

## BRIEF REPORT

# All Great Ape Species (*Gorilla gorilla*, *Pan paniscus*, *Pan troglodytes*, *Pongo abelii*) and Two-and-a-Half-Year-Old Children (*Homo sapiens*) Discriminate Appearance From Reality

Katja Karg and Martin Schmelz

Max Planck Institute for Evolutionary Anthropology, Leipzig,  
Germany

Josep Call

Max Planck Institute for Evolutionary Anthropology, Leipzig,  
Germany and University of St Andrews

Michael Tomasello

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Nonhuman great apes and human children were tested for an understanding that appearance does not always correspond to reality. Subjects were 29 great apes (bonobos [*Pan paniscus*], chimpanzees [*Pan troglodytes*], gorillas [*Gorilla gorilla*], and orangutans [*Pongo abelii*]) and 24 2½-year-old children. In our task, we occluded portions of 1 large and 1 small food stick such that the size relations seemed reversed. Subjects could then choose which one they wanted. There was 1 control condition and 2 experimental conditions (administered within subjects). In the control condition subjects saw only the apparent stick sizes, whereas in the 2 experimental conditions they saw the true stick sizes as well (the difference between them being what the subjects saw first: the apparent or the real stick sizes). All great ape species and children successfully identified the bigger stick, despite its smaller appearance, in the experimental conditions, but not in the control. We discuss these results in relation to the understanding of object permanence and conservation, and exclude reversed reward contingency learning as an explanation.

**Keywords:** appearance–reality, Piagetian conservation, perspective taking, size discrimination, quantity estimation

The ability to distinguish appearance from reality is ecologically significant in a broad variety of situations (Flavell, 1986). Mistaking a snake for a tree branch or an insect for a leaf can have deadly consequences. The ability to avoid judgment mistakes can thus be an evolutionary advantage by helping to effectively forage in the

wild and to protect oneself against predators. Although mimicry and camouflage detection have been studied extensively in the animal kingdom (e.g., Owen, 1982; Wickler, 1968), surprisingly little is known about whether our closest living relatives, the great apes, possess the skill to detect deceptive attempts when their perception differs from reality. As inhabitants of the rainforest, this ability would surely be advantageous for them.

In contrast, a large body of work has been done with human children, starting with the pioneer work of Braine and Shanks (1965a, 1965b). It is Flavell and colleagues however, who should be credited with developing a systematic research program (e.g., Flavell, Flavell, & Green, 1987; Flavell, Green, & Flavell, 1986, 1989; Flavell, Zhang, Zou, Dong, & Qi, 1983). The paradigms to test appearance–reality (henceforth: AR) understanding can be broadly distinguished into two categories. In the first set of studies, the child is confronted with a deceptive object whose true identity is at odds with its visual (or in some cases tactile) appearance. Probably most famous is the so-called *rock-sponge* test (Flavell et al., 1983, 1986). Here, the child is confronted with a sponge that is realistically painted like a rock. To pass the test, the child has to correctly answer two questions: First, what the object “really and truly” is, and second, what it “only looks like.”

In a second set of studies, a distinct feature of an object appears to change while the child is watching, such as an object’s color

---

*Editor’s Note.* Charles Snowdon served as the action editor this article.—JC

---

This article was published Online First August 25, 2014.

Katja Karg and Martin Schmelz, Department of Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; Josep Call, Department of Psychology, Max Planck Institute for Evolutionary Anthropology, and School of Psychology and Neuroscience, University of St Andrews; Michael Tomasello, Department of Psychology, Max Planck Institute for Evolutionary Anthropology.

We thank the German National Academic Foundation for financial support. We thank the children, parents, kindergartens, and animal caregivers at the Wolfgang Köhler Research Center for their cooperation, as well as R. Pieszek for building the apparatuses.

Correspondence concerning this article should be addressed to Katja Karg, Department of Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04109 Leipzig, Germany. E-mail: katja\_karg@eva.mpg.de

(Flavell et al., 1986, 1987), shape (Flavell et al., 1983), or size (Braine & Shanks, 1965b; Elkind, 1966; Flavell et al., 1983). In all cases, the child has to determine the object's real and apparent property after its transformation.

Typically, children come to understand these classic verbal tasks by the age of 4 years, and cross-cultural comparisons with Mandarin-speaking children have found the same absolute levels of performance, error patterns, and age trends (Flavell et al., 1983), leading researchers to conclude that this reflects a deep-seated cognitive deficit in 3-year-old children's ability to hold dual representations of an object (Flavell, Green, & Flavell, 1990).

Several scholars have criticized these early studies for their high linguistic demands that might be the true reason for the failure of 3-year-olds (e.g., Deák, 2006; Siegal, 1997). It is thus surprising that there have been few attempts to test appearance–reality understanding in children nonverbally. King (1971) used the Müller-Lyer illusion to change the relations of a small and a long stick and compared the performance of 4- to 6-year-old children in a verbal and a nonverbal version of the task. He found that 58% of the children were successful in the nonverbal task, whereas only 30% answered both appearance–reality questions correctly. Sapp, Lee, and Muir (2000) and Moll and Tomasello (2012) found 3-year-old children to be successful in appearance–reality tasks when the linguistic demands were reduced. Overall, the current research suggests that although 3-year-olds typically fail in classic verbal appearance–reality tasks, they might not generally lack the understanding that an object's appearance can differ from its true identity or properties.

Analogous results in research on perspective taking support this view. These tasks also require the ability to construe an object or situation in two conflicting ways. Flavell himself found that AR skills correlate highly with perspective-taking skills in 3-year-olds (Flavell, 1986). In addition, Moll and Meltzoff (2011) could show that 3-year-olds know to which of two blue objects an adult is referring to as the “green” one when he or she sees it through a yellow color filter (while to the children it appeared in its true, blue color).

