Comparing humans and nonhuman great apes in the broken cloth problem: Is their knowledge causal or perceptual?

Anna Albiach-Serrano a,b,c,⇑, Carla Sebastián-Enesco b, Amanda Seed a,d, Fernando Colmenares b, Josep Call a,d

⇑ Corresponding author at: Unidad de Etología y Bienestar Animal, Facultad de Veterinaria, Universidad CEU Cardenal Herrera, 46115 Valencia, Spain.

Article info
Article history:
Received 27 March 2015
Revised 6 June 2015
Available online 26 June 2015

Keywords:
Children
Chimpanzees
Orangutans
Broken cloth problem
Causal knowledge
Associative learning

Abstract
When presented with the broken cloth problem, both human children and nonhuman great apes prefer to pull a continuous cloth over a discontinuous cloth in order to obtain a desired object resting on top. This has been interpreted as evidence that they preferentially attend to the functionally relevant cues of the task (e.g., presence or absence of a gap along the cloth). However, there is controversy regarding whether great apes’ behavior is underpinned by causal knowledge, involving abstract concepts (e.g., support, connection), or by perceptual knowledge, based on percepts (e.g., contact, continuity). We presented chimpanzees, orangutans, and 2-, 3-, and 4-year-old children with two versions of the broken cloth problem. The Real condition, made with paper strips, could be solved based on either perceptual cues or causal knowledge. The Painted condition, which looked very similar, could be solved only by attending to perceptual cues. All groups mastered the Real condition, in line with previous results. Older children (3- and 4-year-olds) performed significantly better in this condition than all other groups, but the performance of apes and children did not differ sharply, with 2-year-olds and apes obtaining similar results. In contrast, only 4-year-olds solved the Painted condition.

http://dx.doi.org/10.1016/j.jecp.2015.06.004
0022-0965/© 2015 Elsevier Inc. All rights reserved.
We propose causal knowledge to explain the general good performance of apes and humans in the Real condition compared with the Painted condition. In addition, we suggest that symbolic knowledge might account for 4-year-olds’ performance in the Painted condition. Our findings add to the growing literature supporting the idea that learning from arbitrary cues is not a good explanation for the performance of apes and humans on some kinds of physical task.

Introduction

When presented with a desired object that is not directly reachable but rests on a cloth that is within reach, human infants at around 9 or 10 months of age pull the cloth in order to obtain the object (e.g., Piaget, 1952; Willatts, 1984). This is relevant because, according to Piaget, the deliberate and planful execution of subgoals to achieve goals, or “means–end” behavior, marks the emergence of thinking in infants (Piaget, 1952, 1955). Other species have been shown to be able to pull cloths or strings attached to objects, and although some animals required training (e.g., rats: Tolman, 1937), others pulled spontaneously (e.g., chimpanzees; Köhler, 1925) or learned to do so (e.g., dogs; Osthaus, Lea, & Slater, 2005). Surprisingly, there is a paucity of studies that compare the performance of humans and other species in this and other related tasks, although such comparisons are crucial for understanding the evolutionary history of our cognitive traits.

Several versions of the support task have been designed to explore subjects’ knowledge about the functional properties of cloths as tools. In the on–off problem (Hauser, Kralik, & Botto-Mahan, 1999), subjects are presented with two identical objects: one resting on a cloth and the other resting close to another cloth, both at the same distance from the subjects. To pick the cloth that will bring an object, subjects need to pay attention to the presence or absence of contact between the objects and the cloths. Unlike younger infants, 9- or 10-month-olds can solve similar tasks (for reviews, see Willatts, 1984, 1999), as can nonhuman great apes, monkeys, Asian elephants, and some avian species (Auersperg, Gajdon, & Huber, 2009; De Mendonça-Furtado & Ottoni, 2008; Hauser, Santos, Spaepen, & Pearson, 2002; Hauser et al., 1999; Herrmann, Wobber, & Call, 2008; Irie-Sugimoto, Kobayashi, Sato, & Hasegawa, 2008; Povinelli, 2000; Redshaw, 1978; Russell, Lyn, Schaeffer, & Hopkins, 2011; Spinozzi & Poti, 1989, 1993; Yocom & Boysen, 2010). Several nonhuman species have also mastered a related problem, the broken cloth problem (Hauser et al., 1999), where both cloths hold an object but one cloth is broken into two pieces and cannot bring it into reach (Auersperg et al., 2009; Hauser et al., 1999, 2002; Herrmann et al., 2008; Irie-Sugimoto et al., 2008; Schmidt & Cook, 2006). In this case, subjects need to attend to the presence or absence of a gap along the cloths.

