Cognition 112 (2009) 435–450

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Can chimpanzees (*Pan troglodytes*) discriminate appearance from reality?

Carla Krachun [a,b,*], Josep Call [a,*], Michael Tomasello [a]

[a] *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*
[b] *Carleton University, Ottawa, Canada*

ARTICLE INFO

Article history:
Received 26 July 2008
Revised 17 June 2009
Accepted 19 June 2009

Keywords:
Appearance-reality
Metacognition
Theory of mind
Chimpanzees
Comparative cognition

ABSTRACT

A milestone in human development is coming to recognize that how something looks is not necessarily how it is. We tested appearance-reality understanding in chimpanzees (*Pan troglodytes*) with a task requiring them to choose between a small grape and a big grape. The apparent relative size of the grapes was reversed using magnifying and minimizing lenses so that the truly bigger grape appeared to be the smaller one. Our *Lens* test involved a basic component adapted from standard procedures for children, as well as several components designed to rule out alternative explanations. There were large individual differences in performance, with some chimpanzees' responses suggesting they appreciated the appearance-reality distinction. In contrast, all chimpanzees failed a *Reverse Contingency* control test, indicating that those who passed the Lens test did not do so by learning a simple reverse contingency rule. Four-year-old children given an adapted version of the Lens test failed it while 4.5-year-olds passed. Our study constitutes the first direct investigation of appearance-reality understanding in chimpanzees and the first cross-species comparison of this capacity.

1. Introduction

In the chimpanzee's world, as in our own, appearances can be deceiving: a distant pond vanishes as one gets closer; a snake turns out to be a harmless stick; a scurrying rodent is just a leaf blowing along the ground. Chimpanzees no doubt occasionally experience such mistakes of perception. Whether these experiences eventually lead to the recognition that things are not always as they appear is another question, one that has not yet been addressed experimentally. There is, however, a rich literature on human children's understanding of the appearance-reality distinction, thanks in large part to the work of Flavell and colleagues throughout the 1980s (e.g., Flavell, 1986; Flavell, Flavell, & Green, 1983; Flavell, Green, & Flavell, 1986; Flavell, Green, Wahl, & Flavell, 1987; Taylor & Flavell, 1984). These studies provide a good starting point in devising methods to investigate appearance-reality understanding in chimpanzees.

The tests most commonly used with children have involved presenting them with an object whose appearance is at odds with its true identity, function, or properties. In the well-known *Rock–sponge* test (Flavell, Green, & Flavell, 1986; Flavell et al., 1983), for example, children are shown a sponge that has been painted to realistically resemble a rock. To pass the test, children must correctly say both what the object looks like (rock) and what it really is (sponge). In another classic test, the experimenter moves an object behind a transparent, colored filter and questions children about the object's real and apparent color (Flavell et al., 1986, 1987). Children have also been asked to comment on the real and apparent size or shape of objects whose appearance has been altered by various means, such as by placing them in a glass of water or behind magnifying

* Corresponding authors. Address: Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S 5B6 (C. Krachun), Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany (J. Call). Tel.: +49 (0)341 3550 418; fax: +49 (0)341 3550 444 (J. Call).

E-mail addresses: ckrachun@gmail.com (C. Krachun), call@eva.mpg.de (J. Call).

0010-0277/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2009.06.012

or minimizing lenses (Braine & Shanks, 1965a, 1965b; Flavell et al., 1983; Taylor & Flavell, 1984). Children typically begin to pass these standard verbal tests between about 4 and 5 years of age. Flavell and others (e.g., Gopnik & Astington, 1988; Taylor & Hort, 1990) have argued that younger children's difficulties with appearance-reality tests reflect a deep-seated representational inflexibility that makes them incapable of holding two conflicting representations of the same object in mind at the same time. However, 3-year-olds have succeeded in some appearance-reality tests when more natural questioning and nonverbal, helping responses were used (Sapp, Lee, & Muir, 2000), and also when information-processing requirements were reduced or deceptive elements were added (Rice, Koinis, Sullivan, Tager-Flusberg, & Winner, 1997). And even when 3-year-olds have performed poorly overall in a test, they have sometimes done well in some conditions (see, for example, Study 1 in both Flavell et al., 1983, 1987).

Some researchers have noted that results from appearance-reality tests are often markedly similar to those in research on false belief understanding (e.g., Flavell, 1993; Gopnik & Astington, 1988). In false belief tests, participants must recognize that others will be mistaken about the location of an object moved in their absence (Wimmer & Perner, 1983) or about the contents of a box with misleading packaging (Perner, Leekam, & Wimmer, 1987). As with appearance-reality tests, performance in standard verbal false belief tests shows a developmental progression from 3 to 5 years of age (Flavell, 1993; Wellman, Cross, & Watson, 2001), and younger children tend to perform better in tests involving nonverbal, helping responses (Carpenter, Call, & Tomasello, 2002; Matsui & Miura, 2008) or deceptive elements (Hala, Chandler, & Fritz, 1991; Sullivan & Winner, 1993). For both appearance-reality tests (Deák, 2006; Hansen & Markman, 2005) and false belief tests (Lewis & Osborne, 1990), there is some evidence that young children's difficulties in standard verbal procedures may be due to misinterpreting the test questions. Further, significant positive correlations in performance on appearance-reality and false belief tests have often been observed (e.g., Andrews, Halford, Bunch, Bowden, & Jones, 2003; Frye, Zelazo, & Palfai, 1995; Gopnik & Astington, 1988; Moore, Pure, & Furrow, 1990). The common assumption is that these parallels reflect a reliance, in both tests, on the same underlying cognitive capacity to hold two alternative representations in mind simultaneously. In false belief tests, these are another's mistaken belief versus reality; in appearance-reality tests, they are one's own mistaken perception versus reality.

In the current study, we sought to determine whether our closest evolutionary relatives might also have the capacity to distinguish appearance from reality. Chimpanzees have so far performed poorly in false belief tests (Call & Tomasello, 1999; Kaminski, Call, & Tomasello, 2008; Krachun, Carpenter, Call, & Tomasello, 2009), but it does not necessarily follow that they should perform poorly in appearance-reality tests, for several reasons. First, there are potentially important differences between the two types of test, including the mental states at issue. Gopnik and Astington (1988) speculate that it might be easier to "construct metarepresentations of representations that are more perceptually based than those that are more abstract" (p. 34), and so representing a mistaken perception may be easier than representing a false belief. Perhaps more crucially, false belief tests focus on others' mental contents whereas appearance-reality tests focus on one's own. Chimpanzees could possess a narrower capacity for dual representation than humans – one that is limited to their own sphere of mental experience. This is not an unreasonable prospect, given that one's own mental states can be directly experienced while others' must be indirectly inferred. Intuitively, it would seem easier to recognize when one has a perception in conflict with one's own knowledge of reality than to recognize when someone else has a belief in conflict with reality. Research with apes (Call, 2005; Call & Carpenter, 2001) and rhesus monkeys (e.g., Hampton, 2001; Kornell, Son, & Terrace, 2007; Washburn, Smith, & Shields, 2006) does indeed suggest that they may possess some ability to reflect upon their own mental states, although Carruthers (2008), among others, has challenged the metacognitive interpretation of these findings. Carruthers does grant, however, that more convincing evidence of metacognitive processes might be found by pitting "behavior guided by a current belief against behavior that is guided by a belief about current perception" (p. 86), which is the approach we take in the current study.

False belief tests differ from appearance-reality tests in another way that may give children an unfair advantage over apes. False belief tests typically entail a fairly complex sequence of events played out over time by two or more actors, giving the tests a narrative structure that is not a feature of most appearance-reality tests. With their bedtime stories, cartoons, and so on, even toddlers have amassed far more experience processing such narrative structures than apes. Although chimpanzees do poorly in false belief tests, they might therefore perform better in more simply structured appearance-reality tests.

A final reason to suspect that chimpanzees might succeed in appearance-reality tests is that they have performed well in other paradigms that draw on related cognitive skills. For example, they have demonstrated good visual perspective taking abilities, both with regard to what others can or cannot see (Bräuer, Call, & Tomasello, 2007; Hare, Call, Agnetta, & Tomasello, 2000) and what they themselves would or would not be able to see from different perspectives (Krachun & Call, 2009). In the children's literature, visual perspective taking skills have correlated positively with performance on appearance-reality tasks (Flavell, 1986; Flavell et al., 1986). Further, some investigators (e.g., Braine & Shanks, 1965a, 1965b; Murray, 1968) have theorized that grasping the appearance-reality distinction is foundational to succeeding in Piagetian conservation tasks. Liquid-conservation tasks, for example, test participants' appreciation of the fact that the quantity of a liquid is unchanged when it is poured from one container into another with different proportions. Apes have enjoyed some success in conservation tasks (Call & Rochat, 1996, 1997; Muncer, 1983; Suda & Call, 2004; Woodruff, Premack, & Kennel, 1978), although it is not clear to what extent they used strategies that did not rely upon appearance-reality understanding, such as

tracking the larger amount of liquid through the transfer or estimating quantities by visual inspection after the transfer. Thus, investigations aimed more directly at testing apes' appearance-reality understanding are needed.

To our knowledge, the current study is the first such direct investigation of apes' understanding of the appearance-reality distinction. As in some previous studies with children (Flavell et al., 1983, 1986, 1987; Taylor & Flavell, 1984), our appearance-reality test for chimpanzees involved presenting them with a situation in which an object's real and apparent properties were in conflict with one another. We used magnifying and minimizing lenses to make a small grape appear bigger and a big grape appear smaller. Chimpanzees who recognized that the apparently smaller grape was the truly bigger (and thus more desirable) one should choose that grape, thereby demonstrating that they distinguished appearance from reality. Our Lens test for chimpanzees was comprised of three main components or subtests administered in a fixed order. First, we gave chimpanzees a *Basic* test adapted from procedures used with children. In this test, chimpanzees watched as an experimenter placed a small grape behind a magnifying lens and a big grape behind a minimizing lens. Chimpanzees who could distinguish appearance from reality would be expected to choose the grape that looked smaller but was actually bigger. Second, we administered a *Tracking* test to address the possibility that individuals who passed the Basic test might have done so by visually tracking the bigger grape as it was placed behind the minimizing lens. The Tracking test involved a procedural modification that made it impossible for chimpanzees to solve the task by this method. Briefly, the experimenter again allowed chimpanzees to watch as she placed the small and big grapes behind the lenses, as in the Basic test, but she then blocked their view so that they could not see where she positioned the containers (*Unseen* trials). For comparison, in half the trials (*Seen* trials) chimpanzees were also allowed to watch the placement of the containers. Third, we administered a *Transfer* test to see if chimpanzees who passed the Basic and Tracking tests could transfer their skills over to a different type of food: apple cubes. If so, this would suggest that they understood that the transforming effects of the lenses were not specific to grapes.