Though a large body of literature demonstrates that various nonhuman animal species experience a dissociation between the physical and the perceptual world, particularly when confronted with visual illusions (for a review, see Fujita, 2012), surprisingly little is known about whether animals understand that appearance differs from reality in these illusionary moments. To our knowledge, there has been only one study investigating AR in nonhuman animals. Krachun, Call, and Tomasello (2009) studied whether chimpanzees can distinguish appearance from reality by adapting the classic lens test developed for human children (Braine & Shanks, 1965b; Flavell et al., 1983). In their study, subjects chose between a small and a big grape. However, before their choice the apparent sizes of the grapes were reversed using magnifying and minimizing lenses so that the bigger grape appeared to be the smaller one. Whereas the chimpanzees failed on a group level, several individuals seemed to appreciate the appearance–reality distinction. Four-year-old children who were given an adapted version of the task failed, while 4.5-year-olds succeeded.

Although Krachun and colleagues found some evidence for the AR distinction in chimpanzees, the results were not consistent across subjects. One reason for this might be the unnatural devices used in this study. Chimpanzees do not encounter magnifying or

minimizing lenses in their natural environment (this effect may at best be represented by the magnifying effect of water on objects). However, apparent size changes do occur in apes' natural habitat—by partial occlusion. A big fruit can appear small when covered by a leaf, and an enormous snake can appear much smaller when only its tail is visible. To differentiate between appearance and reality that arises through partial occlusion, at least four challenges must be met: (a) Perceive the real size of the objects, (b) understand that the occluded part of the object persists, even if it is not visible, (c) understand that the occluded object did not undergo an identity change, but only a change in appearance, and (d) overcome the perceptual salience of the changes in appearance at the time of choice (Bruner, 1966). Whereas (a) and (d) relate to attention and inhibition problems, (b) and (c) are logical problems that require an understanding of both object permanence and object conservation (Piaget, 1954, 1961).

Numerous animal species are capable of object permanence in the sense that objects that disappear from sight continue to exist (see Dore & Dumas, 1987; Tomasello & Call, 1997), although only great apes seem to cope well with cases in which objects undergo multiple invisible displacements (Jaakkola, 2014). Additionally, apes seem capable of object conservation (e.g., Suda & Call, 2005; Woodruff, Premack, & Kennel, 1978) defined as the ability to understand that essential properties of physical substances remain invariant regardless of perceptual changes (Piaget, 1952; Piaget & Inhelder, 1941). According to Piaget, children develop conservation skills by the age of 7 to 12 years (Piaget, 1952; Piaget & Inhelder, 1969), but more recent literature suggests that they can succeed in conservation tasks much earlier, when specific difficulties (e.g., language demands) are circumvented (Braine & Shanks, 1965b; Gelman, 1972; McGarrigle & Donaldson, 1975). Thus, both great apes and young children possess the required abilities to allow them to distinguish between appearance and reality. However, whether they are indeed able to do so, is still largely unknown.

We tested AR in all four great ape species—bonobos, chimpanzees, gorillas, and orangutans. We aimed to create a more naturalistic conflict between appearance and reality by partially occluding one large and one small edible pretzel stick such that the size relations were reversed; the bigger reward appeared to be the smaller one and vice versa. We then let the subject choose the reward they wanted to receive. As some chimpanzees appreciated the appearance–reality distinction in Krachun and colleagues' (2009) lens test, we expected chimpanzees to be more successful in this setting. For the other great ape species, we could not make any clear predictions for their success or failure because of the lack of data in appearance–reality tasks. However, because bonobos and orangutans have been proven to possess conservation skills, we hypothesized that these may help them to solve the task. We administered our test also to 2½-year-old children. To our knowledge, in this age group neither AR discrimination nor conservation abilities have been shown so far. We chose this age group because our task had very few verbal demands. Additionally, children of that age already have plenty of experience with partially occluded objects.

All previous “overconservation” AR tasks with children as well as the one study with chimpanzees (Krachun et al., 2009) confronted subjects with the true size relations first, before the relations (in size or quantity) were reversed. Yet we were interested in

the question of how children's and great apes' performance would change if their first impression of the objects was wrong, and the true size relations are revealed thereafter. In this respect, two well-known psychological phenomena in humans are a first impression bias and confirmatory bias. The first impression bias refers to a limitation of human information processing in which people are strongly influenced by the first piece of information they are exposed to, and that they are biased in evaluating subsequent information in the direction of the initial influence (Lim, Benbasat, & Ward, 2000). Confirmatory bias is strongly related to the first impression bias and describes the human tendency to ignore additional relevant information or even misread contradicting evidence if they have already formed an initial hypothesis about a situation or person (Rabin & Schrag, 1999). We were thus expecting that a false first impression might cause more difficulties in identifying the truly bigger stick for human children, and potentially also great apes.

We thus presented two experimental conditions to the subjects: One in which subjects first saw the real stick size relation, and one in which they saw the apparent (deceptive) stick size relation first, but were allowed to see the real size relation thereafter. In a control condition, subjects only saw the apparent stick sizes and chose without further information. Choosing the really bigger stick should therefore be hardest in the control condition, less hard in the condition with a wrong first impression, and easiest in the condition with a true first impression.

## Method

### Subjects

Seven bonobos (*Pan paniscus*), 13 chimpanzees (*Pan troglodytes*), four gorillas (*Gorilla gorilla*), and five orangutans (*Pongo abelii*) participated in this experiment (see Table 1). Ten apes were nursery-reared (together with peers by human caretakers) whereas 19 were mother-reared within a group of conspecifics. All were born in captivity and housed at the Wolfgang Koehler Primate Research Center at the Leipzig Zoo (Germany), where they lived in social groups and had access to indoor and outdoor areas. Subjects were tested in their indoor rooms, were fed according to their daily routine, and were not food or water deprived at any time. Subjects all had previous experience with participating in experimental studies and were therefore familiar with the context.