Interestingly, in both the on–off and broken cloth problems, there were marked species differences in the number of trials required to attain the solution to the task. Humans and many of the nonhuman apes tested, as well as some of the monkeys and a few keas, solved the tasks within the first 6 to 12 trials. In contrast, most of the monkeys and elephants, one blue-fronted parrot, and four pigeons required more than 100 to solve these tasks (Auersperg et al., 2009; De Mendonça-Furtado & Ottoni, 2008; Irie-Sugimoto et al., 2008; Schmidt & Cook, 2006). Because pulling an object to bring food within reach is a situation that is probably encountered by numerous species (e.g., pulling a branch to get an out-out-reach fruit attached to it), the interspecific differences mentioned above are not easily explained by the ecological relevance of the task alone. Apparently, some primate and bird species preferentially attended to those cues in the task that were functionally relevant (i.e., cues that signaled whether the cloth would bring an object or not: contact between objects and cloths and gap along the cloths) while ignoring those that were not. Other species might not have had such a preference, which would explain the large amount of trials they needed to learn to solve these tasks (Kummer, 1995; Schmidt & Cook, 2006).
Hauser and colleagues (1999) trained tamarins to solve both the on–off and broken cloth problems. After subjects had learned to solve the original tasks, successive changes in features irrelevant to the problems were introduced, for example, changes in the color, shape, or size of the cloths. Tamarins transferred their previous knowledge without much problem, showing that they had attended to the functionally relevant cues and ignored the irrelevant ones. Even better results were obtained by capuchin monkeys in a similar study that included only the on–off problem (Yocom & Boysen, 2010). Based on the same idea, Hauser and colleagues (2002) trained tamarins to solve the on–off and broken cloth problems where the correct and incorrect options were arbitrarily associated with an irrelevant cue, the color of the cloth. Subsequently, this association was reversed. Whereas subjects with no previous experience in tool use tasks failed to transfer their good performances, showing that they had paid attention to color, subjects with some experience in such tasks succeeded, indicating that they had attended to the functionally relevant cues. These studies further show that nonhuman primates (or at least the more experienced ones) preferentially attend to those cues signaling the functionality of the cloth as a way to affect the position of the reward. In particular, subjects might be attending to the existence or absence of contact between the food and the cloth in the on–off problem and between the food and the piece of cloth situated within reach in the broken cloth problem (Herrmann et al., 2008; Povinelli, 2000; Spinozzi & Poti, 1989), as has been found for human infants performing similar tasks (Bates, Carlson-Luden, & Bretherton, 1980). However, from these studies it is unclear whether such preferential attention to functionally relevant cues is based on abstract, causal knowledge or not.

Povinelli (2000) presented chimpanzees with several versions of the on–off problem, where one cloth supported a piece of food (correct option) and the other cloth varied in its relation to the food but in no case was suitable to bring it into reach (incorrect option). Specifically, the wrong cloth could be in contact with the food (partly covering it or being wrapped around it), could be lying right next to it (so that a small manipulation by subjects would put them in contact), or could be separated from it. This way, Povinelli not only explored the importance of contact but also explored whether chimpanzees differentiated the type of contact that is causally relevant from the type of contact that is not. His conclusion was that chimpanzees made no such distinction given that they performed at chance level in several of the conditions conducted, indicating that they could not discriminate a supporting cloth from a cloth touching or surrounding—but not supporting—the food. This supported the hypothesis that, unlike humans, nonhuman primates base their decisions on perceptual cues like contact rather than on abstract concepts like support and force. In other words, nonhuman primates might have their causal knowledge limited to learning about observable stimuli. This would have important implications with regard to primate problem-solving abilities because a subject learning to solve a task based only on perceptual cues necessarily will find more difficulty in transferring the acquired knowledge to new structurally similar but perceptually different problems than a subject apprehending something about the “physical laws” underlying such cues (Hauser & Santos, 2007; Penn & Povinelli, 2007; Seed & Call, 2009). Moreover, if Povinelli’s hypothesis held true, this might account for one of the features traditionally thought to distinguish humans from other species: our outstanding technological capacity.