We anticipated the argument that chimpanzees could pass our Lens test by learning a simple reverse contingency rule (choose a smaller grape to get a bigger one), without truly understanding that the appearance of the grapes was misleading. Chimpanzees typically experience serious difficulties with reverse contingency tasks, showing a strong bias toward choosing larger quantities or sizes over smaller ones (Boysen & Berntson, 1995; Boysen, Berntson, Hannan, & Cacioppo, 1996; Boysen, Berntson, & Mukobi, 2001). Individuals who successfully overcome this bias do so only after hundreds of trials (Uher & Call, 2008; Vlamings, Uher, & Call, 2006). Nevertheless, to rule out the possibility that chimpanzees were solving our Lens test by learning a reverse contingency rule, rather than by recognizing that the lenses were transforming the appearance of the grapes, we also ran two Reverse Contingency control tests: *RC-1* and *RC-2* (the reason for including two different tests is explained in the Procedure section). In both of these tests, chimpanzees were required to choose a smaller grape to obtain a bigger one, and vice versa, but there were no misleading appearances involved. If chimpanzees could choose the apparently smaller (truly bigger) grape in the Lens test but could not choose a smaller grape to obtain a bigger one in the Reverse Contingency tests, this would indicate that they were *not* using a simple reverse contingency rule to solve the Lens test.

For comparison and validation purposes, we also tested 4- and 4.5-year-old children with a child-adapted version of the Lens test and with several commonly used appearance-reality tests. We focused on these ages because the transition from failing to passing appearance-reality tests usually occurs during this time (Flavell et al., 1983, 1986; Gopnik & Astington, 1988). The chimpanzees' test is described first below, followed by the test for children.

2. Study 1: Chimpanzees

2.1. Methods

2.1.1. Participants

Fourteen chimpanzees (*Pan troglodytes*) housed socially at the Wolfgang Köhler Primate Research Center in Leipzig, Germany participated (see Table 1). All were born in captivity and they ranged in age from 6 years, 0 months to 31 years, 9 months old at the start of testing (mean = 19 years, 6 months).

Three further chimpanzees were dropped from the study, one for being uncooperative, one because of procedural error, and one for failing the pretest criterion (i.e., the *Preference* test; see below). Participants were not food deprived and they had water available *ad libitum*. Eight chimpanzees had previously taken part in studies employing reverse contingency procedures.

2.1.2. Design

All chimpanzees first received a Preference test to confirm that they preferred bigger grapes to smaller ones. Chimpanzees who did so were then given a two-trial demonstration (*Demo*) of the effects of the magnifying and

Table 1
Chimpanzee participants.

Chimpanzee	Sex	Age (years, months)	Rearing	RC history
Corrie	Female	30, 9	Nursery	None
Dorien	Female	26, 7	Nursery	1, 2
Fraukje	Female	31, 1	Nursery	2
Natascha	Female	27, 5	Nursery	None
Ulla	Female	30, 3	Nursery	None
Patrick	Male	10, 3	Mother	2
Robert	Male	31, 9	Nursery	None
Trudy	Female	13, 11	Mother	2
Frodo	Male	13, 6	Mother	1
Lome	Male	6, 0	Mother	None
Jahaga	Female	14, 3	Mother	1
Sandra	Female	13, 11	Mother	1, 2
Pia	Female	8, 0	Mother	None
Fifi	Female	13, 11	Mother	2

RC history = previous participation in reverse contingency studies: 1 = Vlamings, Uher, and Call (2006); 2 = Uher and Call (2008).

minimizing lenses. Immediately following the Demo were the three components of the Lens test: the (1) Basic, (2) Tracking, and (3) Transfer tests. In the Basic test, all trials were Seen; in the Tracking and Transfer tests half the trials in each session were Seen and half were Unseen, with order determined randomly. Chimpanzees' performance in earlier stages of testing served as a criterion for inclusion in later stages: only those individuals who passed the Basic test were included in the Tracking test, and only those who passed the Tracking test were included in the Transfer test. Passing was defined as choosing the truly bigger food significantly more often than chance ($p < 0.05$) in either the second session or across both sessions combined. In the Tracking and Transfer tests, chimpanzees had to additionally exceed chance performance in the Unseen trials alone in order to pass. In addition to the test trials, all chimpanzees received the first Reverse Contingency control test (RC-1), consisting of four 12-trial sessions. Half the chimpanzees were given this test before the Lens test and half afterwards. Chimpanzees who exceeded chance in the Basic test were also given the second Reverse Contingency control test (RC-2), consisting of two 12-trial sessions, after all other testing was complete.

For all sessions, the bigger and smaller rewards were on the left and right an equal number of times and each container was baited first an equal number of times. The position of the rewards, and also which container was baited first, were randomly determined with the constraint that neither factor could repeat for more than three consecutive trials.

2.1.3. Materials

2.1.3.1. Lens test. For the Lens test, we used two identical plastic containers (16 × 16 × 8 cm), one with a magnifying lens on front and one with a minimizing lens on front (Fig. 1). The lenses were mounted into a hole (6.5-cm diameter) in a removable panel on the front of each container. Because the panels were angled 40° away from participants and the food was tucked in just behind the lenses, the food could not be seen from above (verified independently by three people). Each container had a narrow shelf on top on which the food could be displayed. Small grapes (2.0 × 1.5 cm) and big grapes (2.75 × 1.75 cm) were used in the Basic test and Tracking test (sizes refer to approximate diameters along the longer and shorter dimensions, respectively). A thin slice was cut from the bottom of each grape to keep it from rolling. For the Transfer test, small (1.25 cm^3) and big (1.75 cm^3) apple cubes were used. To maximize the visibility of the food while minimizing glare, we shone a light down into the containers from above.

2.1.3.2. Reverse Contingency (RC) control tests. Containers for the Reverse Contingency tests also required chimpanzees to view the food through a transparent barrier (equivalent to the lenses in the Lens test), but the appearance of the food was not modified. For all trials, we used small and big grapes that approximated the apparent sizes of the grapes as viewed through the lenses in the Lens test. For the RC-1 test, the containers were two identical opaque rectangular tins (16 × 11 × 7 cm) open at the back end (Fig. 2a). A piece of rectangular clear plastic (10 × 8 cm) with a narrow shelf on top was mounted to the front of each container. Chimpanzees could see grapes that were placed onto the surface of the container just behind the plastic but they could not see grapes that were placed inside the container. For the RC-2 test, the containers were two semicircular columns of clear plastic (10 cm wide, 14 cm high), each mounted onto a 10 × 10 cm base (Fig. 2b). The top half of each container was covered in black plastic and contained a hidden shelf onto which a grape could be placed. Chimpanzees could thus see grapes placed in the bottom half but not in the top half of the container.

2.1.4. Pretest procedure

Chimpanzees were tested individually in a familiar enclosure. The experimenter (E) sat facing the chimpanzee through a Plexiglas window (69 × 49 cm) with a sliding table mounted below it, approximately 50 cm above floor level. When E slid the table forward, chimpanzees could choose a container by poking a finger through one of two holes spaced 58 cm apart on the bottom left and right sides of the window.

2.1.4.1. Preference test. Before taking part in the Lens test, chimpanzees had to demonstrate that they preferred big grapes to small ones under normal conditions. The procedure for the Preference test was as follows: E put the containers onto the table, one on the left and one on the right, and simultaneously placed a big grape on top of one and a small grape on top of the other. She then slid the containers within reach of chimpanzees so they could choose (note that the grapes were not lowered behind the lenses in these trials). Chimpanzees who chose the bigger grape

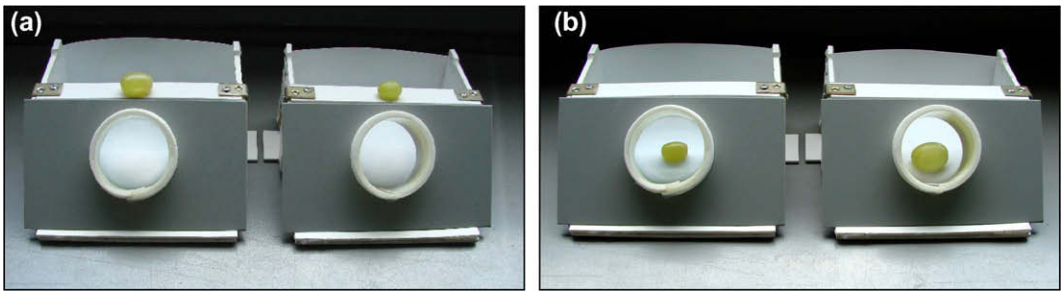

Fig. 1. Containers used in the Lens test, showing the (a) real and (b) apparent sizes of the grapes. The magnifying lens is on the right in both photographs.