Twenty-four 2½-year-old children (13 female, 11 male; age range: 2y, 4m–2y, 7m) from kindergartens in Leipzig, Germany, also participated in this experiment. Five further children (2 fe-

male, 3 male) were dropped from the study due to illness on at least one of the three testing days ( $n = 3$ ), shyness ( $n = 1$ ), or uncooperativeness ( $n = 1$ ).

### Materials

The apparatus consisted of a flat sliding board ( $78.5 \times 30.5$  cm) with two rubber strings attached to it on the left and right side for fixing the pretzel stick rewards (45.5 cm apart, 16.5 cm from the side, 18.5 cm from the subject's bottom side). The rewards could be partially covered by a black board ( $78.5 \times 19$  cm), such that the bigger stick appeared to be smaller than the small stick (see Figure 1).

For the apes, the sliding board was placed in front of a Plexiglas panel ( $64 \times 73$  cm) in the enclosure wall. At the panel's bottom were two holes on the left and right side (6.5 cm diameter, 43 cm apart), corresponding to the two stick rewards. A red occluder ( $80 \times 50$  cm) was used to block the subject's view during baiting. Big and small pretzel sticks (13-cm and 6.5-cm long, 4 mm in diameter) were used as rewards.

The children sat at a table in front of the sliding board. A red curtain could block their view. Instead of pretzel sticks, wooden sticks of the same size as the pretzel sticks in the ape set-up were used for the children, which they could later feed to a duck puppet.

### Design

Each subject received three conditions on different days, with the order of conditions counterbalanced for sex (children) and age and sex (apes). The mean age of the apes in each of the two order groups was 17.6 and 19.6 years.

In each daily test session, we administered 12 trials of the same condition. Apes received two consecutive sessions of each condition, resulting in 24 trials per condition and 72 trials total. Children received one session of each condition, resulting in 12 trials per condition and 36 trials total. There was a minimum of 1 and a maximum of 7 days between sessions.

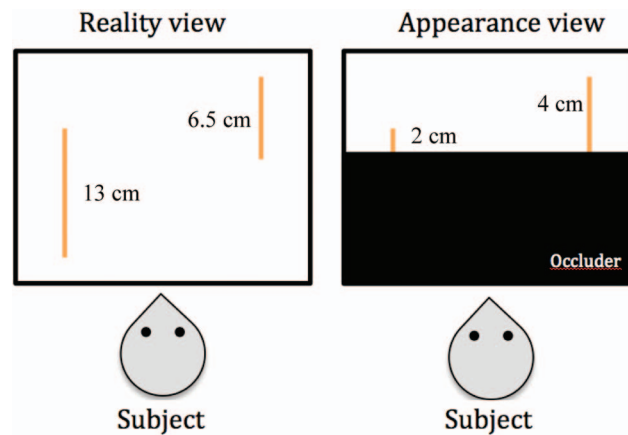
### Procedure

Each subject was tested individually, children in a room in their kindergarten, apes in their sleeping or testing rooms at the zoo. Young ape infants were allowed to stay with their mothers, but trials started only if the infants were not interfering with the test setup.

As a prerequisite for the experiment, subjects needed to be motivated to choose the bigger of two pretzel sticks. Ape subjects are naturally motivated to do so (see Hanus & Call, 2007). To ensure a similar motivation in children, the experimenter established a feeding context. She introduced a duck hand puppet that was explained to be very hungry, and whose favorite food was (wooden) "duck pretzel sticks." The duck liked only the big sticks, and would not eat the small sticks. The experimenter demonstrated the duck's preference by showing it a big and a small stick, and letting it choose its favorite stick. The duck chose the big stick and the experimenter subsequently fed this stick to it through a little hole in its mouth (the experimenter could hold the inserted stick in her hand inside the puppet and remove the stick later in an unobserved moment). The experimenter then tried to feed the duck

Table 1  
Sex and Age Distributions of All Participating Great Ape Species

Species	Sex		Age (years)		Total
	Female	Male	Mean	Range	
Bonobos	4	3	17.4	7–30	7
Chimpanzees	9	4	20.5	7–36	13
Gorillas	3	1	11.0	6–17	4
Orangutans	4	1	20.6	9–32	5
Total	20	9	18.4	6–36	29



*Figure 1.* Reality (left side) and appearance view (right side) of the sticks from the subject's perspective. When the sticks are occluded, the big stick appears smaller than the small stick. The color version of this figure appears in the online article only.

the small stick and it refused. Then she checked the child's understanding of the duck's preference by letting the child choose one stick from between one big and one small stick. Most children (88%) picked the big stick from the very beginning. If not, the experimenter stressed again the duck's preference and repeated the procedure until the child picked the big stick (median of training trials needed across 24 tested children:  $M = 1$ , range 1–3). When the child chose the big stick, the duck jumped happily up and down, and the child was allowed to feed it through the hole in its mouth. When the child chose the small stick, the duck sadly hung her head and said that she did not like the small sticks and that she would eat only the big sticks. Children generally reacted positively to the duck and seemed to enjoy feeding it.