However, one aspect common to all of the above-described studies is that the solutions to the tasks could come both through attending to perceptual cues (e.g., contact, types of contact), apprehending the relations between objects as arbitrary relations (i.e., noncausal) despite the fact that some objects can physically affect the position of other objects (i.e., perceptual knowledge hypothesis), or from a more abstract understanding of the problem involving knowledge about the physical effect that one object can have on another (i.e., causal knowledge hypothesis) (see Seed & Call, 2010). Although, in contrast to causal inference, associative learning based on perceptual cues often requires repeated exposures to the same problem (Kummer, 1995), it is impossible to establish a critical number of trials distinguishing between the two types of knowledge. In Povinelli’s (2000) study, it could be that chimpanzees, despite being able to reason causally, did not understand what they needed to do because they received only four trials per condition, each in a different session and alternated with “spacer” trials where the goal was to grab a cloth and put it in a bucket without considering its physical relation to any food. Moreover, Povinelli’s chimpanzees actually performed significantly above chance in several conditions, and in no case did they perform below chance. This suggests that either they were...
capable of fast learning of perceptual cues or they had another type of knowledge that allowed them to rapidly solve problems they had never encountered before.

In brief, although humans at 9 or 10 months of age and nonhuman primates can solve several support problems, whether they do so by reasoning causally is an open question that needs to be further explored. This requires an experimental design setting the perceptual knowledge hypothesis against the causal knowledge hypothesis (Hauser & Santos, 2007; Seed & Call, 2009). To this end, we presented human children and nonhuman great apes with two versions of the broken cloth problem. The “real” condition, made with paper strips, could be solved based on either perceptual knowledge or causal knowledge. The “painted” condition, which looked very similar to the real condition, could be solved only by learning from perceptual cues. Therefore, if subjects learned to solve the task based just on perceptual cues (e.g., avoiding the image of the gap), they should perform similarly in both conditions and should be able to transfer the solution from one condition to the next. If, alternatively, they had causal knowledge of the task, they should perform better in the real condition than in the painted condition and should not transfer knowledge from the former to the latter. We were interested in studying both the evolution and ontogeny of causal knowledge. To investigate the former, we compared humans with chimpanzees and orangutans (our closest and furthest relatives within the great apes, respectively). To investigate the latter, we tested 2-, 3-, and 4-year-old children (unfortunately, our sample sizes did not allow for age comparisons in nonhuman primates). All subjects were above the age when they start to be able to solve support problems because, more than knowing whether they could solve the task, we were interested in knowing which information they used.

Method

Subjects/participants

Nonhuman participants were 12 chimpanzees (Pan troglodytes) of both sexes (four males and eight females), 5 to 34 years old (mean age = 15 years), and seven orangutans (Pongo abelii) of both sexes (one male and six females), 6 to 36 years old (mean age = 19 years). Most of them had been reared by their mothers, whereas some of them had been reared by humans (see Table 1). The apes were

Table 1
Nonhuman subjects who participated in the study and their species, sex (M = male, F = female), age (in years), rearing history, experience in string-pulling and support tasks, and group assignment.