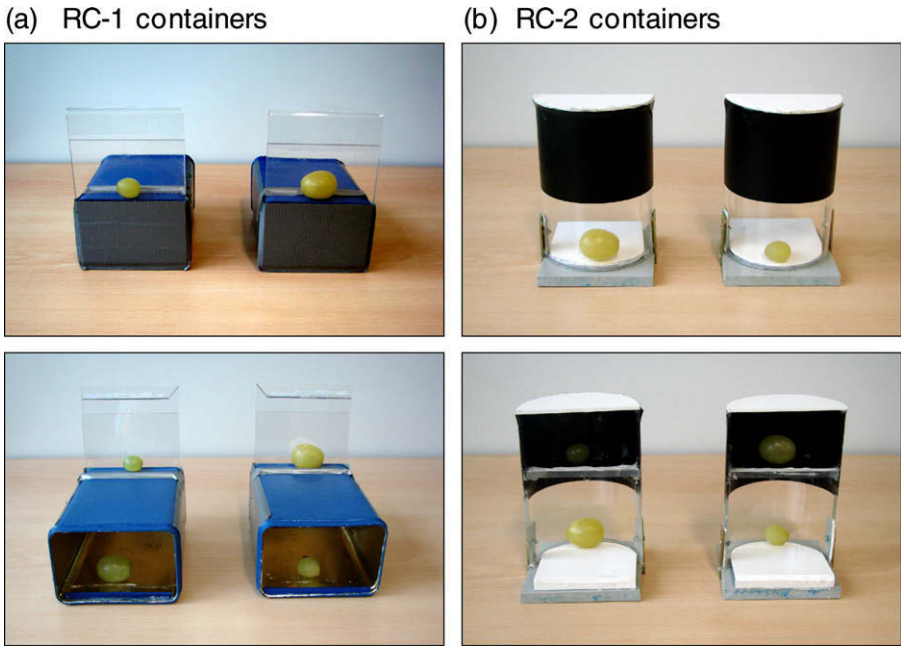

Fig. 2. Containers used in the (a) RC-1 and (b) RC-2 Reverse Contingency control tests, as viewed from the ape's perspective (top row) and the experimenter's perspective (bottom row).

significantly more often than 50% chance ($p < 0.05$) in the first 12-trial session moved directly on to the Demo; chimpanzees who did not exceed chance were given a second Preference session. For the remainder of the study, two Preference trials were repeated at the beginning of each Lens test session. If chimpanzees did not choose the bigger grape in both trials they were not tested that day (this happened only twice).

2.1.4.2. Demo. Our chimpanzees had no prior experience with magnifying or minimizing lenses. Before administering the Lens test we therefore gave them two Demo trials, with two objectives in mind. First, we wanted to verify that chimpanzees were in fact subject to the visual illusions created by the lenses. If so, they should choose the magnified grape in the first Demo trial, before they had any reason to suspect that the appearance of the grapes was misleading. Second, we needed to familiarize chimpanzees with the effects of the lenses before testing. The procedure was as follows: E put an occluder in front of the window to block the chimpanzee's view. She placed a small grape behind the magnifying lens and a big grape behind the minimizing lens. When viewed through the lenses, the apparent relative size of the grapes was thus the reverse of reality. Note also that because of their spheroid shape, the grapes scaled upward and downward in appearance with minimal distortion. When E was sure the chimpanzee was attending, she removed the occluder, waited for the chimpanzee to look at both containers, and then slid them forward so the chimpanzee could choose. Before handing over the chosen grape, however, E slowly removed it from behind the lens and then replaced it again, repeating this action until the chimpanzee had witnessed its visual transformation at least twice. E next performed the demonstration with the other grape behind the other lens. Two such Demo trials were repeated at the beginning of every Lens test session immediately before the test trials, in order to refresh chimpanzees' memory of the effects of the lenses.

2.1.4.3. Additional pretest experience with different foods. We were curious to know if more varied experience with the lenses, beyond that chimpanzees received in the Demo trials, would make a difference to their performance in later test trials. Thus, before testing, we gave six chimpanzees two additional pretest sessions (12 trials each) in which they experienced the effect of the lenses on a variety of foods including pretzels, peanuts, tea biscuits, and pieces of banana, kiwi, and orange.[1] Chimpanzees experienced the effects of the lenses on at least four of these foods depending on individual preferences, and the foods were presented in random order. Grapes were still used for the two Preference and Demo trials at the beginning of these sessions. The procedure for these additional pretest sessions was the same as for the Basic test described below.

2.1.5. Lens test procedure

The three components of our appearance-reality Lens test are depicted in Fig. 3 and are also described in detail below, in the order in which they were administered.

2.1.5.1. Basic test. This first component of the Lens test was our simplest test of chimpanzees' ability to distinguish appearance from reality. E placed a small grape atop the

[1] Two chimpanzees who were ultimately dropped from the study also received these sessions. Also note that the eight chimpanzees who did not participate in these sessions proceeded directly from the Demo to the Lens test.

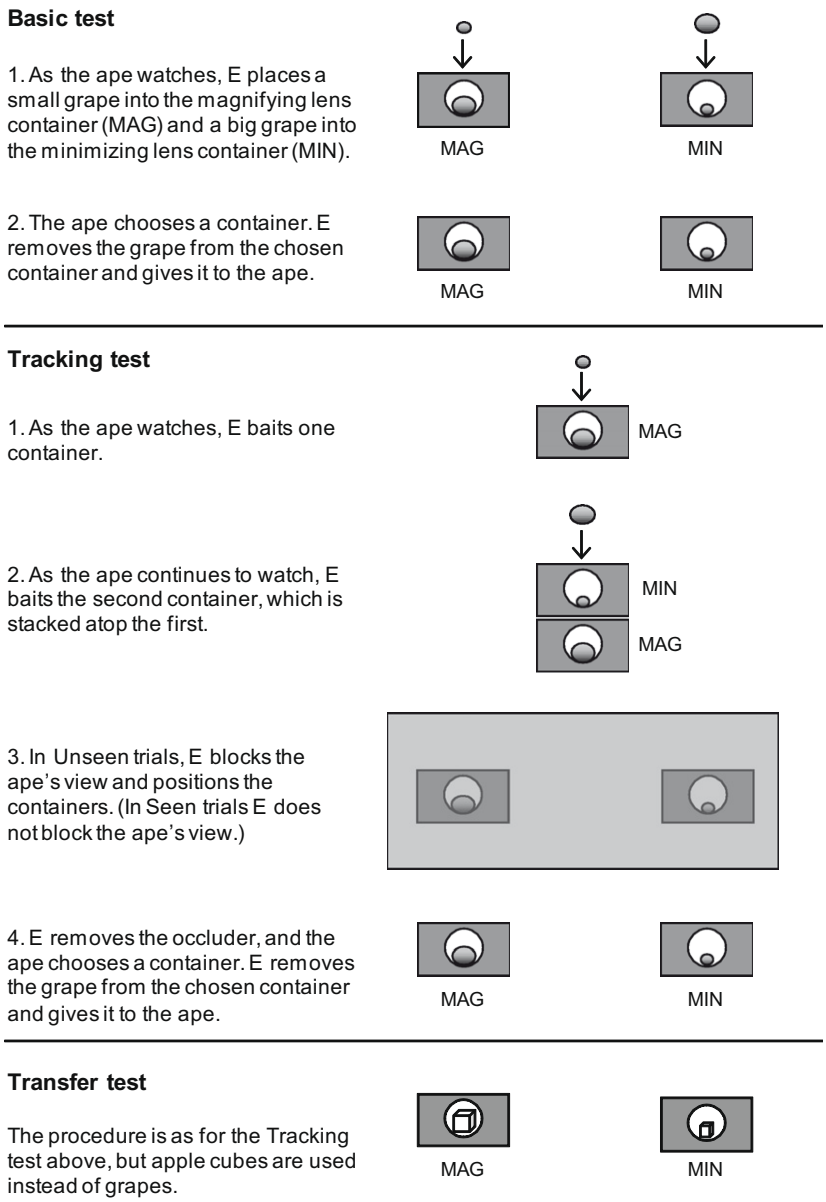

Fig. 3. The various components of the Lens test for chimpanzees. MAG = magnifying lens; MIN = minimizing lens.

magnifying lens container and a big grape atop the minimizing lens container. As chimpanzees watched, she slowly lowered each grape in turn into the container and behind the lens. When E was confident the chimpanzee had looked at both containers (either during baiting or afterwards), she allowed the chimpanzee to choose. E then removed both grapes and rewarded the chimpanzee with the chosen grape. To ensure that chimpanzees would not just choose the first or last container E touched, she simultaneously touched both containers before sliding them forward. Also, so that chimpanzees could not learn to avoid the magnifying lens by noticing minute differences in the appearance of the containers, halfway through each session E surreptitiously switched the front panels of the containers, along with the lenses mounted in them. As a further precaution, before each session E also removed each lens and remounted it into the other panel.

Chimpanzees who exceeded chance in the Basic test proceeded directly to the Tracking test. Those who failed the Basic test, however, may have done so because they were unable to inhibit reaching for the grape that appeared bigger. We therefore gave these individuals a *Reduced Inhibition* test, consisting of two sessions the same as before except that E covered the lenses just before letting chimpanzees choose, so that they could not see the grapes at the moment of choice. Thus, if being confronted by the image of the grapes at the moment of choice was causing inhibition problems for these chimpanzees in the Basic

test, they would be expected to perform better in the Reduced Inhibition test. Any chimpanzee who exceeded chance in this test would then be allowed to proceed to the Tracking test. This never occurred, however, because all six chimpanzees who failed the Basic test also failed the Reduced Inhibition test.

2.1.5.2. Tracking test. We recognized that chimpanzees who did well in the Basic test could have done so by visually tracking the bigger grape. Of course, to do so would necessarily involve ignoring the misleading appearance of the grapes behind the lenses, suggesting some understanding that the grapes' appearance conflicted with reality. However, stronger evidence that chimpanzees recognized that the apparently bigger grape was in reality the smaller one would be if they could succeed without the opportunity to visually track the grapes. There were two conditions in the Tracking test: Seen and Unseen. Seen trials were similar to the Basic test in that chimpanzees could still visually track the movement of the bigger grape, although they now had to pay closer attention in order to successfully do so. In Unseen trials, however, we made it impossible for chimpanzees to track the movement of the grapes. In both conditions, E began by placing one container in the center of the table and lowering a grape into it as the chimpanzee observed. E then stacked the second container on top of the first and lowered the second grape into it. Again, the small grape always went behind the magnifying lens and the big grape behind the minimizing lens. In Seen trials, E positioned the containers on the left and right sides of the table in full view of chimpanzees and then let them choose. In Unseen trials, however, before positioning the containers E blocked chimpanzees' view with an occluder so that they could not visually track the movement of the grapes. Chimpanzees who exceeded chance performance in the Tracking test proceeded to the Transfer test.