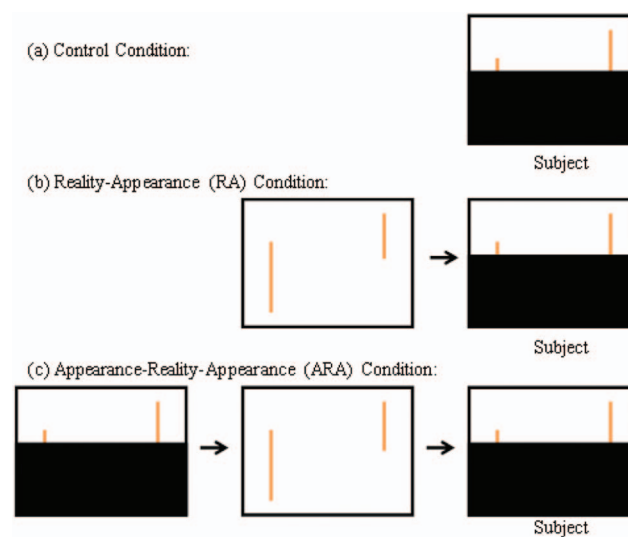
Before each trial, the sliding table was pulled away from the subject, the subject's view was blocked and the two rubber strings were baited with one big and one small pretzel stick according to

a predefined scheme. For all sessions, both stick lengths were presented on the left and right side an equal number of times, and the side that the big stick appeared on first was also counterbalanced between subjects. The position of the sticks was randomly determined with the only constraint being that the same constellation could not be presented on more than three consecutive trials.

Each trial started with the experimenter removing the occluder, allowing the subjects to observe the stick constellation. What subjects saw next differed between the conditions in the following way (see also Figure 2):

(a) Control condition: Only the covered version ("appearance view") of the sticks.

(b) Reality–Appearance (RA) condition: First the real length of the sticks ("reality view"), then the board was placed on the sticks, leading to the appearance view.



*Figure 2.* Experimental procedure in each of the three conditions. The big and the small stick were presented either covered by a black board or uncovered. The color version of this figure appears in the online article only.



(c) Appearance–Reality–Appearance (ARA) condition: First the covered version of the sticks. Next, the board was removed and the subjects were allowed to see the real stick lengths. The board was repositioned thereafter, leading to the appearance view.

Then the sliding table was pushed toward the subject in all conditions.

Each view (appearance or reality) was presented for about 3 seconds before the next step began. As soon as the sliding table was pushed toward the subject, the subject was allowed to choose one of the two sticks by poking a finger through the corresponding hole (apes) or by pointing to one of the two sticks (children). The table was only pushed toward the subject if the subject did not have their fingers in one or both holes (apes) or on the table (children). Crucially, when subjects made their choice, the sticks were always covered with the board. The experimenter avoided potential cueing in the following ways: First, when she pushed the table toward the subject, she used both hands ensuring that each stick had a hand close to it; second, while pushing the table, the experimenter looked down at the center between the two sticks; and third, in case she had to wait for the subject to make a choice, she folded her hands in her lap and kept her eyes down.

Since in all of our conditions the truly bigger stick appeared to be the smaller, it is conceivable that subjects might learn the rule of always choosing the apparently smaller stick. Although we could have conducted a control condition in which the larger stick also remained the apparently larger one after occlusion, we decided against this possibility because previous studies have shown that our apes (e.g., Hanus & Call, 2007), just like those included in other studies (e.g., Boysen & Berntson, 1995; Beran, 2001), have a strong preference for larger food quantities. Thus, we suspected that this condition would not be very informative. Instead, we decided to compare subjects with different levels of learning experience (due to different numbers of preceding experimental conditions) in their performance in the control condition. If subjects just learned a rule to always choose the apparent smaller stick, we expected that those with extensive learning experience would be more successful in the control condition than naive subjects.

### Data Scoring and Analysis

For ape subjects, a choice was considered made when she poked her finger through one of the holes after the table was pushed toward them. Children were asked which stick they would want to feed to the duck before the table was pushed toward them. They made their choice by either touching one of the two sticks or by unambiguously pointing to one of them. All choices were live coded, but trials were also videotaped for later analysis. A second independent observer coded a random sample of 20% of all the sessions for reliability. The interrater agreement was excellent (Cohen's  $\kappa = .92$ ,  $p < .0001$ , for apes and  $= 1.0$ ,  $p < .0001$ , for children).

As apes and children had differing procedures, we initially analyzed them separately. We used repeated-measures ANOVAs with Greenhouse-Geisser adjustment with condition as a within-subject factor and sex and order of conditions as between-subjects factors. Additionally, we analyzed species as a factor for the ape data. We also conducted a repeated-measures ANOVA to directly compare the performance of apes (with all species pooled together)

and children (while being aware that differences might partially result from variations in the procedure). We used one-way ANOVAs to compare the performance of subjects in the control condition after different amounts of exposure to apparent size changes (control first/second/last) and to compare their performance in their first conditions only. In addition, we used one-sample  $t$  tests to examine whether their performance differed from chance in the various conditions.

### Results

Figure 3 shows the proportion of trials in which apes and children chose the bigger stick across the three conditions. Focusing on the apes revealed a highly significant effect of condition,  $F(1, 4) = 127.45$ ,  $p < .001$ ,  $\eta^2 = .98$ , and pairwise comparisons revealed that this effect was driven by them choosing the bigger stick significantly less often in the control condition than in the RA or ARA conditions (Bonferroni-corrected  $p = .002$  and  $p = .003$ , respectively). In contrast, the apes were equally successful in the RA and ARA condition ( $p = 1.0$ ). No other factors, including species,  $F(3, 3) = .429$ ,  $p = .75$ ,  $\eta^2 = .30$ , nor any interactions were significant ( $p > .05$ ). Overall, apes were highly successful in identifying the bigger stick in the RA ( $M = 85.9\%$ , 95% CI [80.0, 93.3]) and ARA condition ( $M = 84.5\%$ , 95% CI [77.0, 92.6]), but much less so in the control condition ( $M = 20.5\%$ , 95% CI [12.9, 26.2]).

