condition because they saw the experimenter's current actions as trying to open the box, that is, as having the (so far unsuccessful) goal of opening the box.

The key to this explanation is that apes in the control condition were expecting the food from the pocket in the same way that apes in the experimental condition were expecting the food from the box. When they did not get it—the script had changed—the question is why they

decided to stay or leave. We think that the apes in the control condition had little on which to base this decision: the human did not behave as expected, and it is unclear what she was doing still twirling the metal on top of the box. They thus had very little reason to expect a piece of food other than blind hope, and indeed, it is possible that the experimenter had simply changed her mind about giving them food at all (thus being unwilling). In contrast, apes in the experimental condition had seen the human in the past using the metal on the top of boxes to open them, and so they saw the current behavior as another version of this—albeit so far unsuccessful—with the first experimenter's desire to give them food still intact (but being unable to open this box). For them, then, it might thus pay to wait for the experimenter's success. This result fits well with the study by Call et al. (2004) who found that chimpanzees waited longer for an experimenter who was willing but unable to give them food compared to one who was able but unwilling.

A possible alternative explanation is that the apes in the control condition might simply have given up on the food from the pocket and they had no specific reason to expect the food from the box, but the apes in the experimental condition had experienced boxes opening in the past and so they expected this one to open too—without any consideration of the human's goals. This alternative, however, cannot explain why apes in the control condition gave up on the grape they were expecting (from the pocket), whereas in the experimental condition did not give up.

Nevertheless, the methodological point is that the control condition of this study could have been instantiated in various ways. Thus, another possibility would be to have the first experimenter open each of the context boxes in the control condition, as in the experimental condition, but then still give food from her pocket. We chose not to do this initially because we were afraid that if apes were not paying close attention, they would see the experimenter open the boxes and then give them food and assume it came from the boxes. That is, we chose the way we did because we wanted the difference between conditions to be clear to the apes. However, since they did differentiate between conditions in this first study, we decided to try this second, more demanding version of the control condition.

## Study 1b

In this study, we replicated the general procedure of the previous study, but now in the control condition, the first experimenter opened each box before she gave the ape food from her pocket. We predicted the same general pattern of findings as in Study 1a. We tested chimpanzees and bonobos, the two species that showed the largest

differences between conditions in latency in Study 1a (see Fig. 3). The mean delay between Study 1a and Study 1b was for chimpanzees 54 weeks and for bonobos 23 weeks.

## Methods

### Participants

The same five bonobos (mean age = 11.6 years, age range = 8–22 years; 3 males and 2 females) and 14 of the 15 chimpanzees that had participated in Study 1a participated in this experiment (one male chimpanzee that had been tested in Study 1a refused to participate and therefore was not tested). In addition, four other adult female chimpanzees that had not participated in Study 1a participated in this experiment, for a total of 18 chimpanzees (mean age = 14.8 years, range = 4–29 years; 5 males and 13 females), see Table 1 for details.

### Materials and design

The materials used in Study 1a were modified slightly for this experiment to reduce carryover effects. Specifically, the positions of the boxes were rearranged on the table and they were painted a different color (yellow). Apart from that, the materials, general setup, and the design of Study 1b matched that of Study 1a. Again, each ape participated in both conditions.

### Procedure

The same general procedure was used as in Study 1a. What differed were the details of the conditions presented to apes in the context phase. In this study, after the first experimenter (E1) opened each context box, she stood up before giving a grape to the apes in both conditions. We had E1 stand up because of the spatial closeness of E1's chest pocket and the boxes—to help apes see clearly where the grapes were taken from—either out of E1's chest pocket or out of the context box.

**Experimental condition** E1 knelt in front of the table supporting the boxes. She opened the latch of the first box, lifted the lid, stood up, retrieved a grape from out of the box, closed the lid, and handed the grape to the ape through the middle hole of the window. She then got down on her knees again and started the same procedure with the next box.

**Control condition** E1 opened the latch of the first box in the same way as in the other condition, opened the lid, stood up, closed the lid, retrieved a grape from out of her chest pocket, and handed it to the ape through the middle hole of the window. She then got down on her knees again and started the same procedure with the next box.

As before, E1's behavior during the test phase was identical in both conditions: she twisted a wheel on top of the test box (see Study 1a).

### Coding and analyses

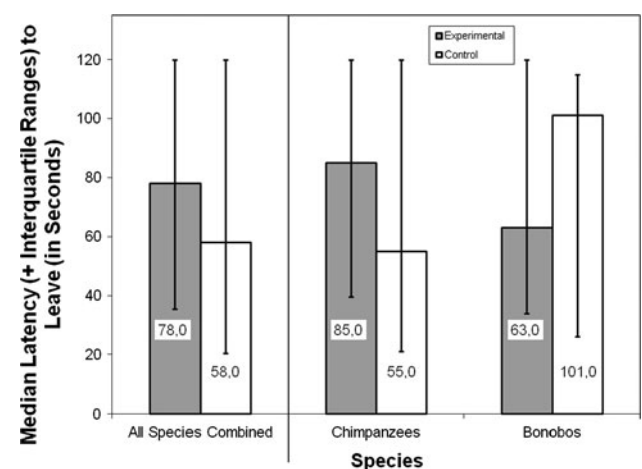
The apes' behavior was scored by the second author and analyzed in the same way as in Study 1a, with the exception that Mann–Whitney  $U$  tests were used to analyze for differences between species in this study. Again, a second, naïve coder scored a randomly chosen 20 % of the trials from all apes blind to condition, and interobserver reliability was excellent: Spearman correlations (all  $p$ 's two-tailed) for attendance:  $r_s = 1.00$ ,  $N = 12$ ,  $p < 0.001$ , for latency:  $r_s = 0.997$ ,  $N = 12$ ,  $p < 0.001$ , and for behavioral rate:  $r_s = 0.985$ ,  $N = 12$ ,  $p < 0.001$ . The coders' judgments did not differ on any of the measures: Wilcoxon tests, all  $p$ 's  $\geq 0.104$ .