<table>
<thead>
<tr>
<th>Group</th>
<th>Species</th>
<th>Subject</th>
<th>Sex</th>
<th>Age</th>
<th>Rearing</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Chimpanzees (Pan troglodytes)</td>
<td>Alex</td>
<td>M</td>
<td>8</td>
<td>Hand</td>
<td>a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alexandra</td>
<td>F</td>
<td>10</td>
<td>Hand</td>
<td>a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pia</td>
<td>F</td>
<td>10</td>
<td>Mother</td>
<td>a,c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Swela</td>
<td>F</td>
<td>14</td>
<td>Mother</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fraukeje</td>
<td>F</td>
<td>23</td>
<td>Hand</td>
<td>a,b,c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robert</td>
<td>M</td>
<td>34</td>
<td>Hand</td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Lobo</td>
<td>M</td>
<td>5</td>
<td>Mother</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tai</td>
<td>F</td>
<td>7</td>
<td>Mother</td>
<td>b,c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annett</td>
<td>F</td>
<td>10</td>
<td>Hand</td>
<td>a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patrick</td>
<td>M</td>
<td>12</td>
<td>Mother</td>
<td>a,c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fifi</td>
<td>F</td>
<td>16</td>
<td>Mother</td>
<td>a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Corrie</td>
<td>F</td>
<td>33</td>
<td>Hand</td>
<td>a</td>
</tr>
<tr>
<td>A</td>
<td>Orangutans (Pongo abelii)</td>
<td>Raaja</td>
<td>F</td>
<td>6</td>
<td>Mother</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kila</td>
<td>F</td>
<td>9</td>
<td>Mother</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dokana</td>
<td>F</td>
<td>20</td>
<td>Mother</td>
<td>a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dunja</td>
<td>F</td>
<td>36</td>
<td>Mother</td>
<td>a,b</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Padana</td>
<td>F</td>
<td>12</td>
<td>Mother</td>
<td>a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pini</td>
<td>F</td>
<td>21</td>
<td>Mother</td>
<td>a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bimbo</td>
<td>M</td>
<td>29</td>
<td>Hand</td>
<td>a,b</td>
</tr>
</tbody>
</table>

*Herrmann and colleagues (2008): 18 trials in a string-pulling task and 18 trials in a support task involving cloths plus, for all subjects except Fifi, Corrie, and Dunja, 18 trials in a support task involving canes.

*Albiach-Serrano and colleagues (2012): 72 trials in a support task involving paper strips.

*Mayer and colleagues (2014): up to 240 trials in a connection task involving broken and complete strings.*
housed at the Wolfgang Köhler Primate Research Center in the Leipzig Zoo (Germany) in social groups inside large enclosures with earth ground, rocks, streams, and natural vegetation. They had trees and other vertical and horizontal structures for climbing and finding shelter, and they had enrichment devices such as artificial termite mounds. Their feeding routines did not change during the study, and water was always available during testing. Subjects were tested individually in special observation rooms that they were free to enter or not. All apes had prior experience with experiments on physical and social cognition, and most of them had some experience with string-pulling or support tasks (see Table 1 for details).

Human participants were 72 children (Homo sapiens) of 2, 3, and 4 years of age (each age group: \( n = 24 \), of which 12 were boys and 12 were girls). The 2-year-old group comprised children from 23 to 29 months, the 3-year-old group comprised children from 32 to 40 months, and the 4-year-old group comprised children from 43 to 53 months. The median ages of the groups were 27.9, 36.5, and 48.7 months, respectively. Children were recruited from kindergartens in Madrid, Spain (see Acknowledgments). There, they were tested individually in a private room. Originally, we received consent from the parents of 30 2-year-olds, 36 3-year-olds, and 29 4-year-olds. However, four 2-year-olds, four 3-year-olds, and one 4-year-old refused to participate or did not complete the test due to illness, and two, eight, and four children of the respective age groups were discarded (attending to their ages, to minimize deviation from the ideal median age of each group: 24, 36, and 48 months) in order to keep the group sizes even. None of the children had previous experience with cognitive experiments.