Although children's success rates were lower in the RA ( $M = 65.6\%$ , 95% CI [57.9, 73.3]) and ARA ( $M = 68.8\%$ , 95% CI [62.7, 75.5]) condition compared to the apes, the general pattern was the same. Their performance in the control condition ( $M = 23.6\%$ , 95% CI [18.7, 28.6]) was similar to that of the apes and there was

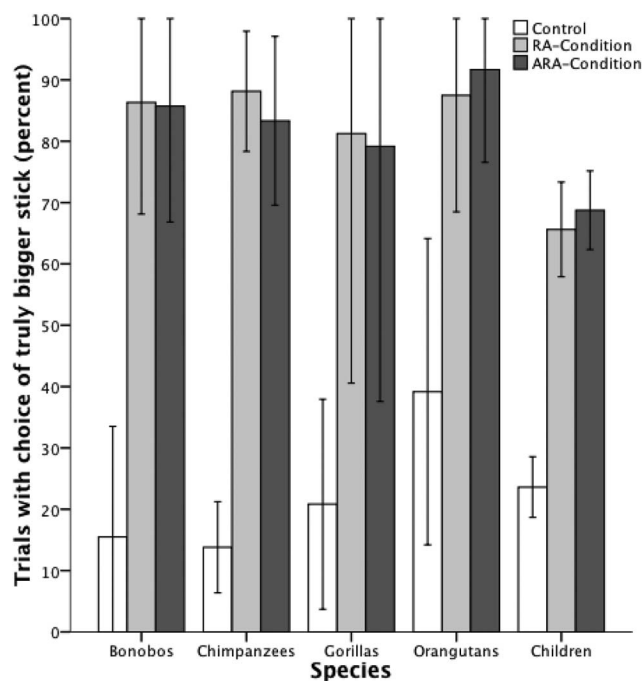


Figure 3. Mean percent of trials in which subjects chose the truly bigger stick, separated for species. Error bars refer to 95% confidence intervals.

a significant difference between conditions,  $F(2, 20) = 54.30$ ,  $p < .001$ ,  $\eta^2 = .82$ , driven by their performance in the RA and ARA condition being significantly better than in the control condition (Bonferroni-corrected  $p < .001$  in both comparisons). Children did not perform differently in the RA and ARA condition ( $p = 1.0$ ).

A repeated-measures ANOVA including condition as within-subject factor and sex, order of conditions, and group (apes vs. children) as between-subjects factors revealed a significant effect of group,  $F(1, 29) = 13.37$ ,  $p = .001$ ,  $\eta^2 = .32$ , condition,  $F(1, 43) = 150.82$ ,  $p < .001$ ,  $\eta^2 = .84$ , and no effect of sex,  $F(1, 29) = .07$ ,  $p = .79$ ,  $\eta^2 = .002$ , or order of conditions,  $F(5, 29) = 1.01$ ,  $p = .43$ ,  $\eta^2 = .15$ . There was a significant interaction between condition and group,  $F(1, 43) = 5.64$ ,  $p = .012$ ,  $\eta^2 = .16$ . Post hoc pairwise comparisons revealed that apes performed significantly better than children in the RA and the ARA condition, one-sample  $t$  test:  $t(51) = 4.27$ ,  $p < .001$  and  $t(51) = 3.17$ ,  $p = .003$ , respectively, whereas there was no difference in the control condition,  $t(49) = 1.01$ ,  $p = .32$ . Both apes and children performed above chance in the RA, one-sample  $t$  test: apes:  $t(28) = 11.31$ ,  $p < .001$ ; children:  $t(23) = 4.19$ ,  $p < .001$ , and ARA conditions, one-sample  $t$  test: apes:  $t(28) = 9.14$ ,  $p < .001$ ; children:  $t(23) = 6.04$ ,  $p < .001$ , and below chance in the control condition, one-sample  $t$  test: apes:  $t(28) = 9.39$ ,  $p < .001$ ; children:  $t(23) = 11.05$ ,  $p < .001$ .

We checked our data for possible learning effects by comparing subjects' performance in their first and their last experimental session (RA or ARA condition). Performance levels did not significantly increase with experience, apes, first session:  $M = 85.3\%$ , 95% CI [77.2, 93.5], last session:  $M = 86.2\%$ , 95% CI [79.5, 92.8];  $t$  test:  $t(28) = .31$ ;  $p = .76$ ; children, first session:  $M = 63.9\%$ , 95% CI [56.1, 71.6], last session:  $M = 67.7\%$ , 95% CI [61.0, 74.4],  $t(23) = .96$ ;  $p = .35$ .

We also analyzed whether subjects learned the rule of always choosing the apparent smaller stick by comparing their performance in the control condition as a function of the number of preceding conditions in which they could have experienced apparent size changes (RA and ARA condition). There was no significant difference between more and less experienced apes,  $F(2, 26) = .29$ ,  $p = .75$ , see Figure 4. In contrast, children differed significantly in their performance in the control condition depending on their prior experience,  $F(2, 21) = 7.3$ ,  $p = .004$ . However, this difference arose from their better performance without experience (control first), compared to the extensive experience (control last; Bonferroni-corrected post hoc test:  $p = .003$ ).

To control for any other confounding learning effects across sessions, we compared the performance between subjects in their first condition only. Despite the smaller cell sizes, the observed patterns remained the same both for apes,  $F(2, 28) = 36.28$ ,  $p < .001$  and children,  $F(2, 23) = 5.88$ ,  $p = .009$ . Subjects chose the bigger stick significantly less often in the control condition than in the RA or ARA condition (Bonferroni-corrected post hoc test: apes: both  $p < .001$ ; children: Control-RA:  $p = .034$ , Control-ARA:  $p = .015$ ), and there was no difference for neither group between the RA and ARA condition ( $p = 1.0$ ).