## Results

### Participation

**Latency** Overall, there were no significant differences in latency between conditions,  $T^+ = 117.00$ ,  $N = 19$  (4 ties),  $p = 0.195$ ,  $r = 0.185$  and no significant differences between species,  $U = 34.0$ ,  $N = 23$ ,  $p = 0.434$  (see Fig. 5, Table 3).

**Attendance** Overall, there were also no significant differences between conditions on the time apes spent at the testing site,  $T^+ = 127.00$ ,  $N = 19$  (4 ties),  $p = 0.103$ ,  $r = 0.269$  and no significant differences between species,  $U = 25.50$ ,  $N = 23$ ,  $p = 0.154$  (Fig. 6, Table 3).

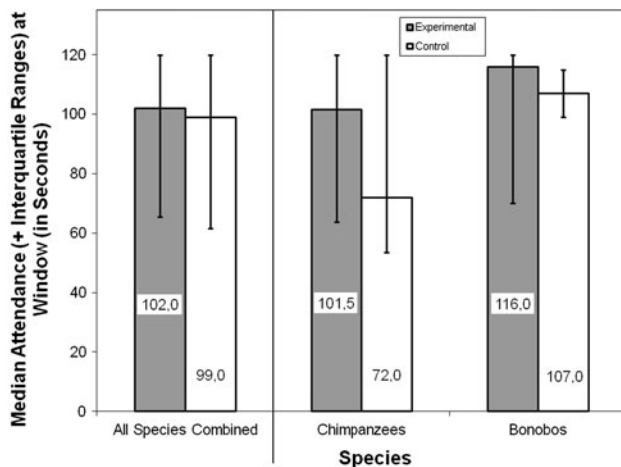


**Fig. 5** Medians (+interquartile ranges) for latency to leave for both species combined and for each species separately in Study 1b



**Table 3** Means (and SEM) for all three measures in Study 1b

	Measure					
	Latency (in sec.)		Attendance (in sec.)		Behavioral rate	
	Experimental condition	Control condition	Experimental condition	Control condition	Experimental condition	Control condition
All species	74.61 (8.80)	67.39 (9.63)	89.39 (7.01)	84.61 (7.57)	0.055 (0.011)	0.048 (0.010)
Chimpanzees	76.44 (9.58)	65.17 (10.88)	87.83 (8.13)	79.11 (9.14)	0.056 (0.012)	0.053 (0.011)
Bonobos	68.00 (23.25)	75.40 (22.69)	95.00 (14.89)	104.40 (6.85)	0.053 (0.031)	0.026 (0.019)

**Fig. 6** Medians (+interquartile ranges) for the number of seconds spent in front of the window for both species combined and for each species separately in Study 1b

**Behavioral rate** There were also no significant differences in behavioral rate between conditions, with apes showing the behaviors we measured in an average of 3–4 % of the time in both conditions,  $T^+ = 145.50$ ,  $N = 20$  (3 ties),  $p = 0.275$ ,  $r = 0.129$ . There were no differences between species,  $U = 32.00$ ,  $N = 23$ ,  $p = 0.363$  (Table 3).

There were no significant effects of gender (all  $p$ 's  $\geq 0.265$ ) or order (all  $p$ 's  $\geq 0.093$ ) for any of the measures, Mann–Whitney  $U$  tests.

#### Comparison of studies

Because most of the apes in this study had already participated in Study 1a, to see whether there were any carryover effects (e.g., habituation to the testing situation over the two studies), we compared the latency and attendance measures across the two studies. Wilcoxon exact tests on attendance showed that apes spent significantly less time at the testing site in Study 1b than in Study 1a in the experimental condition, suggesting some habituation,  $T = 82.00$ ,  $N = 13$  (6 ties),  $p = 0.008$ ,  $r = 0.585$ . This was not the case in the Control condition,  $T = 94.50$ ,  $N = 17$  (2 ties),  $p = 0.411$ ,

$r = 0.195$ . No difference between studies was found for latency in either condition, both  $p$ 's  $\geq 0.194$ .

#### Discussion

The results of Study 1a were not replicated in this study. This is probably due to three factors. First, the experimental and control conditions in this study were extremely similar, and any lapses of attention on the apes' part could easily lead to the two conditions appearing identical. Second, because this study was run after the first study, apes may have habituated to the situation. Indeed, overall, apes waited in the critical area less in the experimental condition in this study than they did in the first study. Third, apes in the control condition of this study had at least some reason to expect two pieces of food (the grape from the test box and one from the experimenter's pocket), whereas apes in the experimental condition had reason to expect only one (the one from the test box). One or more of these reasons might have prevented the apes from showing an understanding of others' goals in the absence of immediate behavioral cues in this experiment.

A possible alternative explanation of both studies is that the apes were not attending to the first experimenter's goals, but only to the causal relations involved in the situation. Thus, in the control condition, there was no causal relationship between the experimenter opening the box and apes' receipt of food from his pocket, whereas in the experimental condition, there was a causal relation. However, this causal relation is not physical causality. In the experimental condition, the opening of the box is causally related to the ape getting food only in the sense that it enables the experimenter to go on to perform the goal-directed action of grasping the food and giving it to the ape. To say that in the experimental condition, opening the box was causally related to the experimenter's giving the food (in a way that it was not in the control condition) is simply to say that there were physical conditions that enabled the experimenter's goal-directed action. Given this analysis, we might then predict that at test, when the experimenter is manipulating the mechanism on top of the box and it does not work in the normal amount of time, if the apes were

only focused on the causality of the box, they should conclude that the normal causal sequence was a failure—and so they should leave. The fact that they stayed suggests that they understood that so long as the experimenter had the goal of giving them food, and opening the box was simply an enabling condition for that, then they should stay to see if her further attempts (driven by her goal of giving them food) would succeed.