2.1.5.3. Transfer test. We questioned whether individuals who performed well in previous sessions would transfer their skills over to a different type of food. If they could not do so, it would suggest that they believed the transforming effects of the lenses to be specific to grapes. In contrast, succeeding in the Transfer test would suggest that chimpanzees had come to understand something about the lenses more generally – that they caused whatever was placed behind them to have a misleading appearance. Thus, we administered two sessions identical to the Tracking test described above except that small and big apple cubes were used instead of grapes, for the Preference and Demo trials as well as the test trials.

2.1.5.4. Big-to-bigger control trials. At this point, we considered that individuals who succeeded in the Lens test could have done so because they learned to distinguish the magnifying and minimizing lenses from one another, despite our efforts to make them identical. They could have then avoided the magnifying lens because it had become associated with negative outcomes. Alternatively, they could have succeeded in the Lens test by avoiding the magnified food because it looked strange to them (e.g., grapes are not usually that big), without truly understanding that its appearance was misleading. To test these possibilities, after the Transfer test we gave our first three participants (Fifi, Jahaga, and Trudy) one session of six Seen and six Unseen trials in which we put the *bigger* piece of food behind the magnifying lens and the smaller piece behind the minimizing lens. The result was that the truly bigger food now looked even bigger and the truly smaller food now looked even smaller. The procedure was otherwise exactly the same as the previous Transfer test procedure. In every single trial, all three chimpanzees chose the food that was now both apparently and truly bigger than the other piece. They were clearly not avoiding the magnifying lens or the magnified food. Because the results were unequivocal, and because we were concerned that this session might hinder chimpanzees' performance in the Reverse Contingency trials that followed by getting them into the habit of choosing the bigger piece of food, we did not administer it to the remaining participants.

2.1.6. Reverse Contingency test procedure

If chimpanzees could pass the Lens test but not the Reverse Contingency tests, this would indicate that they were not solving the Lens test by learning a reverse contingency rule. The procedures for the two Reverse Contingency tests (RC-1 and RC-2) are depicted in Fig. 4 and described below.

2.1.6.1. RC-1 test. To minimize any possible transfer effects (positive or negative) that could occur if chimpanzees connected this test with the Lens test, we had a different experimenter administer these trials. E blocked the chimpanzee's view with an occluder and placed a big grape onto the shelf of one container and a small grape onto the other. E also placed a grape of the opposite size inside of each container. Thus, the container with the big grape displayed on its shelf had a small grape hidden inside of it, and vice versa. E then removed the occluder and slowly lowered each grape in turn down from the shelf and onto the surface of the container, just behind the transparent barrier. This action was analogous to lowering the grapes behind the lenses in the Lens test, but the appearance of the grapes was not transformed. After sliding the containers forward so the chimpanzee could choose, E removed the grape from behind the transparent barrier of the chosen container and dropped it into a nearby bucket. She then removed the hidden grape from inside the container and gave it to the chimpanzee. Removing the chosen grape from view just before revealing the other grape was analogous to lifting the chosen grape out from behind the lens in the Lens test.

After running RC-1, we realized that this test differed from the Lens test in one potentially important respect. In RC-1, at the beginning of each trial, E blocked chimpanzees' view before placing one grape on top of each container and one (of opposite size) inside each container. Thus, chimpanzees did not have the opportunity to see the big grape that they would receive if they picked the small one, and vice versa. In the Lens test, however, chimpanzees watched as E lowered the grapes into the containers, so they were able to see what their reward would be if they were to choose a given container. This difference

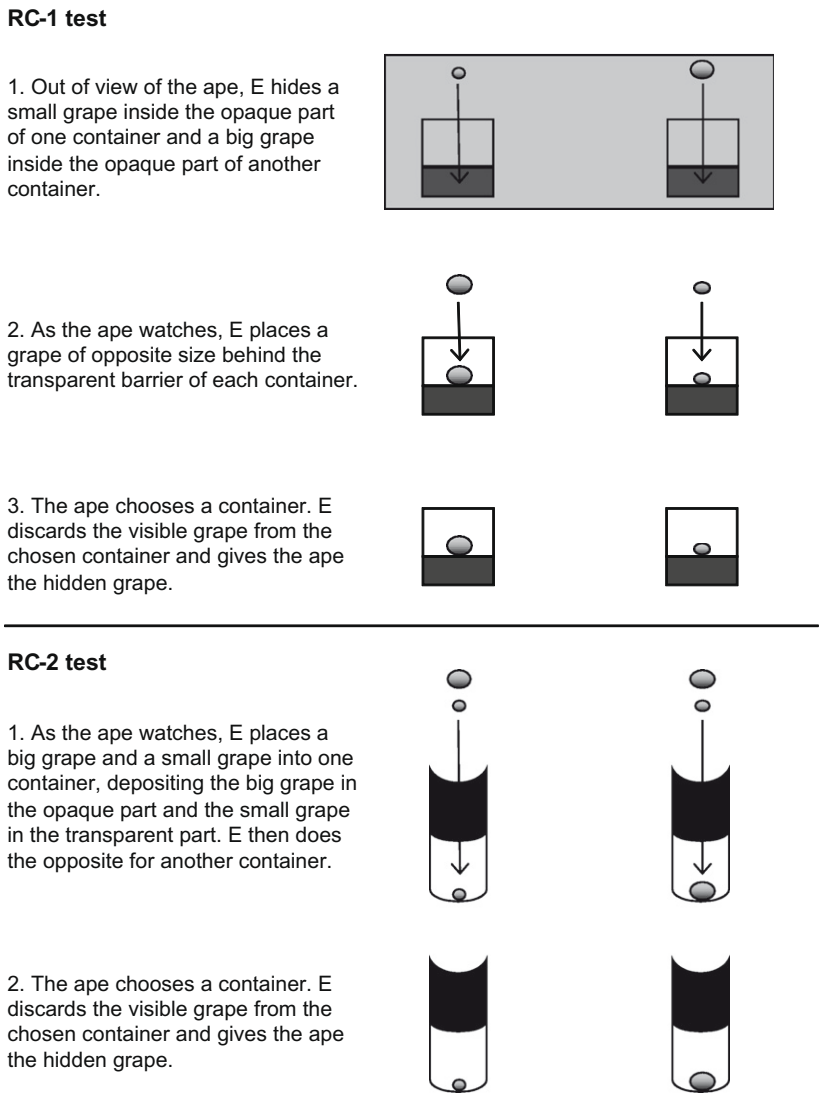

Fig. 4. Procedure for the two Reverse Contingency control tests: RC-1 and RC-2.

could have made the RC-1 test harder for chimpanzees than the Lens test. To address this potential problem, we created a modified Reverse Contingency test (the RC-2 test), described below.

2.1.6.2. RC-2 test. A third experimenter administered this test, again to minimize possible transfer effects. To make RC-2 more similar in procedure to the Lens test than RC-1, E did not block chimpanzees' view of the containers while placing the hidden grapes inside. Instead, E held up both a small and big grape together in one hand for the chimpanzee to see and then moved the grapes downward, depositing one into the opaque top part of the container and the other into the transparent lower part. E then repeated this with the other container, reversing the locations of the small and big grapes. Thus, analogous to the Lens test, chimpanzees saw a small grape disappear into one container while a big grape became visible and a big grape disappear into the other container while a small grape became visible.

2.1.7. Reliability and analyses

An independent coder naïve to the hypotheses of the study and blind to condition coded chimpanzees' choice in 10% of trials. Sessions were chosen randomly, with the constraint that each chimpanzee and each type of test, including Reverse Contingency tests, be represented at least once. Reliability was excellent (Cohen's $\kappa = 0.98$, $p < 0.001$). All statistical analyses were nonparametric and all p-values reported below are exact and two-tailed.

2.2. Results and discussion

2.2.1. Preference test and Demo

Chimpanzees unequivocally demonstrated that they preferred big grapes to small ones. Thirteen out of 14 chim-

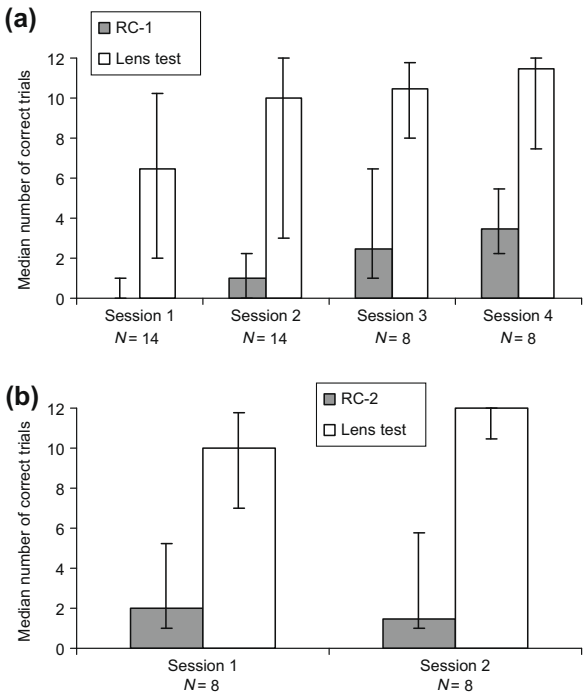

Fig. 5. Chimpanzees' median number of correct trials (± inter-quartile ranges) in the two Reverse Contingency tests. (a) RC-1 and (b) RC-2, and in the corresponding sessions of the Lens test. Note that the RC-2 test was administered only to the eight chimpanzees who passed the Basic test.

panzees chose the bigger grape in at least 10/12 trials of the Preference test (Binomial test: $p < 0.02$). The remaining chimpanzee scored 9/12 and was thus given a second session, in which she scored 12/12. Chimpanzees clearly also perceived the visual illusion created by the lenses. In the very first Demo trial, when they still had no experience whatsoever with the lenses, 13 chimpanzees were given the opportunity to indicate which grape they wanted.[2] Twelve of these 13 chimpanzees wrongly chose the magnified grape, and 11 also choose the magnified grape in their second Demo trial.