## Discussion

Our study resulted in two key findings. First, all great ape species were successful in identifying the bigger stick if they had

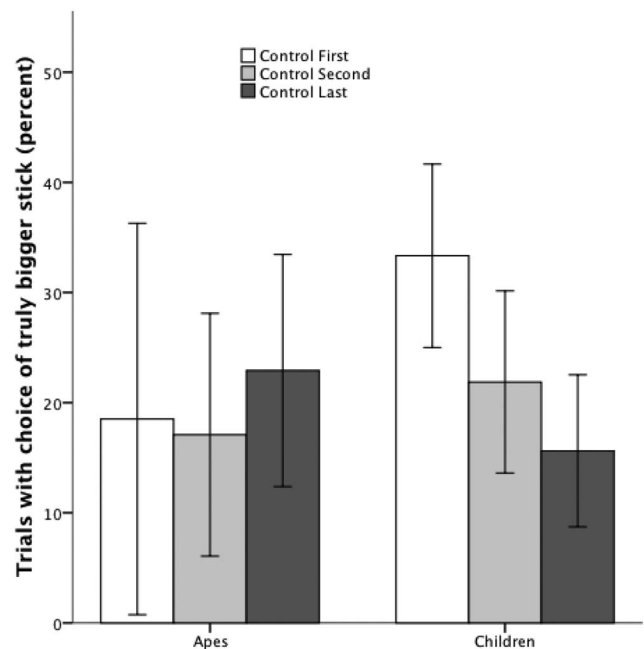


Figure 4. Mean percent of trials in which subjects chose the truly bigger stick in the control condition, compared between subjects with differing amount of learning experience (through differing numbers of preceding experimental conditions). Subjects with extensive experience with the size reversal (control last) did not perform better than subjects without such experience (control first or second). Error bars refer to 95% confidence intervals.

already seen the real stick size, even though their perceptions did not reflect reality in the moment of choice. Also, 2½-year-old children performed above chance, but were less successful than the apes. Second, we found that in the control condition apes and children performed well below chance: They consistently selected the apparently bigger, but really smaller stick. This behavior did not change with increasing number of trials (not even after they sometimes picked the truly bigger stick by accident) and did not depend on the order of conditions. Thus, even subjects who received the control condition last and had plenty of opportunity to learn about the reversed size relation in the previous two conditions did not choose the truly bigger stick. This finding is in line with previous research that shows great apes' general difficulties in learning reversed reward contingency rules. In several studies, chimpanzees were trained to choose the smaller of two food quantities to receive the bigger one (Boysen & Berntson, 1995; Uher & Call, 2008; Vlamings, Uher, & Call, 2006). None of the 12 naive subjects tested performed above chance before completing 250 trials. Generally, chimpanzees have significant difficulties with inhibitory control when confronted with such perceptual seductions (Boysen, 1996). These difficulties rise with increasing ratio of the presented food quantities (Boysen, Berntson, & Mukobi, 2001; Uher & Call, 2008). Three-year-old children also have difficulties inhibiting to reach for a larger food quantity (Russell, Mauthner, Sharpe, & Tidswell, 1991). Against this background, the performance of great apes and children in the RA and ARA conditions is even more remarkable.

Previous research has found evidence that individual chimpanzees are able to discriminate appearance and reality. Krachun et al.

(2009) reversed the size relations of a big and a small grape by using magnifying and minimizing lenses, and some of the tested individuals were successful in this task. However, as chimpanzees failed on a group level, it remained unclear whether this ability was generalizable or due to some exceptional individual abilities. The current study supports the former view by showing that chimpanzees show group-level success when they are confronted with more naturally occurring perceptual challenges. Furthermore, for the first time all other ape species also demonstrated their ability to discriminate appearance from reality. Especially their conservation skills could have helped them solve the task (Suda & Call, 2005).

The poorer performance of children compared to apes might be explained by differences in motivation, as children did not eat the pretzel sticks themselves, but only fed them to a duck puppet. Also, the limited executive attention of 2½-year-olds, that starts improving only by the age of 2 years up to the age of 7, could have constrained their performance (Rueda, Posner, & Rothbart, 2005). Recent AR literature suggests that children can accomplish nonverbal AR tasks by the age of 3 years (Moll & Tomasello, 2012; Sapp et al., 2000), but to our knowledge this is the first study reporting AR skills at an even younger age.

One main question arising from these findings is whether our task could have been solved by a simpler mechanism other than through an understanding of the AR distinction. We would like to discuss three alternative explanations. First, subjects could have succeeded by learning the simple rule of always choosing the smaller-looking stick or always avoiding the bigger-looking stick (reversed reward contingency learning). Indeed, this would be a successful strategy; across all conditions, the smaller-looking stick was always the bigger stick in reality at the moment of choice. However, we can rule out this explanation for several reasons. (a) Even subjects who received the control condition last and therefore had the opportunity to learn such a rule in 48 (apes) or 24 (children) preceding trials did not perform better than subjects without such extensive experience. (b) We would expect individual improvement over the sessions when learning such a rule. However, we did not observe an improvement across experimental sessions. (c) The results remain the same even when comparing naïve subjects in their first condition only. (d) Krachun and colleagues (2009) included two reversed reward contingency control tests in their experiment, in which chimpanzees had to point to the smaller of two grapes to receive the bigger one. Subjects performed poorly and did not show any sign of learning over the 12 trials. We thus conclude that reversed reward contingency learning cannot explain the success of subjects in our study.

Second, subjects may have solved the task because of a proximity bias or by simple visual tracking. In the case of a proximity bias, subjects in the experimental conditions could have disregarded the stick sizes and acted according to the following rule: Choose the stick that was closest to you before (although both sticks were at the same distance from the subject at the moment of choice, see Figure 1). However, we would then expect that they would choose randomly in the control condition, when both sticks are at the same distance from the subject. In contrast, we found that they had a clear preference for the apparently bigger stick in the control condition.