It is also relevant that in the control condition, the experimenter finishing manipulating the box was, at least in a weak sense, an enabling condition for the experimenter reaching into her pocket to get the food—as she could not do two things at once with her hands—and so the two conditions were not so different in terms of causality after all. Given this analysis, if the apes were only focused on how free hands enable food acquisition then in the control condition, at test they should wait until the experimenter finished manipulating the box, as that would enable her to get and give the food; but, at least in the first experiment, they did not do this. In any case, we think that any causal relations that might have played a role in the current study did so within the context of the experimenter's goal-directed actions. However, recognizing that it is difficult to determine whether the apes were using causal or intentional understanding in this study, and in Study 2, we adopted a different paradigm in which there was basically no physical causality involved.

## Study 2

To further investigate the question of whether great apes can understand the goals of others in the absence of concurrent behavioral cues based only on context, we ran another study with chimpanzees using a second, very different methodology. The basic idea was that animals quite often anticipate what an individual will do based only on an “intention movement” at the start of the behavior (Tinbergen 1953). We manipulated the context to see whether chimpanzees would interpret what a human was doing differently—even though her behavior during the test was identical in all conditions—as a function of what potential goals were available. In fact, we manipulated the situation so that chimpanzees would normally be expecting one thing to happen, based on past experience, but then some novel event intervened that could potentially lead to a different prediction, if the chimpanzees understood that this might lead to a change in the actor's goal.

Thus, a human sat on a stool giving chimpanzees food from a bucket through a mesh panel. There was also a second bucket containing food in front of another mesh panel some meters away. On several occasions in a pretest, the human stood up from her stool, went over to this other

bucket, and gave the chimpanzee food from it. In this situation, the chimpanzees quickly learned to anticipate as soon as the human stood up what she was going to do, so they headed for the second bucket straightaway. In the test phase of the experiment—identically in all conditions—the experimenter again stood up from her stool and turned her body in the direction of the second bucket. However, in the experimental condition (which was instantiated in three different ways), something happened immediately prior to the experimenter standing up, for example, a call came from a walkie-talkie in the same direction as the second bucket, or another human threw a clipboard to the experimenter and it landed short (again, in the same direction as the second bucket). The question was whether the chimpanzees would predict that the human was again heading for the other food bucket, as she had done previously in the same situation, or whether they would instead correctly assume that that goal had been preempted by a new goal to get the walkie-talkie or the clipboard as more immediate and salient needs.

As a comparison to the experimental condition, we included two other conditions. In the control condition, again a similar event occurred immediately prior to the experimenter standing up, but the difference was that the experimenter herself caused this event, and so it should not be expected to preempt her goal of feeding the chimpanzee from the other bucket. This condition was included in order to check whether chimpanzees' response behavior could be influenced only by the pure presence or distraction of an event occurring. In the other condition, the baseline condition, no such event happened at all.

If our conclusions from Study 1 were correct, chimpanzees would interpret exactly the same behavior of the experimenter differently depending on their understanding of the goal the experimenter was pursuing by standing up—even though their past experience in this situation should lead them to expect that the experimenter would be headed for the other food bucket. Specifically, they should leave for the other bucket more quickly in the control and baseline conditions than in the experimental condition (in which the experimenter seemingly had another goal in standing up). In contrast, if chimpanzees paid attention only to concurrent behavioral cues, they should not show any difference in reaction between conditions: the experimenter's action during the test phase was identical, thus subjects should leave equally quickly in all conditions.

## Methods

### *Participants*

Thirteen of the chimpanzees (*Pan troglodytes*) that had participated in Study 1b and two additional chimpanzees

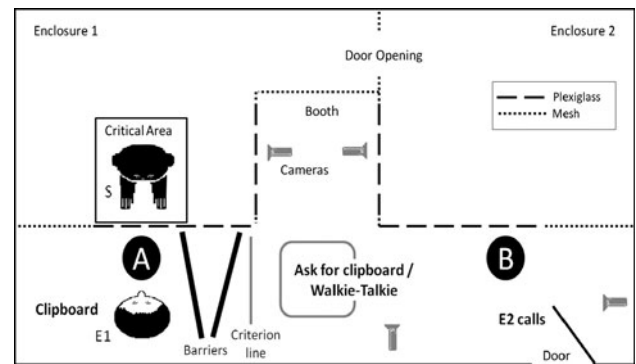
participated in this study. However, the youngest male and one of the two additional chimpanzees, a female, had to be dropped after they had received the first trials because of inattentiveness. The final group of 14 chimpanzees ranged in age from 7 to 31 years, with a mean age of 19.2 years (females: mean age = 20.4 years, range = 7–30 years; males: mean age = 15.8 years, range = 9–31 years).

### Materials and design

Test sessions took place in two of the chimpanzees' familiar enclosures (approximately 15 m<sup>2</sup> each). The cages were connected by a hydraulic door that could be adjusted at different heights. Since all subjects were quite familiar with crossing through this door, they knew that if the door was opened completely, it was easy for them to cross through very quickly. However, if one closed the door partially, the apes' speed in crossing could be reduced until a point where they would still fit through the door but would refuse to cross.

See Fig. 7 for a sketch of the setup. Two plastic buckets (25 cm in diameter), unfamiliar to the chimpanzees, with lids made of pieces of cardboard (22 × 31 cm) attached on top, were placed in front of plexiglass windows at opposite sides (location A and B) of the zookeepers' hallway in front of the chimpanzees' enclosures. The buckets were filled with grapes. A plastic stool (25 cm high) was placed behind (from chimpanzees' perspective) each of those buckets. Two plastic barriers (100 × 50 cm each) were put between location A and location B right next to location A. The barriers served to increase the time E1 needed to walk from one location to the other and thus give chimpanzees more time to make a prediction about where E1 might go in each trial. One clipboard was placed behind the stool at location A and a second clipboard was placed outside the test room (to be used in the test session). A pair of walkie-talkies (13 × 5 cm each) was used in one of the conditions. Both clipboards and walkie-talkies were familiar to the chimpanzees. We marked the edges of an area of approximately 1 square meter with yellow paint on the floor of the chimpanzees' side of the window at location A—the "critical area" to be used for coding. Each session was videotaped from four different angles (see Fig. 7).