Subjects of each species, and participants of each age class in the case of children, were divided into two groups. Groups A and B were matched as far as possible for age, gender, and, in the case of apes, rearing history (see Tables 1 and 2 for ape and children data, respectively). Members of Group A received the real condition first followed by the painted condition, whereas members of Group B received the reverse order of conditions. This was true for all of the apes and for 40 of the 72 children (shown in parentheses in Table 2), whereas the remaining 32 children received only one condition (Group A: real; Group B: painted) (see “Procedure” section).

**Materials**

Two types of brown paper strips, continuous (6 × 30 cm) and discontinuous (same measures except for one 1.5-cm gap at 6.5 cm from the further end, which divided the strip into two pieces), could be placed on top of a dark blue plastic panel (78 × 32 cm). A single continuous strip centered on the panel was used in the pretest. Two parallel strips situated 14 cm from each other, one continuous and one discontinuous, were used in the test (Fig. 1A). Other plastic panels had the same strip shapes painted with brown spray (Plastikote 2118/RAL 1019 grey beige) (Fig. 1B). The similar color plus the perimeter thin shadows drawn with pencil made it difficult for the human eye to distinguish painted strips from real strips at first sight. However, we did not want to mislead our subjects into thinking that the painted strips were real ones (because they should be able to discriminate the causal and noncausal structures of each problem). To this effect, the materials were set up in full view of subjects (i.e., subjects did not have their backs turned and were not looking elsewhere).

Fig. 2 shows the experimental setup for apes and children. The task was presented to the apes on a sliding table (80 × 39 cm) fixed to the outside of a mesh window (69 × 48 cm). In the case of children,

<table>
<thead>
<tr>
<th>Age group</th>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
<td>Sex (M/F)</td>
</tr>
<tr>
<td>2-Year-olds</td>
<td>12 (6)</td>
<td>6/6</td>
</tr>
<tr>
<td>3-Year-olds</td>
<td>12 (6)</td>
<td>6/6</td>
</tr>
<tr>
<td>4-Year-olds</td>
<td>12 (6)</td>
<td>6/6</td>
</tr>
</tbody>
</table>

Note. The number of children who, like the apes, did both conditions appears in parentheses, whereas the remaining participants did only one condition (Group A: real; Group B: painted). None of the children had any previous experience in cognitive experiments.
the apparatus lay on a table, and a transparent Plexiglas barrier (82 × 17.5 cm) elevated 2.5 cm prevented them from getting the reward directly while allowing them to access the strips. Banana slices (approximately 8 mm high) were used as rewards for the apes, whereas plastic coins (1 cm high and 4.4 cm in diameter) were used as rewards for the children. This was done to maximize the subjects’ interest in the task because apes are mostly motivated by food, whereas children are highly motivated by toys. Apes could consume the food as soon as they retrieved it. Children could insert the coins into a colorful piggy bank (approximately 22 × 16 × 13 cm) equipped with a device that switched on music for a few seconds. Alternatively, they could keep the coins in a transparent jar in order to use them later. The transparent jar and the piggy bank were kept within reach of the children. At the end of each experimental session, a sticker was given to the children as a reward for having participated in the task.

During testing, the experimenter sat behind the table facing the subjects (see Procedure section). In the case of some 2- and 3-year-old children, a second experimenter sat by the children (slightly behind in order to be out of the children’s sight and prevent any cuing) to help them handle the piggy bank. Other children could handle the piggy bank alone, and with the apes we used food rewards. Therefore, in these cases there was no need for a second experimenter.

**Procedure**

The pretest was run immediately before the first test session. It consisted of six trials where the subject was presented with one single continuous strip with a reward placed on the far end of it (Group A: real; Group B: painted). The aim of the pretest was to ensure that subjects pulled the paper strip (real condition) or touched the painted strip (painted condition) in order to obtain the reward. A trial started with the experimenter placing the panel on the table, in the case of apes in a retracted position to prevent the subject from touching it. In every case, the subject witnessed the setup of the materials, including the placement of the paper strip (in the real condition) and the baiting. Then the experimenter pushed the table forward, in the case of apes, or encouraged children to “get a coin for the piggy.” Children had been shown the functioning of the piggy bank beforehand. No verbal instructions were given, but if subjects did not know how to perform in the pretest the
The experimenter acted as a model by pulling/touching the real/painted strip, after which subjects still needed to successfully complete the six trials of pretest. This procedure was necessary only for one orangutan (in the painted pretest), nine 2-year-olds (four in the real pretest and five in the painted pretest), and two 3-year-olds (one in each condition).