With regard to visual tracking, subjects could have followed the bigger stick with their eyes until the moment of choice. We cannot completely exclude this possibility. For example, in the RA con-

dition, they could visually identify the big stick, ignore the subsequent covering and reversal of the size relations while continually tracking the location of the big stick, and then act according to their original preference. However, applying this same rule in the ARA condition is problematic because the stick that appears first as the bigger stick turns out, in reality, to be the smaller one in the next step. If subjects just tracked the stick that looked bigger in the first place, they would end up choosing the smaller one in reality. But this is not what we observed. We found that subjects performed equally well in the RA and ARA conditions, meaning that their first impressions did not necessarily determine their choices. In fact, we often observed that subjects in the ARA condition first pointed to the smaller stick when they saw the deceiving situation, but then corrected their pointing when the sticks were fully uncovered. However, when the sticks were covered again, subjects did not change back to their original choice. The two representational situations thus did not coexist as equally relevant choices. Instead, subjects evaluated one as being more informative for guiding their decisions. Hence, even if they did not judge their first impression as “wrong” or “only apparent,” the competing representation of the real stick length in the second step was strong enough to overwrite their first memory trace. Furthermore, Krachun and colleagues (2009) excluded the explanation of tracking in a control condition in which the subjects’ visual access to the rewards was blocked between the initial baiting event behind the magnifying/minimizing lens and the choice situation. Several subjects still successfully chose the truly bigger grape here.

A third alternative is that conservation abilities, not AR abilities, are sufficient to solve the task. We agree that conservation may be enough for solving our task. However, the same applies to all appearance–reality tests with seemingly changing object properties. Subjects have to conserve the true property of the object to overcome the lure of the apparent property later. This property to conserve refers to size in the lens test, object color in the color filter test, or the stick straightness after putting it in water—all examples of classic appearance–reality test in the child literature (Braine & Shanks, 1965b; Flavell et al., 1986). Actually, some researchers have proposed that understanding the appearance–reality distinction is even a prerequisite for succeeding in conservation tasks, and indeed Murray (1968) empirically supported this claim (Braine & Shanks, 1965a, 1965b; however, see Langer & Strauss, 1972, for contradicting results). Although the demands of conservation tasks and AR tasks overlap to a large extent, the typical AR questions are more abstract than the concrete questions about, for example, color or quantity in conservation tasks (see Russell & Mitchell, 1985, for an example). Naturally, these questions are not applicable to nonlinguistic animals. Hence, we cannot conclude that our test required a dual representation of the same object, as in classic AR tasks. Future studies based on nonverbal paradigms will be required to unambiguously distinguish between AR and conservation abilities in animals and children.

The fact that subjects in our study did not transfer their knowledge about the size relation reversal to the control condition allows several conclusions about the underlying mechanism that they used to solve the task. First, they did not have a general awareness or mistrust that appearance can differ from reality, and they also did not acquire such suspiciousness after 48 (apes) or 24 (children) trials. We thus conclude that they did not grasp the general concept of an appearance–reality conflict, as Flavell’s Stage 3 children (11-



to 12-year-olds) would with their explicit knowledge about the AR distinction (Flavell et al., 1986); rather, their choice behavior was driven by and depending on recent perceptual input. However, they were able to rate the real and apparent representation as at least more or less meaningful for their decision (if not as more or less “true”). Great apes and 2½-year-old children thus seem to have some implicit and limited understanding of the appearance-reality distinction, but might lack a fully blown general understanding of this concept.

Overall, one cannot underestimate the importance of the nature of the task for investigating cognitive skills of both children and great apes. Great apes successfully master tasks that are ecologically significant to them, whereas they fail in similar tasks presented in a less relevant context (e.g., Hare & Tomasello, 2004). Young children succeed in nonverbal AR tests, whereas they fail in verbal versions. The introduction of nonverbal methods in false belief research has led to a remarkable shift in the age when children pass the test. While children below 4 years typically fail in verbal versions, nonverbal tasks are mastered already by the age of 2 years and younger (see Baillargeon, Scott, & He, 2010, for a review). As false belief tasks also require subjects to hold a dual representation of the same situation or object, it is plausible that nonverbal AR tests will reduce the passing age, and lead to a new comprehension of how young children perceive AR conflicts. The adverbs “really” and “truly” that are commonly used in the central question of classic AR tests are abstract adverbs that are rarely used in children’s everyday lives. However, children encounter practical, nonverbal AR problems from an early age. Thus, it could well be that the ability to solve practical problems develops earlier than a linguistic AR understanding. According to our study, such a basic, practical AR understanding may be present in 2½-year-olds, and is shared with our closest living primate relatives, the great apes. To which extent the performance in such “really and truly” nonverbal AR tests correlates with pure conservation skills remains a question for future research, but we believe that the current study is a valuable step in this direction.