A within-subjects design was used. There were two sessions: a warm-up session and a test session. In the test session, each chimpanzee participated in all conditions (see below). For each condition, there were different numbers of trials with one to four trials each, for a total of 11 trials. These trials were presented in counterbalanced order in such a way that not more than two trials in a row were of



**Fig. 7** The setup used in Study 2: E1 feeding *S* from one of two locations (A and B) with locations of the materials used in the control and the three experimental conditions indicated

the same condition. The test session lasted about 20 min per subject.

### Procedure

**Warm-up session** All chimpanzees were given a warm-up session in which they were fed by Experimenter 1 (E1) at the two locations (A and B). Because of constraints on available testing time, this session took place 3 months before chimpanzees participated in the test session. In this warm-up session, only the stools and the covered buckets containing food were present (there were no barriers, walkie-talkies, or clipboards). First, E1 fed the chimpanzee from bucket A, then got up and walked over to location B, sat down, and fed the chimpanzee from bucket B. After feeding the chimpanzee, there she got up again and walked back to location A to start a new trial. The number of grapes given each time at each location varied between one and three (randomized), so that chimpanzees could not predict when feeding at one location was over within each trial. Each time E1 walked from location A to location B, Experimenter 2 (E2) observed the chimpanzee's moving behavior and adjusted the height of the door opening between the two cages accordingly. That is, if the door was fully open, chimpanzees were likely to pass through immediately, whereas if the opening was smaller, they hesitated more to cross through (and at some point refused to cross). During these trials, E2 thus adjusted the height of the door up or down until each chimpanzee met the following criterion: they should not have entered Enclosure 2 before E1 had crossed the criterion line but should have entered this enclosure completely by the time E1 reached location B. Thus, for each chimpanzee, a different door height was determined at which he or she entered Enclosure 2, meeting the criterion. As soon as the chimpanzee met this criterion twice in a row, after feeding the

chimpanzee from bucket A once, E1 stood up, walked in the direction of location B, but this time turned left and pretended to arrange the cameras in the booth. She then went back to location A and continued feeding. Thus, this “camera check” was an interruption of the feeding process at location A and was performed in order to inform chimpanzees that E1 did not go immediately to location B every time she got up at location A.

*Test session* For the test session, when chimpanzees entered the test room, the two covered buckets filled with grapes were already in place in front of the windows at locations A and B, with the stools behind them and the barriers near location A. There was a clipboard behind stool A. The door was adjusted to the specific criterion for that chimpanzee and then E1 and E2 started to talk using the walkie-talkies for about five minutes to ensure that chimpanzees noticed that these were functionally equivalent to the ones usually used by their keepers and other researchers. During that time, they set up the four cameras used to videotape the test session. After setting up the cameras, E1 placed her walkie-talkie at the location shown in Fig. 7 and sat down on stool A to start feeding the chimpanzee from that bucket.

To ensure that the chimpanzees remembered the basic feeding procedure from the warm-up session and to introduce the barriers, chimpanzees were given further warm-up trials. In these warm-up trials, E1 fed the chimpanzee from bucket A, then got up and stepped over the barriers, looking only down at the barriers or at the wall behind them while doing so. She then briefly paused in front of an inconspicuously marked line (the “criterion line”; see Fig. 7) before she stepped over it and walked over to location B to feed the chimpanzee there. After feeding from location B, E1 walked back to location A again and started the next warm-up trial. As in the warm-up session, the number of grapes given to a chimpanzee at each location varied between one and three for each trial. During this warm-up period, E2 again observed the chimpanzee during the time it took E1 to walk from location A to location B (approximately 17 s) to ensure that they did not enter Enclosure 2 before E1 had passed the criterion line but had entered this enclosure before E1 had sat down on the stool at location B. As soon as this criterion was reached (but after a minimum of at least five warm-up trials), E1 repeated the “camera check” as described for the warm-up session. After this camera check, the test trials began.

As in the warm-up trials, E1 began by feeding the chimpanzee at location A. However, during test trials, in some conditions (experimental and control), some event made E1 stop feeding, whereas in other conditions (baseline), no external event could be observed before she stopped feeding. In all conditions, E1 then turned her head

toward the direction of location B, looking toward the stool at this location, looked back to the chimpanzee, got up, turned toward the barriers, stepped over each of them while looking at them or the wall behind them, and paused for a second in front of the criterion line. What she did then varied between the conditions. However, it is important to note that up until E1 paused in front of the criterion line—during which time chimpanzees’ behavior was measured—E1’s behavior in every condition (including baseline and control conditions) was identical.

*Experimental conditions* In all of the trials of the experimental condition, there was an observable event that caused E1 to stop feeding the chimpanzee at location A and get up. We expected that this event would lead chimpanzees to be less likely to predict that E1 was getting up to go to location B to feed them there. In order to avoid repeating the same event multiple times, we used three different events corresponding to the following three conditions:

*Walkie-talkie* While E1 was feeding the chimpanzee, E2, who was outside the room, called E1 through the walkie-talkie placed between location A and location B. E1 looked ambiguously in the direction of the stool at location B and the walkie-talkie, looked at the chimpanzee, got up, crossed the barriers, and paused in front of the criterion line. Then, she took one more step, bent down, picked up the walkie-talkie from the floor and said, “Yes, I can hear you.” She then turned off the walkie-talkie and put it down again.