2.2.2. Lens test

The data were bi-modally distributed, with most chimpanzees doing either very well or very poorly in our test. We therefore examined medians rather than group means. We first compared chimpanzees' performance in the two Reverse Contingency control tests with their performance in the corresponding sessions of the Lens test, including all Seen and Unseen trials (Fig. 5). The median number of trials correct was significantly lower in all four sessions of RC-1 than in the first four sessions of the Lens test (Wilcoxon tests: Sessions 1 and 2: $T^+ \geqslant 77.00$, $N = 12$[2 ties], $p \leqslant 0.005$; Sessions 3 and 4: $T^+ \geqslant 35.00$, $N = 8$[0 ties], $p \leqslant 0.016$). Likewise, it was significantly lower in the two sessions of RC-2 than in the first two sessions of the Lens test ($T^+ = 36.00$, $N = 8$[0 ties], $p = 0.008$ in both sessions). The fact that chimpanzees did so poorly in the Reverse Contingency tests, both in absolute terms and relative to the Lens test, indicates that they were not using a simple reverse contingency rule to solve the Lens test.

As a group, chimpanzees did not perform significantly better than chance in the Basic test, whether we collapsed across Sessions 1 and 2 ($T^+ = 50.00$, $N = 12$[2 ties], $p = 0.42$) or analyzed Session 2 alone ($T^+ = 53.00$, $N = 12$[2 ties], $p = 0.26$). However, because of the bi-modal distribution of the data, and because chimpanzees who did poorly in one test were eliminated from subsequent tests, it made more sense to examine individual performance. We did so using Binomial tests to determine how many individuals passed or failed the various test components (see Table 2).

Eight of the 14 chimpanzees tested (57%) passed the Basic test, and five of them even exceeded chance within the first session. Further, in the first six trials of Session 1, five chimpanzees got at least 5/6 trials correct. Thus, in the Basic test, more than half the chimpanzees were capable of ignoring the misleading appearance of the magnified grape to choose the truly bigger one, and about a third were capable of doing so from the earliest trials. Of course, the fact that chimpanzees watched the baiting introduces the possibility that they succeeded by visually tracking the bigger grape, without attending to the appearance of the grapes behind the lenses. The eight chimpanzees who passed the Basic test were therefore given a Tracking test, with the Unseen trials being the true test of chimpanzees' ability to succeed without visually tracking the food. Chimpanzees were considered to have passed the Tracking test (and also the subsequent Transfer test) only if they exceeded chance performance in the Unseen trials separately, in addition to exceeding chance in the Seen trials. Five of the eight chimpanzees who received the Tracking test passed, and three of them got all six Unseen trials correct in the first session, thereby demonstrating that they could choose the truly bigger grape (despite its misleading appearance) without tracking. Four of these chimpanzees also passed the subsequent Transfer test with apples, indicating that they did not consider the effect of the lenses to be specific to grapes. The fifth chimpanzee was marginally significant with a total of 9/12 Unseen Transfer trials correct across Sessions 1 and 2. While no chimpanzee got all six Unseen trials correct in the first Transfer session, one chimpanzee got 5/6 correct and two chimpanzees got 4/6 correct. Thus, of the 14 chimpanzees tested, 4 individuals (29%) passed all three components of the Lens test. It should be noted that, in sharp contrast to the Lens test, not one chimpanzee exceeded chance in either of the Reverse Contingency control tests (RC-1 or RC-2), whether we analyzed all sessions combined within each test or just the last session (Binomial test: all $ps > 0.05$).

2.2.3. Other variables

Performance in all test components (Seen or Unseen trials) was not significantly affected by: (1) additional pretest experience with different foods, (2) previous reverse contingency experience, (3) order of administration of the Lens test and the RC-1 test, or (4) sex (Mann-Whitney U tests:

[2] One chimpanzee, Fifi, was inadvertently not given the opportunity to make a choice before the first Demo was performed. She was nevertheless retained for testing because this minor procedural error would not be expected to influence her performance in subsequent trials.

Table 2
Number of correct trials for individual chimpanzees in the various components of the Lens test.

Ape	Basic test		Tracking test				Transfer test			
	All trials Seen		Seen trials		Unseen trials		Seen trials		Unseen trials	
	Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 2	Sess 1	Sess 2
Corrie	2/12	4/12	–	–	–	–	–	–	–	–
Dorien	1/12	0/12	–	–	–	–	–	–	–	–
Fraukje	2/12	0/12	–	–	–	–	–	–	–	–
Natascha	6/12	6/12	–	–	–	–	–	–	–	–
Ulla	3/12	6/12	–	–	–	–	–	–	–	–
Patrick	1/12	0/12	–	–	–	–	–	–	–	–
Robert	7/12	10/12	2/6	0/6	1/6	1/6	–	–	–	–
Trudy	7/12	12/12	5/6	4/6	5/6	3/6	–	–	–	–
Frodo	12/12	12/12	3/6	5/6	5/6	4/6	–	–	–	–
Lome	10/12	12/12	6/6	6/6	2/6	6/6	6/6	6/6	4/6	5/6
Jahaga	11/12	12/12	6/6	6/6	6/6	6/6	3/6	6/6	1/6	6/6
Sandra	10/12	12/12	5/6	6/6	6/6	6/6	5/6	6/6	4/6	6/6
Pia	12/12	12/12	6/6	6/6	5/6	6/6	6/6	6/6	5/6	6/6
Fifi	2/12	10/12	6/6	5/6	6/6	6/6	1/6	6/6	1/6	6/6

Note: For Session 2 of each test, cells are highlighted in grey if the chimpanzee achieved significantly better than chance performance ($p < 0.05$), either in that session or in combination with Session 1. Individuals who did not exceed chance are excluded from subsequent tests.

collapsing across Sessions 1 and 2 for each test, all $ps > 0.07$). In contrast, the eight mother-reared chimpanzees did perform significantly better as a group than the six nursery-reared chimpanzees in the Basic test ($U = 7.00$, $p = 0.027$). However, the mother-reared chimpanzees were also significantly younger on average than the nursery-reared group (11 years, 8 months versus 29 years, 7 months; $U = 0.00$, $p < 0.001$). In fact, performance in the Basic test was inversely correlated with age in years (Pearson $r_{12} = -0.57$, $p = 0.034$). There was only one individual that clearly did not fit this pattern (Patrick was younger but scored just 1 of 24 trials correct). Re-running this analysis without this outlier confirmed and strengthened the inverse relation between age and performance ($r_{11} = -0.79$, $p = 0.001$). Other researchers have found inverse age effects for chimpanzees in other tasks, such as mirror self-recognition (Povinelli, Rulf, Landau, & Bierschwale, 1993) and episodic-like memory tasks (Martin-Ordas, Call, Haun, & Colmenares, submitted for publication), with adolescents and young adults performing better than older individuals. However, regressing both rearing and age against performance in the Basic test revealed that rearing explained a larger proportion of the variance than age, which did not enter the model after rearing was in it ($R^2 = 0.65$, $F_{1,11} = 20.39$, $p = 0.001$). This was somewhat surprising given that generally, when differences are detected, human-reared chimpanzees usually outperform mother-reared apes (see review by Call & Tomasello, 1996).

3. Study 2: Children

In order to compare chimpanzees' and children's performance on a similar appearance-reality task, we adapted our Lens test for children. While the children's test included essentially the same components as the chimpanzees', we made several changes that allowed us to administer the test in one session, to avoid repeatedly disturbing children in their daycare centers. First, the children's test consisted of just eight trials in total, and we did not administer the Basic test component. The Basic test was deemed unnecessary for children because the Seen tracking trials also allowed for visual tracking of the food, so they served the same function as the Basic trials.[3] Second, we did not administer the Reverse Contingency tests to children. While these were essential for the chimpanzees because of the large number of trials they received, it seemed highly unlikely that children would learn to use a reverse contingency rule to pass our Lens test within their relatively small number of trials.

As children could not be given food for safety reasons, we had them choose between big and small pieces of artificial food to give to a toy duck. For comparison with other commonly used procedures, we also gave children several additional appearance-reality tests similar to ones used by other researchers, including a *Standard-Format Lens* test in which children were simply asked which food looked bigger and which was really bigger. Children could answer by pointing in this test just as they could in our Lens test, and so both were nonverbal in that sense. However, the direct questioning about appearance and reality and the two-question structure made the Standard-Format Lens test more similar to other standard verbal tests such as the Rock–sponge test and tests using colored filters (see Flavell et al., 1983, 1986, 1987).

3.1. Methods

3.1.1. Participants

Thirty-two children from kindergartens in Leipzig, Germany were recruited for participation by letters sent to parents. There were 16 children in each of two age groups: 4-year-olds (range = 4 years, 0 months to 4 years, 4 months; mean = 4 years, 2 months) and 4.5-year-olds

[3] For the chimpanzees, we nevertheless included the Basic trials because we wanted to eliminate early on those individuals who could not pass even the simplest version of the test. Further, we were concerned that moving directly into the Tracking test might be too confusing for the chimpanzees, given that the experimenter could not even minimally explain the task for them as she could for the children.

(range = 4 years, 6 months to 4 years, 11 months; mean = 4 years, 8 months). The 4-year-olds included 8 males and 8 females; the 4.5-year-olds included 9 females and 7 males. Eight further children were dropped from the study, five because of uncooperativeness and three because of procedural error.

3.1.2. Design

All children received an introduction to the test followed by two Demo trials and then the Lens test. As with the chimpanzees, the children's test involved several components administered in a fixed order: (1) a Tracking test consisting of alternating Seen and Unseen trials, two of each, (2) a Transfer test consisting of two Unseen trials, and (3) a Post-test consisting of two Unseen trials similar to the initial Demo trials. Given the small number of trials administered, children participated in all components regardless of their performance. Within each test component, the truly bigger food was on the left and right an equal number of times and each container was baited first an equal number of times, with side and order determined randomly.