## References

- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*, 110–118. doi:10.1016/j.tics.2009.12.006
- Beran, M. J. (2001). Summation and numerosness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115*, 181–191. doi:10.1037/0735-7036.115.2.181
- Boysen, S. T. (1996). “More is less”: The elicitation of rule-governed resource distribution in chimpanzees. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 177–189). Cambridge, UK: Cambridge University Press.
- 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. doi:10.1037/0097-7403.21.1.82
- 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. doi:10.1037/0735-7036.115.1.106
- 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/Revue Canadienne de Psychologie, 19*, 197–207. doi:10.1037/h0082903
- Braine, M. D. S., & Shanks, B. L. (1965b). The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior, 4*, 227–242. doi:10.1016/S0022-5371(65)80025-1
- Bruner, J. S. (1966). On the conservation of liquids. In J. S. Bruner, R. R. Older, & P. M. Greenfield (Eds.), *Studies in cognitive growth: A collaboration at the center for cognitive studies* (pp. 183–207). New York, NY: Wiley.
- Deák, G. O. (2006). Do children really confuse appearance and reality? *Trends in Cognitive Sciences, 10*, 546–550. doi:10.1016/j.tics.2006.09.012
- Doré, F. Y., & Dumas, C. (1987). Psychology of animal cognition: Piagetian studies. *Psychological Bulletin, 102*, 219–233. doi:10.1037/0033-2909.102.2.219
- Elkind, D. (1966). Conservation across illusory transformations in young children. *Acta Psychologica, 25*, 389–400. doi:10.1016/0001-6918(66)90020-5
- Flavell, J. H. (1986). The development of children’s knowledge about the appearance-reality distinction. *American Psychologist, 41*, 418–425. doi:10.1037/0003-066X.41.4.418
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1987). Young children’s knowledge about the apparent-real and pretend-real distinctions. *Developmental Psychology, 23*, 816–822. doi:10.1037/0012-1649.23.6.816
- 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*, i–v, 1–87. doi:10.2307/1165866
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1989). Young children’s ability to differentiate appearance-reality and Level 2 perspectives in the tactile modality. *Child Development, 60*, 201–213. doi:10.2307/1131085
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1990). Developmental changes in young children’s knowledge about the mind. *Cognitive Development, 5*, 1–27. doi:10.1016/0885-2014(90)90010-Q
- Flavell, J. H., Zhang, X.-D., Zou, H., Dong, Q., & Qi, S. (1983). A comparison between the development of the appearance-reality distinction in the People’s Republic of China and the United States. *Cognitive Psychology, 15*, 459–466. doi:10.1016/0010-0285(83)90015-4
- Fujita, K. (2012). Amodal completion and illusory perception in birds and primates. In O. F. Lazareva, T. Shimizu, & E. A. Wasserman (Eds.), *How animals see the world: Comparative behavior, biology and evolution of vision* (pp. 101–116). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195334654.003.0008
- Gelman, R. (1972). Logical capacity of very young children: Number invariance rules. *Child Development, 43*, 75–90. doi:10.2307/1127873
- Hanus, D., & Call, J. (2007). Discrete quantity judgments in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus*): The effect of presenting whole sets versus item-by-item. *Journal of Comparative Psychology, 121*, 241–249. doi:10.1037/0735-7036.121.3.241
- Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour, 68*, 571–581. doi:10.1016/j.anbehav.2003.11.011
- Jaakkola, K. (2014). Do animals understand invisible displacement? A critical review. *Journal of Comparative Psychology, 128*, 225–239. doi:10.1037/a0035675
- King, M. (1971). The development of some intention concepts in young children. *Child Development, 42*, 1145–1152. doi:10.2307/1127799
- Krachun, C., Call, J., & Tomasello, M. (2009). Can chimpanzees (*Pan troglodytes*) discriminate appearance from reality? *Cognition, 112*, 435–450. doi:10.1016/j.cognition.2009.06.012
- Langer, J., & Strauss, S. (1972). Appearance, reality, and identity. *Cognition, 1*, 105–128. doi:10.1016/0010-0277(72)90048-0



- Lim, K. H., Benbasat, I., & Ward, L. M. (2000). The role of multimedia in changing first impression bias. *Information Systems Research, 11*, 115–136. doi:10.1287/isre.11.2.115.11776
- McGarrigle, J., & Donaldson, M. (1975). Conservation accidents. *Cognition, 3*, 341–350. doi:10.1016/0010-0277(74)90003-1
- Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development, 82*, 661–673. doi:10.1111/j.1467-8624.2010.01571.x
- Moll, H., & Tomasello, M. (2012). Three-year-olds understand appearance and reality—just not about the same object at the same time. *Developmental Psychology, 48*, 1124–1132. doi:10.1037/a0025915
- Murray, F. B. (1968). Phenomenal-real discrimination and conservation illusion-distorted length. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 22*, 114–121. doi:10.1037/h0082748
- Owen, D. F. (1982). *Camouflage and mimicry*. Chicago, IL: University of Chicago Press.
- Piaget, J. (1952). *The child's conception of number*. London, UK: Routledge.
- Piaget, J. (1954). *The construction of reality in the child*. Oxford, UK: Basic Books. doi:10.1037/11168-000
- Piaget, J. (1961). The genetic approach to the psychology of thought. *Journal of Educational Psychology, 52*, 275–281. doi:10.1037/h0042963
- Piaget, J., & Inhelder, B. (1941). *Le développement des quantités physiques chez l'enfant* [The development of physical quantities in the child]. Neuchâtel, Switzerland: Delachaux et Niestlé.
- Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York, NY: Basic Books.
- Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics, 114*, 37–82. doi:10.1162/003355399555945
- Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28*, 573–594. doi:10.1207/s15326942dn2802\_2
- Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The “windows task” as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*, 331–349. doi:10.1111/j.2044-835X.1991.tb00881.x
- Russell, J., & Mitchell, P. (1985). Things are not always as they seem: The appearance–reality distinction and conservation. *Educational Psychology, 5*, 227–238. doi:10.1080/0144341850050306
- 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. doi:10.1037/0012-1649.36.5.547
- Siegal, M. (1997). *Knowing children: Experiments in conversation and cognition*. Hove, UK: Psychology Press.
- Suda, C., & Call, J. (2005). Piagetian conservation of discrete quantities in bonobos (*Pan paniscus*), chimpanzees (*Pan troglodytes*), and orangutans (*Pongo pygmaeus*). *Animal Cognition, 8*, 220–235. doi:10.1007/s10071-004-0247-6
- Tomasello, M., & Call, J. (1997). *Primate cognition*. New York, NY: Oxford University Press.
- 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. doi:10.1037/0735-7036.122.2.204
- Vlamings, P. H., 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. doi:10.1037/0097-7403.32.1.60
- Wickler, W. (1968). *Mimicry in plants and animals*. New York, NY: McGraw-Hill.
- Woodruff, G., Premack, D., & Kennel, K. (1978). Conservation of liquid and solid quantity by the chimpanzee. *Science, 202*, 991–994. doi:10.1126/science.202.4371.991

Received January 29, 2014

Revision received May 22, 2014

Accepted May 29, 2014 ■