*Ask for clipboard* E1 stopped feeding the chimpanzee, looked ambiguously in the direction of the stool at location B and the door by looking at a point in between these two locations, and called E2’s name. E2 opened the door and answered “Yes?”, and E1 asked him to give her a clipboard, performing a chimpanzee-like request gesture with her right arm. E2 threw a clipboard into her direction such that it landed at an inconspicuously marked spot between location A and B. E1 turned back to the chimpanzee and got up as before. After she had paused in front of the criterion line, she made one more step, bent down, and picked up the clipboard.

*E2 calls* From the door, E2 called E1 by name and asked her to come outside the test room with a beckoning gesture. As soon as E2 called her E1 stopped feeding, looked ambiguously in the direction of the stool at location B and the door while talking to E1, then looked back to the chimpanzee, got up, and crossed the barriers as before. After the short pause in front of the criterion line, she walked directly toward location B but before getting there turned right and left the test room for about 40 s. She then returned.

After each of these trials, E1 went back to location A and continued feeding the chimpanzee there. Each



chimpanzee received each of these trials once (to avoid learning effects) in counterbalanced order.

**Control condition (drop clipboard)** In the one trial in this condition, a similar observable event occurred that caused E1 to stop feeding chimpanzees at location A, but we expected it not to alter chimpanzees' prediction that E1 would go feed them at location B, because it was caused directly by E1 herself. After E1 sat down on stool A, she reached behind her, picked up the clipboard, and put it on her lap. She then fed the chimpanzee, stopped, took the clipboard with both hands, and dropped it intentionally on the floor at her left side. Then, she looked ambiguously in the direction of the stool in location B, back to the chimpanzee, got up, stepped over the barriers, paused, and walked over to location B to feed the chimpanzee there.

**Baseline conditions (baseline and post-experimental baseline)** In the baseline conditions, no observable event occurred before E1 stopped feeding, so in general we expected chimpanzees to predict that E1 was getting up to go feed them at location B. E1 simply stopped, looked ambiguously in the direction of the stool at location B, back to the chimpanzee, got up, stepped over the barriers, paused, and walked over to location B to feed the chimpanzee there. There were two types of baseline trials, those which immediately followed experimental trials (post-experimental baseline; three trials) and those which did not (baseline; four trials). The procedure for both was identical; we distinguished the two types simply because we thought that in the post-experimental baseline trials, the fact that chimpanzees had just previously experienced E1 not approaching location B in the experimental conditions might result in a lower likelihood of leaving the critical area quickly in succeeding trials. The post-experimental baseline trials also served to boost chimpanzees' performance back up to pre-experimental trials levels.

We measured chimpanzees' leaving behavior from the time E1 looked at stool B before getting up from stool A until she paused briefly in front of the criterion line. Thus, E1's behavior during the period in which we measured chimpanzees' response behavior was identical in all conditions, and chimpanzees could make no predictions about where E1 might go solely from her actions during that period.

#### *Coding, reliability, and analysis*

The first author coded chimpanzees' leaving behavior from the videotapes, blind to condition. We analyzed the data in two different ways. First, the main analysis was done on the number of seconds chimpanzees waited before leaving the critical area in each response phase (i.e., from when E1 looked toward stool B to when E1 paused after having

**Table 4** The scale used for coding chimpanzees' performance in relation to E1's actions

Step	E1's behavior
1	E1 looks at stool B
2	E1 looks at S again
3	E1 gets up
4	E1 turns
5	1st leg lifted (in front of barrier 1)
6	1st foot visible above barrier 1
7	1st leg lowered
8	E1 stands above barrier 1
9	2nd leg lifted
10	2nd foot visible above barrier 1
11	2nd leg lowered
12	E1 stands between the barriers
13	1st leg lifted (in front of barrier 2)
14	1st foot visible above barrier 2
15	1st leg lowered
16	E1 stands above barrier 2
17	2nd leg lifted
18	2nd foot visible above barrier 2
19	2nd leg lowered
20	E1 pauses after having stepped over both barriers
21	1st step
22	2nd step
23	3rd step
24	4th step
25	5th step
26	E1 turns (in order to sit down)
27	E1 takes a seat
28	E1 sits on stool B
29	S does not leave critical area

Explanation: a step lasts from lifting the foot to setting the foot down again. In between steps add .5 to the step just finished

stepped over both barriers). Second, we matched chimpanzees' response behavior to E1's action more precisely by coding E1's behavior at the moment the chimpanzees left the critical area using a scale consisting of 29 different steps (see Table 4). Leaving the critical area was defined as the moment when all four of the chimpanzees' limbs had left the marked square by the window. To assess interobserver reliability, a naïve coder also watched the videotapes, blind to condition, and rated 25 % of the trials in both ways. An excellent level of interobserver agreement was reached: Spearman's correlations (all  $p$ 's two-tailed) for seconds:  $r_s = 0.946$ ,  $p < 0.001$  and for the scale:  $r_s = 0.933$ ,  $p < 0.001$ . The coders' judgments did not differ on any of the measures: Wilcoxon tests, all  $p$ 's  $\geq 0.097$ .

For statistical analysis and comparison across conditions, we calculated means for the three experimental, the

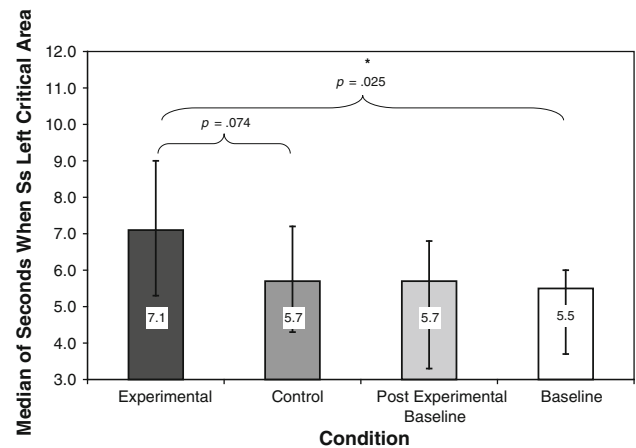


four baseline, and the three post-experimental baseline trials for each chimpanzee. Since during testing, it became obvious that females seemed to pay more attention than males, and we also analyzed both genders separately. Unfortunately, statistical tests could only be calculated for females ( $n = 10$ ) because of the males' small sample size ( $n = 4$ ). Friedman tests were run to test for differences in the chimpanzees' performance across conditions overall. Then, based on those results, protected Wilcoxon tests were used to determine specific differences between conditions (see Cohen and Cohen 1983). Given our clear predictions regarding the direction of the expected effects (as based on previous studies and the results of Study 1a), again, we report one-tailed  $p$  values throughout.