Upon completion of the Lens test, all children were also administered four additional appearance-reality tests (one trial of each) based on procedures developed by other researchers. They were first given a Standard-Format Lens test, in which they were directly asked which piece of food looked bigger and which was really bigger, as in standard verbal appearance-reality tests. They were then given a Rock–sponge test and two tests involving transparent colored filters: the *Nonverbal Color Filter* test and the *Verbal Color Filter* test (adapted from Flavell et al., 1983, 1986). The two Color Filter tests were always grouped together with the nonverbal version administered first. Half the children received the Rock–sponge test before the two Color Filter tests and half received it afterwards.

3.1.3. Materials

The magnifying and minimizing containers were similar to the chimpanzees' containers except that the back was left open to make it easier for the experimenter to reach inside. In place of grapes, we used small and big balls of green modeling clay flattened on the bottom to keep them from rolling. In place of apples, we used small and big cubes of white clay. The balls and cubes were approximately the same size as the grapes and apple cubes used with the chimpanzees. They served as artificial food for a small plush duck creatively named "Duck". Materials for the Rock–sponge test included a sponge painted to look like a rock. Materials for the two Color Filter tests included a transparent yellow filter (20×12 cm) mounted into a cardboard A-frame that could be made to stand upright at an angle of approximately 45-degrees, a hole puncher that made flower-shaped holes, and yellow and white pieces of paper.

3.1.4. Pretest procedure

Children were tested individually in a quiet room in their kindergarten by a female experimenter who sat opposite the child at a small table.

3.1.4.1. Introduction. The experimenter (E) first introduced the child to Duck and explained that Duck had a fat belly because she liked to eat a lot. E told children that when Duck got a lot to eat she became happy and quacked, but that when she did not get enough to eat she became sad and remained quiet. E then demonstrated to children that Duck preferred bigger pieces of food to smaller pieces. E put three small and three big green balls onto the table, explaining that they were Duck's food (we hereafter refer to them as food). E then put one big piece of food in front of the child, saying, "Let's see what happens when we give Duck this." Duck approached the food, 'ate' it, and then quacked and danced about happily. E explained that Duck was happy because she had gotten lots to eat. E then put a small piece of food in front of the child. Duck approached and ate it, then hung her head sadly as E said, "Duck is sad because she didn't get enough to eat." Finally, E put both a big and a small piece of food in front of the child and said "Okay, now Duck can choose between two pieces herself, and we'll see which piece she takes." Duck looked back and forth between the two pieces, then approached the bigger piece, ate it, and began quacking and dancing about. E explained again that Duck was so happy because she had gotten lots to eat.

3.1.4.2. Demo. As with the chimpanzees, E baited the containers out of view of children, putting a small piece of food behind the magnifying lens and a big one behind the minimizing lens. She then presented the containers and asked the child to indicate which piece of food they should give Duck so that she would be happy and quack. When the child pointed to one of the containers, E said, "Okay, but before we give Duck the food, let me show you something." E then slowly removed the food from the container and replaced it again until the child had seen its visual transformation at least twice. She then repeated this with the other container. Finally, E removed both pieces of food from the containers and gave Duck the one the child had indicated. Duck reacted appropriately, quacking and dancing when she got the truly bigger piece and hanging her head sadly when she got the smaller piece. E did two Demo trials, one with the magnifying container on the left and one with it on the right, with order counterbalanced across participants.

3.1.5. Lens test procedure

As with the chimpanzees, the Lens test was comprised of a number of components, described below in the order in which they were administered.

3.1.5.1. Tracking test. Children proceeded directly to the Tracking test from the Demo. As mentioned earlier, although children did not receive the Basic test that chimpanzees received, the Seen trials of the Tracking test performed the same function as the Basic trials. Thus, for the children's Tracking test, we always began with a Seen trial to preserve the same fundamental sequence of testing as for chimpanzees. The procedure was the same as for the chimpanzees, with E baiting the containers and stacking them one atop the other, then positioning them with children's view blocked in Unseen trials but not in Seen trials.

As before, E asked children which food they should give Duck so that she would be happy and quack.

3.1.5.2. Transfer test. The Transfer test was the same as the Tracking test except that white cubes were used instead of green balls, and only Unseen trials were run to keep the testing time within reasonable limits. Because the white cubes were very similar in color to the interior of the containers, E placed a black piece of paper on the floor of each container before administering this test. This was not necessary for the chimpanzees because the apple cubes were off-white and lit with supplementary lighting from above, so they were clearly visible through the lenses.

3.1.5.3. Post-test 'demo' trials. Once children had had more experience with the lenses, we were interested in knowing how they would perform when they could longer witness the baiting of the containers, as was the case in the initial Demo trials. These trials were thus equivalent to the two Demo trials repeated at the beginning of every test session for chimpanzees. The procedure for the two Post-test trials was the same as for the initial Demo trials except that E did not perform the demonstration but simply removed the food the child indicated and gave it to Duck.

3.1.6. Procedure for additional appearance-reality tests

3.1.6.1. Standard-Format Lens test. We were curious to know how children's performance in our Lens test might differ if we asked them directly about appearance and reality, as in standard verbal tests. Thus, after children indicated a container in the second Post-test trial E said "Okay, but before I give Duck the food, I would like to ask you something." Pointing simultaneously to the front of each container, E asked children two questions, in counterbalanced order: (1) When you look at the food through here, which piece looks bigger? and (2) Which piece is really bigger? Children had to answer both questions correctly to pass.

3.1.6.2. Rock–sponge test. In this standard verbal test, E first placed the Rock–sponge on the table and asked, "Can you tell me what this is?" When children identified the object as a rock/stone/etc. (which they did in every case), E allowed them to handle it and discover that it was a sponge. E then put the Rock–sponge back onto the table and asked the child what the object looked like and what it really was, counterbalanced for order. Children had to answer both questions correctly to pass. Occasionally a child was hesitant to answer, in which case E repeated the question(s) while stating the two possible options (rock or sponge), in varying order. Thus, we began with open-ended questions and only provided the two response options if necessary. (Sapp et al., 2000, used a similar open-ended question format in their verbal response paradigm.) We avoided providing the two response options whenever possible because of evidence that doing so may lead children to err by giving the same answer to both questions (see Deák, 2006).

3.1.6.3. Nonverbal Color Filter test. In this test, E stood the frame containing the yellow filter onto the table and made sure the child watched as she moved a piece of white paper behind the filter and out again at least twice. E then set the paper down behind the filter and used the hole puncher to make a flower-shaped hole in it. E next removed the paper flower in her closed hand and put two paper flowers, one white and one yellow, in front of the child. Pointing to the hole in the paper behind the filter, E asked "Can you show me which of these flowers I just took out of here?" Children were correct if they indicated the white flower.

3.1.6.4. Verbal Color Filter test. This test followed immediately upon the Nonverbal Color Filter test. With the paper still behind the filter, E asked children what color the paper looked like behind the filter and what color it really was, in counterbalanced order. As with the Rock–sponge test, if children were hesitant to respond E restated the question(s) including both possible answers, yellow or white. Children had to answer both questions correctly to pass.

3.1.7. Reliability and analyses

An independent coder naïve to hypotheses and blind to condition coded the choice responses of 25% of randomly chosen participants in each age group. There was 100% agreement with the main coder. A greater percentage of trials was coded for children than for chimpanzees because children received a much smaller number of trials overall. Unless otherwise noted, all statistical analyses were nonparametric and all reported p-values are exact and two-tailed.

3.2. Results and discussion

3.2.1. Demo

Like the chimpanzees, children clearly demonstrated that they perceived the magnified food as bigger: 27 out of 32 children wrongly chose the magnified food in their first Demo trial. Of the five children who chose correctly in the first Demo trial, four were incorrect in the second Demo trial.

3.2.2. Lens test

To be consistent with the chimpanzees' analysis, we examined medians rather than group means (Fig. 6). There were no significant differences across the various components of the Lens test for either age group (Friedman tests: 4-year-olds: $X^2 = 0.69$, $df = 3$, $N = 16$, $p = 0.88$; 4.5-year-olds: $X^2 = 2.46$, $df = 3$, $N = 16$, $p = 0.53$), so we analyzed all eight test trials combined. As a group, 4.5-year-olds succeeded in a significantly greater number of trials (median = 7) than 4-year-olds (median = 5) (Mann-Whitney U test: $U = 75.00$, $N_{\text{each group}} = 16$, $p = 0.042$). Additionally, 4.5-year-olds performed significantly better than 50% chance (Wilcoxon test: $T^+ = 109.50$, $N = 15$[1 tie], $p = 0.003$) but 4-year-olds did not ($T^+ = 66.50$, $N = 14$[2 ties], $p = 0.39$). This is in keeping with previous studies in which dramatic improvements in appearance-reality test performance have been observed between 4 and 5 years of age.

Because of the small number of trials children received in any given test component, it was not meaningful to look at individual performance within the separate components. However, taking all eight test trials into account, 4/16 children in the younger age group (25%) and 10/16

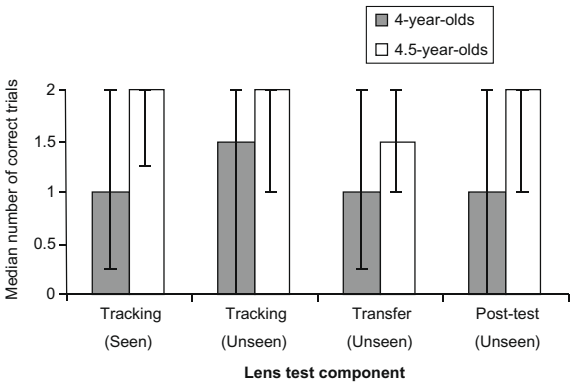

Fig. 6. Children's median number of correct trials (± inter-quartile ranges) in the various components of the Lens test ($N = 16$ for each bar).

children in the older age group (63%) exceeded chance performance by getting seven or more trials correct (Binomial test: $p \leqslant 0.035$).