## Results

Mean latencies for each condition are presented in Table 5. The chimpanzees behaved differently in the different conditions in terms of latency to leave, Friedman's  $\chi^2_{(3)} = 10.63$ ,  $N = 14$ ,  $p = 0.011$  (exact, two-tailed). When single conditions were compared, a large effect was found only between the experimental and the baseline condition, Wilcoxon  $T^+ = 84.00$ ,  $N = 14$ ,  $p = 0.025$ ,  $r = 0.528$ , but there was a moderate (but nonsignificant) effect for the comparison between the experimental and the control condition,  $T^+ = 76.00$ ,  $N = 14$ ,  $p = 0.074$ ,  $r = 0.394$ , see Fig. 8. Similar results were found when we analyzed the data using the scale: a large effect between the experimental (mean: 11.81 steps) and the baseline (mean: 9.65 steps) condition, and a moderate effect for differences between the experimental and the control (mean: 10.71 steps) condition (with the same  $p$  values as found for the analysis on seconds).

When we analyzed the females' performance separately, we found that they showed significant differences between conditions, Friedman's  $\chi^2_{(3)} = 10.32$ ,  $N = 10$ ,  $p = 0.012$  (exact, two-tailed). When single conditions were compared, we found that as a group females significantly waited longer in front of the window in the experimental condition than in the baseline, the post-experimental baseline, and the control conditions (experimental baseline:  $T^+ = 46.00$ ,  $N = 10$ ,  $p = 0.032$ ,  $r = 0.596$ ; experimental control:  $T^+ = 45.00$ ,



**Fig. 8** Medians (+interquartile ranges) for seconds chimpanzees waited before leaving the critical area at location A in Study 2 (For baseline, post-experimental baseline, and experimental trials means were calculated for each subject; presented here is the median of the subjects' means)

$N = 10$ ,  $p = 0.040$ ,  $r = 0.564$ , and experimental post-experimental baseline:  $T^+ = 46.00$ ,  $N = 10$ ,  $p = 0.030$ ,  $r = 0.597$ ), see Fig. 9. No significant differences were found between the baseline, the post-experimental baseline, and the control conditions (all  $r$ 's  $\leq 0.194$ ). Again, similar results were found when we analyzed the data using the scale: significant differences between the experimental (mean: 10.54 steps) condition and the baseline (mean: 8.10 steps), post-experimental baseline (mean: 9.50 steps), and control (mean: 9.40 steps) conditions (with the same  $p$  values as found for the analysis on seconds).

Figure 9 also shows the results for the four male chimpanzees. In general, in all conditions, they waited longer before they left the critical area than did female chimpanzees and, unlike the females, showed fewer signs of distinguishing between the conditions.

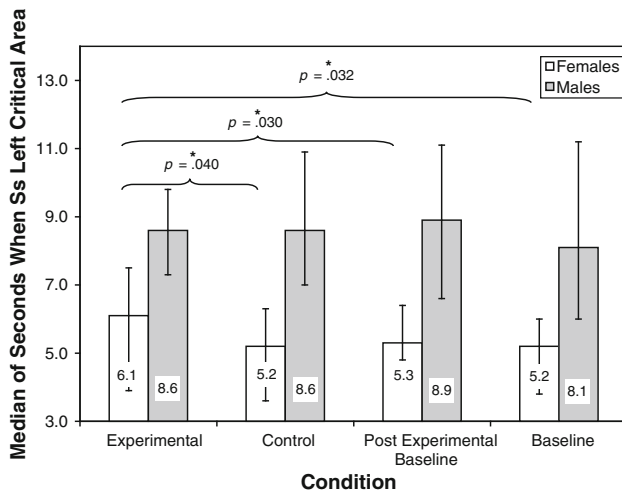
## Discussion

This study provides converging evidence, using a different methodology, for the findings of the first study. The chimpanzees we tested, in particular the female chimpanzees,

**Table 5** Mean latency to leave (in seconds) for each trial in Study 2

Sex	Condition										
	Experimental			Control	Post-experimental baseline			Baseline			
	WT	AFC	E2C		DRC	PB1	PB2	PB3	B1	B2	B3
Both	8.2	5.8	7.7	6.7	7.6	5.8	7.2	5.7	6.0	6.2	5.7
Females	7.2	5.3	7.5	6.2	7.0	4.6	6.3	4.9	5.0	5.5	5.6
Males	10.8	7.2	8.1	7.9	9.1	8.6	9.3	7.7	8.1	8.7	6.2

WT walkie-talkie, AFC ask for clipboard, E2C E2 calls, DRC drop clipboard, PB post-experimental baseline, B baseline



**Fig. 9** Medians (+interquartile ranges) for seconds chimpanzees waited before leaving the critical area at location A for the groups of female and male chimpanzees separately in Study 2 (For baseline, post-experimental baseline, and experimental trials means were calculated for each subject; presented here is the median of the subjects' means)

interpreted what the human experimenter was doing differently in the different conditions, even though her actions were identical in all cases during the test phase. Indeed, in two of the experimental conditions (the walkie-talkie and E2 calls conditions), the experimenter's behavior was even identical to that in the baseline condition *before* the test—the only thing that differed was what happened in the environment. The use of concurrent behavioral cues was thus not possible at any phase in these conditions.