3.2.3. Additional appearance-reality tests

Fig. 7 shows, for each of the additional appearance-reality tests we administered, the number of children in each age group who passed. Order of administration of the Rock–sponge test and the two Color Filter tests did not affect performance in either test (Fisher exact tests: $p \geqslant 0.08$ in all cases), and there were no effects of question order within each test (Fisher exact tests: $p \geqslant 0.12$ in all cases). We therefore collapsed across these variables.

Within the younger group, there were no significant differences across tests (Cochran's $Q = 3.34$, $df = 3$, $N = 16$, $p = 0.39$), and performance did not differ from chance in any of the tests (Binomial test: $p \geqslant 0.21$ in all cases). However, for the 4.5-year-olds, there were significant differences across tests ($Q = 11.65$, $df = 3$, $N = 16$, $p = 0.007$). This was due mainly to the older children's poor performance in the Standard-Format Lens test, which was significantly worse than in the Rock–sponge test (Sign test: $p = 0.008$) and marginally worse than in the Verbal Color Filter test (Sign test: $p = 0.063$). All other paired comparisons were nonsignificant at $p \geqslant 0.13$. Older children's performance did not exceed chance in the Standard-Format Lens test (Binomial test: $p = 0.80$), but it was better than chance in the Rock–sponge test (Binomial test: $p = 0.001$) and approached significance in both Color Filter tests (Binomial test: $p = 0.077$ in both cases). It is curious that the 4.5-year-olds did so poorly in the Standard-Format Lens test, given how well they did in the Lens test. We suspect they were confused over why the experimenter was suddenly questioning them about which piece of food looked bigger and which was truly bigger, when in most cases their responses up to that point clearly indicated that they knew this (see Siegal, 1997). Children's somewhat weaker performance in both Color Filter tests relative to the Rock–sponge test may also seem surprising, but it must be kept in mind that the sample size was not particularly large ($N = 16$). Furthermore, many 4.5- to 5-year-old children may be in a transitional phase with regard to appearance-reality understanding, and so varying task factors might exert a significant influence in this age range. Both the Rock–sponge test and the Color Filter tests were administered after the relatively lengthy Lens test, and children's motivation to attend may have been waning by that point. This could have had more of an effect on performance in the Color Filter tests, which were somewhat longer and more involved than the Rock–sponge test.

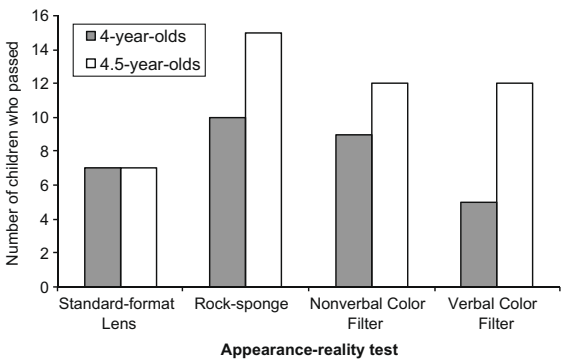

Fig. 7. Number of children who passed the various additional appearance-reality tests.

3.2.4. Correlations among tests

To examine relations among the various tests, we collapsed across groups and controlled for age. To make the Lens test consistent with other tests, we converted the data to pass/fail scores using a passing criterion of six or more trials correct. As expected, performance in the Lens test was significantly correlated with performance in the Standard-Format Lens test (Pearson's partial $r_{29} = 0.42$, $p = 0.018$). The Standard-Format Lens test was also positively correlated with the Verbal Color Filter test (partial $r_{29} = 0.36$, $p = 0.048$), but not with the Rock–sponge test (partial $r_{29} = -0.15$, $p = 0.41$). This is not surprising, given the greater similarities between the Lens and Color Filter tests (i.e., in those tests, but not the Rock–sponge test, children witnessed the visual transformation of the object and viewed the object through a transparent barrier). There were no other significant correlations (all $ps \geqslant 0.07$).

As children received one trial only of each of the additional appearance-reality tests we administered, it is possible that their performance in any given test did not truly reflect their capacities. Momentary distraction might have caused them to fail a test they would have otherwise passed, for example. Alternatively, they could have chosen the correct answer by chance, without understanding the distinction between appearance and reality. Reasons such as these, as well as differing task factors, may have been responsible for the lack of correlation found among the various tests. Thus, we created an aggregate proportion score for each child from the four additional appearance-reality tests we administered (the Standard-Format Lens test, the Rock–sponge test, and the two Color Filter tests), and then compared these scores to the proportion of trials correct in the Lens test using a Pearson correlation test with approximate significance values. The correlation was highly significant ($r_{30} = 0.44$, $p = 0.011$) and it remained marginally significant when controlling for age (partial $r_{29} = 0.33$, $p = 0.070$). This positive correlation was

even more apparent when we included only the Unseen trials from the Lens test, which were the more meaningful trials ($r_{30} = 0.46$, $p = 0.009$; partial $r_{29} = 0.42$, $p = 0.018$).

To summarize, performance in the Lens test was typical for children in the age range we tested. Four-year-olds did not exceed chance performance as a group whereas 4.5-year-olds did. Such a shift in appearance-reality test performance between 4 and 5 years of age is common in research using standard verbal tests (e.g., Flavell et al., 1986). This shift was also apparent in the additional appearance-reality tests we administered, with the exception of the Standard-Format Lens test. Furthermore, there were positive correlations between the Lens test and the aggregate of the four additional appearance-reality tests. All of these factors suggest that the Lens test is a valid measure of appearance-reality understanding.

4. General discussion

Like the young 4-year-olds we tested, chimpanzees did not do well as a group in our Lens test. Also like the children, however, some individual chimpanzees did succeed – and they did so continuously over a series of tests. Thus, while the capacity to distinguish appearance from reality may not be widespread among chimpanzees, it may nevertheless be present in some proportion of the population, perhaps especially among adolescents and young adults. Similar developmental patterns have been found for chimpanzees regarding other abilities, such as mirror self-recognition (Povinelli et al., 1993) and episodic-like memory (Martin-Ordas et al., submitted for publication). However, the relation we found in the current study between age and performance in the appearance-reality test should be interpreted cautiously because age was confounded with rearing history.

Of course, it is also possible that appearance-reality understanding is common among chimpanzees but that they found our task difficult for other reasons. Similarly, some researchers have suggested that younger children's problems with appearance-reality tasks are more a function of problematic task factors rather than of any inability to distinguish appearance from reality (e.g., Deák, 2006; Hansen & Markman, 2005; Sapp et al., 2000). One possibility is that the chimpanzees who failed even our Basic test may have been unable to inhibit choosing the grape that appeared bigger, even if they did recognize that it was not truly bigger (Boysen et al., 2001). Although we attempted to remove this 'pull of the apparent' by covering the lenses just before letting these individuals choose in the Reduced Inhibition test, this may not have effectively removed the impact of what they had seen a brief moment before. Completely eliminating response-inhibition issues in tests with young children or nonhuman animals is a continuing challenge (see, for example, Bialystok & Senman, 2004).

The other side of the coin is that the chimpanzees who passed our test may have done so not because they were capable of distinguishing appearance from reality but for some other reason. Based on the results of our control tests and other manipulations, we can rule out a number of alternative explanations. We know that successful chimpanzees did not simply learn to choose the small grape to get the big one because they could not learn to do this in the Reverse Contingency control tests. We also know that they did not just learn to avoid the magnifying lens or the magnified food because they did not do so when the truly bigger, rather than smaller, piece of food was magnified in the Big-to-bigger control trials. And some individuals clearly did not solve our task by visually tracking the grapes because they still succeeded when we made this impossible in the Unseen trials of the Tracking test. Our data, at the very least, show that some chimpanzees were able to overcome the misleading perceptual appearance of the grapes with regard to their size. Additionally, data on object individuation (Mendes, Rakoczy, & Call, 2008) has demonstrated that apes who watch an object disappear into a box and then later retrieve it from that box know that it is the very same object. Thus, we conclude that several of our chimpanzees did indeed appreciate that what they saw was at odds with what they knew to truly be the case. They recognized that the very same grape that looked smaller was really and truly the bigger one, and they chose accordingly. Future studies could examine whether chimpanzees are also capable of recognizing when the appearance of an object is at odds with its true identity or function, as in the standard Rock–sponge test.

Our findings have important implications for current debates on metacognition and mental state attribution in nonhuman animals. For example, can the successful chimpanzees in our study be described as showing evidence of metacognition, in the sense of reflecting upon their own mental states? Flavell et al. (1983) suggest that appearance-reality understanding is indeed metacognitive: "although always susceptible to being deceived by appearances, we have acquired the metacognitive knowledge that appearance-reality differences are always among life's possibilities" (p. 96). Carruthers (2008) also implies that if chimpanzees were to pass an appearance-reality test, this would be a more convincing indicator of metacognition than previous studies have offered. Put very simply, Carruthers' issue with the evidence to date is that the behaviors claimed to be indicative of metacognitive processes in animals can be explained without invoking any kind of reflection on one's own mental states. An animal need only have different mental states of varying strengths competing with one another, with the stronger exerting more influence over behavior than the weaker. When no potential response is more strongly motivated than any other, for example because the mental states involved are of similar strength, animals tend to react by seeking more information or choosing randomly.

Succeeding in our appearance-reality test is difficult to explain in Carruthers' terms because the beliefs and perceptions pitted against one another would seem to have equal strengths (i.e., a strong belief that the big grape is in container A is matched by a strong perception that the big grape is in container B). Under these conditions, and when further information seeking is not an option, chimpanzees would be expected to choose randomly. Conceivably, beliefs could win out over perceptions if they were for some reason naturally stronger; but, if anything, we would expect perceptions to exert the stronger influence.