One could argue that the chimpanzees could have been using some learned context cues involving walkie-talkies and the like. This seems unlikely, however. There exists no specific information on the chimpanzees' exact history with these or similar situations in the past, but watching one person stand up to go retrieve a clipboard that another person has thrown too short is almost certainly novel for all chimpanzees—clipboards are almost always right next to the human or even in their laps. And although they have seen humans answer walkie-talkies, it is usually by retrieving it from their own belt or pocket. Most importantly, in the control condition, the experimenter threw down her clipboard and stood up, but in this case, the chimpanzees apparently saw things differently since she threw it down herself. So context here cannot be defined as a clipboard on the ground because they behaved differently if the experimenter threw it on the ground herself or another human threw it (short) to her. This condition also argues against the possibility that the pure presence or distraction of an event occurring is what made chimpanzees wait longer in the experimental condition.

The key point is this. When the experimenter stood up from her stool in the test phase, based on their past experience in exactly the same physical context with exactly the same experimenter, chimpanzees should have expected her to go over to the other bucket of food (and indeed their behavior in the pretest shows that they did come to expect this). However, they did not: Instead, they changed their prediction of what the experimenter was doing based on a novel event—one with which they had limited if any experience in the current context. In order to predict that the experimenter was doing anything other than going to the other bucket of food, they would have had to override recently learned behavioral rules (the experimenter going to the food bucket) in favor of something else. Our proposal for that something else would be that they understood the experimenter's new goal based on a general understanding of the kinds of things that humans want and find salient.

Another way of highlighting the difference between the two kinds of explanation—concurrent behavioral cues versus understanding goals—is this. In the current study, the chimpanzees could have been predicting what the human would do only on the basis of context. The question is what we mean by context. Our proposal would be that it does not mean something simple like the presence of a particular object, or else there would have been no difference in performance in the experimental and control conditions. What it means, in this case, is a problem facing the human—an out-of-reach clipboard she wants, a far-away walkie-talkie she must answer, or the beckoning of another human she must respond to. Understanding the context not in terms of surface features like objects but rather in terms of problems for an actor means precisely understanding that for her the current situation is not the desired situation, and so her goal is to eliminate this discrepancy.

The fact that in this study, significant results were shown mainly by the female chimpanzees is reminiscent of findings of sex-based learning differences in chimpanzees in the wild (Boesch-Achermann and Boesch 1993; Lonsdorf 2004, 2005). Female chimpanzees start to fish for termites earlier in their ontogeny than male chimpanzees do, and they are also more proficient at this skill after its acquisition than their male conspecifics, perhaps because they are more attentive to their mothers' performance in social learning situations. It thus may be that females generally pay more attention in feeding situations and/or are less distractible than males and that gives them an advantage in both these situations.

## General discussion

In the current studies, we found that the great apes tested responded differently to exactly the same human behavior

in different contexts. Our proposal is that they were using these contexts to determine the human's goal. Previous studies have always compared situations in which the human's behavior was slightly different when he had one goal rather than another—and so the apes could always have been reacting to some kind of concurrent behavioral cues. However, in the current studies, the use of such behavioral cues was not possible during the test phase—or at any time in some conditions in Study 2.

Still, as noted above, there could conceivably be an analog to behavioral cues that we might call contextual cues. The apes could have learned some kind of association between specific contexts (behavioral cues that precede the test phase) and an actor's behavior. As with behavioral cues, this has some plausibility in familiar, repetitive contexts, but not in more novel contexts, and in the current studies, we created novel contexts for the apes. Moreover, the way these contextual cues might have worked was very different in the two studies. In Study 1, we arranged things so that if the apes were simply associating the human's previous behavior or events with outcomes, they should have waited in anticipation for food equally long in both conditions. A more plausible interpretation, therefore, is that in the experimental condition, the apes simply saw the experimenter as trying to open the box—which meant that they should wait, which they did.

In the second study, we first led the apes to establish an expectation that when the human stood up from her stool, she was headed for the other food bucket. Thus, if the apes predicted what the experimenter was likely to do based only on her previous behavior, they should have predicted this same sequence of events always and equally in all conditions, but they did not. In the key conditions, they took into account a unique event that had just happened: someone threw the experimenter a clipboard (short), or someone called the experimenter, or the experimenter's walkie-talkie made noise. On the basis of this unique event—not previously experienced with this experimenter in this situation—they now predicted that she would do something different, because she now had a different goal. Of course, it is possible that in their past experience, the apes have seen someone retrieve a clipboard, and so they are predicting that is what will happen now. However, when the experimenter threw down her own clipboard, they did not predict that she would retrieve it but rather thought that now she would be going to the other bucket. Thus, unlike Study 1 in which context was operationalized as just-previous behavior, here context had something to do with what was happening in the situation; in our interpretation, given the results of the control condition in which the experimenter threw down her own clipboard, context really means a new goal-relevant possibility in the situation for the human.

It is useful to recall here again two other studies that may be interpreted as apes understanding others' goals. First, Warneken and colleagues (2006, 2007) found that when a human had a problem such as an out-of-reach object, chimpanzees sometimes helped her by retrieving it for her, whereas they did not do this if she had thrown it away. Second, Tomasello and Carpenter (2005), Buttelmann et al. (2007) both conducted studies in which chimpanzees imitated human behavior, and they did so differently depending on how they interpreted the human's goals and intentions. The point is that both of these sets of studies have response measures—retrieving objects and imitating—that are not easily explained as an outcome of behavioral or contextual cues because the response is to help or to imitate the human appropriately—not just to anticipate.