When young children are asked about the real and apparent properties of objects, such as size or color, they most often err by answering both questions based on the object's perceptual appearance (Flavell et al., 1983; Taylor & Flavell, 1984). In keeping with this, in our Standard-Format Lens test children made 15 such 'phenomenism' errors but only three 'realism' errors (i.e., giving reality answers to both questions). Thus, if chimpanzees did not recognize that the apparently bigger grape was the truly smaller one, we would expect them to be heavily biased towards choosing that grape. Some chimpanzees did so, but some did not, and those that did not are evidence that at least some individual chimpanzees are capable of grasping the appearance-reality distinction.

That some chimpanzees are able to pass appearance-reality tests is exciting news, given that they have performed consistently poorly in false belief tests (Call & Tomasello, 1999; Kaminski et al., 2008; Krachun et al., 2009). It is true that in most false belief studies, group effects only have been analyzed, and so any positive results at the individual level would not have been detected. However, Krachun, Carpenter, Call, and Tomasello (submitted for publication) analyzed individual performance in five chimpanzees given a novel false belief test and they still found no evidence of false belief understanding. While representing others' false belief states may be beyond chimpanzees, the fact that some chimpanzees can recognize when they themselves are experiencing a false visual perception shows that they are not entirely incapable of recognizing conflicting mental states. The puzzle, then, is why some chimpanzees are capable of passing appearance-reality tests but not false belief tests, especially given how closely related these abilities appear to be in human children (Flavell, 1993). The answer to this question may lie in the enormous difference in complexity between these two situations. The first involves variation along just one dimension: the real versus apparent size of the grape. In contrast, the second varies along a number of dimensions: one's own mental state versus someone else's mental state, the true location of the object versus the false representation of its location, the particular events in a sequence that oneself has witnessed versus the ones that someone else has witnessed, and so on. The sheer complexity of false belief tests, at least the ones developed so far, may just be too cognitively taxing for chimpanzees.

It is also possible that some particular features of false belief tests are more problematic for chimpanzees than other features. For example, the kinds of mental states that chimpanzees are required to infer in false belief tests may be more difficult for them than the kinds of mental states to be inferred in other tests, such as in visual perspective taking tests. Gopnik and Astington (1988) speculated that this is why children show good visual perspective taking abilities earlier in development than they begin to pass false belief tests. Chimpanzees may similarly find perceptually based mental states easier to infer than more abstract ones such as beliefs. Supporting this is the fact that chimpanzees have performed well in tasks requiring them to take into account others' visual perceptions (Bräuer et al., 2007; Hare et al., 2000; Hare, Call, & Tomasello, 2006) or others' knowledge construed as what they have seen in the recent past (Hare, Call, & Tomasello, 2001; Kaminski et al., 2008).

Another possibility is that dealing with another individual's mental states, of any kind, may be harder than dealing with one's own. In this case, a chimpanzee who was capable of holding two conflicting representations of an object in mind simultaneously (e.g., a grape as truly big but apparently small) might nevertheless be incapable of simultaneously representing the real location of an object and someone else's mistaken belief about its location. In the literature on mental state attribution in humans, there is much debate about the relation between self-knowledge and knowledge of other minds; but a recent meta-analysis of a large number of studies suggests that there is no consistent difference between children's performance on self and other false belief tests (Wellman et al., 2001). Nevertheless, it could still be the case that chimpanzees find it easier to recognize their own mental states than to recognize others', even if this is not so for human children.

Comparative cognitive research may provide important perspectives on human cognitive evolution and development. If it indeed turns out to be a reliable finding that our nearest primate relatives (at least a reasonable proportion of them) can deal with problems involving appearance versus reality but not with problems involving false beliefs, this would provide important new constraints on our theory building. It would also suggest new lines of research aimed at understanding both apes' and children's ability to compare different kinds of mental states with one another (e.g., perceptions and beliefs), as well as their ability to compare their own mental states with those of others. Because chimpanzees do not have a natural language, it would also have important implications for the role of language (and other culturally constructed forms of representation) in human social-cognitive development in general.

Acknowledgements

We are grateful to Katrin Riedl and Stefanie Keupp for their assistance with the chimpanzee study and to Cornelia Schulze, Johanna Uebel, Henrik Roethel, and Kathrin Greve for their help with the child study.

References

Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development, 74*, 1476–1499.

Bialystok, E., & Senman, L. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75*, 562–579.

Boysen, S. T., & Berntson, G. G. (1995). Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 21*, 82–86.

Boysen, S. T., Berntson, G. G., Hannan, M. B., & Cacioppo, J. T. (1996). Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 22*, 76–86.

Boysen, S. T., Berntson, G. G., & Mukobi, K. L. (2001). Size matters: Impact of item size and quantity on array choice by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115*, 106–110.

Braine, M. D. S., & Shanks, B. L. (1965a). The conservation of a shape property and a proposal about the origin of the conservations. *Canadian Journal of Psychology, 19*, 197–207.

Braine, M. D. S., & Shanks, B. L. (1965b). The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior, 4*, 227–242.

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition, 10*, 439–448.

Call, J. (2005). The self and other: A missing link in comparative social cognition. In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 321–341). Oxford: Oxford University Press.

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition, 3*, 207–220.

Call, J., & Rochat, P. (1996). Liquid conservation in orangutans (*Pongo pygmaeus*) and humans (*Homo sapiens*): Individual differences and perceptual strategies. *Journal of Comparative Psychology, 110*, 219–232.

Call, J., & Rochat, P. (1997). Perceptual strategies in the estimation of physical quantities by orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 111*, 315–329.

Call, J., & Tomasello, M. (1996). The effect of humans on the cognitive development of apes. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought* (pp. 371–403). New York: Cambridge University Press.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70*, 381–395.

Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology, 20*, 393–420.

Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language, 23*, 58–89.

Deák, G. O. (2006). Do children really confuse appearance and reality? *Trends in Cognitive Sciences, 10*, 546–550.

Flavell, J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist, 41*, 418–425.

Flavell, J. H. (1993). The development of children's understanding of false belief and the appearance-reality distinction. *International Journal of Psychology, 28*, 595–604.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology, 15*, 95–120.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*. Serial No. 212.

Flavell, J. H., Green, F. L., Wahl, K. E., & Flavell, E. R. (1987). The effects of question clarification and memory aids on young children's performance on appearance-reality tasks. *Cognitive Development, 2*, 127–144.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*, 26–37.

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of three-year-olds' understanding of false belief. *Child Development, 62*, 83–97.

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences, 98*, 5359–5362.

Hansen, M. B., & Markman, E. M. (2005). Appearance questions can be misleading: A discourse-based account of the appearance-reality problem. *Cognitive Psychology, 50*, 233–263.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour, 59*, 771–785.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*, 139–151.

Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition, 101*, 495–514.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition, 109*, 224–234.

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint-seeking in monkeys. *Psychological Science, 18*, 64–71.

Krachun, C., & Call, J. (2009). Chimpanzees (*Pan troglodytes*) know what can be seen from where. *Animal Cognition, 12*, 317–331.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science, 12*, 521–535.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (submitted for publication). A new change-of-contents false belief test: Children and chimpanzees compared.

Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development, 61*, 1514–1519.

Martin-Ordas, G., Call, J., Haun, D., & Colmenares, F. (submitted for publication). Great ape memories for past events show a human-like developmental pattern.

Matsui, T., & Miura, Y. (2008). Pro-social motive promotes early understanding of false belief. In *Nature proceedings*. <hdl:10101/npre.2008.1695.1>.

Mendes, N., Rakoczy, H., & Call, J. (2008). Ape metaphysics: Object individuation without language. *Cognition, 106*, 730–749.

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development, 61*, 722–730.

Muncer, S. J. (1983). "Conservations" with a chimpanzee. *Developmental Psychobiology, 16*, 1–11.

Murray, F. B. (1968). Phenomenal-real discrimination and the conservation of illusion-distorted length. *Canadian Journal of Psychology, 22*, 114–121.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125–137.

Povinelli, D. J., Rulf, A. B., Landau, K. R., & Bierschwale, D. T. (1993). Self-recognition in chimpanzees (*Pan troglodytes*): Distribution, ontogeny, and patterns of emergence. *Journal of Comparative Psychology, 107*, 347–372.

Rice, C., Koinis, D., Sullivan, K., Tager-Flusberg, H., & Winner, E. (1997). When 3-year-olds pass the appearance-reality test. *Developmental Psychology, 33*, 54–61.

Sapp, F., Lee, K., & Muir, D. (2000). Three-year-olds' difficulty with the appearance-reality distinction: Is it real or is it apparent? *Developmental Psychology, 36*, 547–560.

Siegal, M. (1997). *Knowing children: Experiments in conversation and cognition* (2nd ed.). London: Psychology Press.

Suda, C., & Call, J. (2004). Piagetian liquid conservation in the great apes (*Pan paniscus, Pan troglodytes,* and *Pongo pygmaeus*). *Journal of Comparative Psychology, 118*, 265–279.

Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56*, 135–148.

Taylor, M., & Flavell, J. H. (1984). Seeing and believing: Children's understanding of the distinction between appearance and reality. *Child Development, 55*, 1710–1720.

Taylor, M., & Hort, B. (1990). Can children be trained in making the distinction between appearance and reality? *Cognitive Development, 5*, 89–99.

Uher, J., & Call, J. (2008). How the great apes (*Pan troglodytes, Pongo pygmaeus, Pan paniscus, Gorilla gorilla*) perform on the reversed reward contingency task II: Transfer to new quantities, long-term retention, and the impact of quantity ratios. *Journal of Comparative Psychology, 122*, 204–212.

Vlamings, P. H. J. M., Uher, J., & Call, J. (2006). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, and *Gorilla gorilla*) perform on the reversed contingency task: The effects of food quantity and food visibility. *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 60–70.

Washburn, D. A., Smith, J. D., & Shields, W. E. (2006). Rhesus monkeys (*Macaca mulatta*) immediately generalize the *uncertain* response. *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 185–189.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*, 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Woodruff, G., Premack, D., & Kennel, K. (1978). Conservation of liquid and solid quantity by the chimpanzee. *Science, 202*, 991–994.