One might wonder why some of our results are not particularly strong. One possibility is that in both studies, our apes might have thought they had some chance of getting food in both conditions. In Study 1, even though the experimenter did not previously have the goal of opening boxes, she might be willing to do so and therefore waiting in the critical area might be useful. In Study 2, it was not very costly to walk over to the other food bucket, so apes might have taken a chance and gone there just in case the experimenter would give them some food there on her way out of the room or when coming back after getting the walkie-talkie. However, the main reason that the results were not particularly strong is surely the relative ambiguity of the situations: we deliberately stripped the situations of all the typically present behavioral cues that could be used to help infer the experimenter's goal: gaze direction, facial expressions, and effortful action. Without these cues, apes' task is surely much more difficult and this is reflected in the results.

All in all, it is difficult to imagine that the totality of ape social interaction depends on their learning specific behavioral and contextual cues for predicting what others will do in every situation. It is difficult because, first, they would have to learn probably thousands or tens-of-thousands of such specific cues, maybe differently for different individuals. Many of these cues would require learning complex conditional discriminations between arbitrary stimuli which, as Call (2007) has noted, chimpanzees learn only with much difficulty or not at all. It is also difficult because, second, those rules would be of no help in novel situations—and apes seem to adapt immediately to novel situations such as those in the current studies as well as in previous studies. Of course, some learning is involved in all of this. An individual can discern the particular goal of another individual in a particular situation only on the basis of some kind of behavioral or contextual cue. However, the fact that some experience and/or learning is involved does

not mean that the outcome is not an understanding of goals. We think a very reasonable way to explain the apes' behavior in the current study, and other studies of the same and similar phenomena, is to credit them with an understanding of others as goal-directed agents who act when and because the currently perceived situation does not match their goal.

**Acknowledgments** We thank Anke Bullinger, Johannes Grossmann, Yvonne Rekers, Katja Große, Franziska Zemke, Katharina Kirchhofer, Denise Göhler, and Alexandra Rosati for help conducting this study, and Roger Mundry for statistical assistance, and Thomas Bugnyar for helpful comments on an earlier version of this manuscript. Please address correspondence to David Buttelmann at david.buttelmann@uni-erfurt.de.

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Boesch-Achermann H, Boesch C (1993) Tool use in wild chimpanzees: new light from dark forests. *Curr Dir Psychol Sci* 2:18–21. doi:10.1111/1467-8721
- Buttelmann D, Carpenter M, Call J, Tomasello M (2007) Enculturated chimpanzees imitate rationally. *Dev Sci* 10:F31–F38. doi:10.1111/j.1467-7687.2007.00630.x
- Buttelmann D, Carpenter M, Call J, Tomasello M (2008) Rational tool use and tool choice in human infants and great apes. *Child Dev* 79:609–626. doi:10.1111/j.1467-8624.2008.01146.x
- Call J (2007) Past and present challenges in theory of mind research in primates. In: van Hofsten C, Rosander K (eds) *Progress in brain research*, vol 164. Elsevier, Amsterdam, pp 341–354. doi:10.1016/S0079-6123(07)64019-9
- Call J, Tomasello M (1998) Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and human children (*Homo sapiens*). *J Comp Psychol* 112:192–206. doi:10.1037/0735-7036.112.2.192
- Call J, Hare B, Carpenter M, Tomasello M (2004) 'Unwilling' versus 'unable': chimpanzees' understanding of human intentions. *Dev Sci* 7:488–498. doi:10.1111/j.1467-7687.2004.00368.x
- Carpenter M, Akhtar N, Tomasello M (1998) Fourteen- to 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behav Dev* 21:315–330. doi:10.1016/S0163-6383(98)90009-1
- Carpenter M, Call J, Tomasello M (2002) Understanding 'prior intentions' enables 2-year-olds to imitatively learn a complex task. *Child Dev* 73:1431–1441. doi:10.1111/1467-8624.00481
- Cohen J, Cohen P (1983) *Applied multiple regression/correlation analysis for the behavioral sciences*, 2nd edn. Lawrence Erlbaum, New Jersey
- Gergely G, Bekkering H, Király I (2002) Rational imitation in preverbal infants. *Nature* 415:755. doi:10.1038/415755a
- Haccou P, Meelis E (1994) *Statistical analysis of behavioral data. An approach based on time-structured models*. Oxford University Press, Oxford
- Lonsdorf EV (2005) Sex differences in the development of termite-fishing skills in wild chimpanzees (*Pan troglodytes schweinfurthii*) of Gombe National Park, Tanzania. *Anim Behav* 70:673–683. doi:10.1016/j.anbehav.2004.12.014
- Lonsdorf EV, Pusey AE, Eberly L (2004) Sex differences in learning in chimpanzees. *Nature* 428:715–716. doi:10.1038/nature02527
- Meltzoff AN (1995) Understanding the intentions of others: reenactment of intended acts by 18-month-old children. *Dev Psychol* 31:1–16. doi:10.1037/0012-1649.31.5.838
- Povinelli DJ, Perilloux HK, Reaux JE, Bierschwale DT (1998) Young and juvenile chimpanzees' (*Pan troglodytes*) reactions to intentional versus accidental and inadvertent actions. *Behav Process* 42:205–218. doi:10.1016/S0376-6357(97)00077-6
- Tinbergen N (1953) *Social behavior in animals*. Methuen, London
- Tomasello M, Carpenter M (2005) The emergence of social cognition in three young chimpanzees. *Monogr Soc Res Child Dev* 70: Serial No. 279
- Warneken F, Tomasello M (2006) Altruistic helping in human infants and young chimpanzees. *Science* 313:1301–1303. doi:10.1126/science.1121448
- Warneken F, Hare B, Melis A, Hanus D, Tomasello M (2007) Spontaneous altruism by chimpanzees and young children. *PLoS Biol* 5:e184. doi:10.1371/journal.pbio.0050184
- Woodruff G, Premack D (1979) Intentional communication in the chimpanzee: the development of deception. *Cognition* 7:333–362. doi:10.1016/0010-0277(79)90021-0