The test was similar to the pretest except that there were two baited strips: one continuous and one discontinuous. To retrieve a reward, subjects needed to choose the continuous strip. Pulling or touching a paper strip and touching a painted strip were considered as choices. When the continuous strip was pulled, subjects could attain the reward on their own. When the continuous strip was touched, the experimenter gave the reward placed on that same strip. After an incorrect choice, the experimenter pulled the apparatus back to avoid second choices. The experimenter never acted as a model in the test. Apes in Group A received the real condition first and then the painted condition, whereas apes in Group B received the painted condition first and then the real condition. Like the apes, some children \((n = 40)\) received both conditions, whereas other children \((n = 32)\) only underwent either the real condition (Group A) or the painted condition (Group B) (see Table 2). In addition, each condition consisted of 36 trials divided into three sessions of 12 trials for the apes and two sessions of 18 trials for the children. The reason for this procedural difference between the apes and the children was that a pilot study with children (carried out after testing with the apes had started) showed that some of the younger children had limited motivation to return for multiple sessions. Consequently, the first children tested received only one of the two conditions in two sessions \((n = 32)\). As the study progressed, however, it became clear that it was possible to test children in both conditions \((n = 40)\). The position of the discontinuous strip (left/right) was counterbalanced within sessions and randomly presented, with the restriction that it could not be the same in more than two consecutive trials. Different sessions were held on different days, consecutive as far as possible.

Scoring and data analysis

Responses were noted in situ using data sheets, and all sessions were videotaped. The first author tested the apes, whereas the second author and one collaborator tested the children (see Acknowledgments). We randomly selected 20% of the ape and children sessions, and a second observer coded them to assess interobserver reliability, which was excellent (Cohen’s kappas: apes, .97, \(p < .01\); children tested in both conditions, .96, \(p < .01\); children tested in one condition, .97, \(p < .01\)). We used SPSS to analyze the data. Some of our variables did not meet the normality and homogeneity assumptions for parametric tests; therefore, we used nonparametric tests. Given our small sample sizes, we used exact probability tests in all of the analyses (Mundry & Fischer, 1998). All tests were two-tailed unless stated otherwise.

Our main dependent measure was the percentage of trials in which subjects selected the correct alternative, whereas our independent variables were group, rearing history (mother vs. hand), condition (real vs. painted), and condition order (first vs. second). The data were analyzed in three steps. First, for each group we analyzed the effect of each of the independent variables on the percentage of correct responses, and we tested whether subjects responded at above chance levels in each condition. Second, for each group we analyzed performance on the first trial of the first condition received, and we subsequently assessed learning by calculating the correlation of performance across trials. Third, we compared groups’ performance in each condition and classified subjects in each group according to their success, both overall and in the first 12 trials of each condition.

Results

Pretest

Both apes and children spontaneously pulled a paper strip and readily learned to touch a painted strip in order to obtain a reward placed on top. Table 3 shows the mean number of trials needed to succeed in the pretest by 2-, 3-, and 4-year-olds, with the criterion of success being the correct performance in two consecutive trials within the six trials given. The data are based on 40 of 72 children
who participated in the study. Although we are missing detailed data on the remaining children and apes in the pretest, their performance was similar (if anything, the apes solved the pretest faster than the children given that the mesh prevented them from trying to reach the reward directly and, unlike the children, they generally engaged in manipulative behaviors as soon as the apparatus was within reach without expecting further instructions from the experimenter).

Order of presentation

Table 4 shows the groups’ performance in both conditions of the test as a function of the order of presentation. There were no significant differences in performance between those apes that received the real condition first and those that received it second (Mann–Whitney tests [one-tailed]: chimpanzees, $U = 15.5, n = 12, p = .36$; orangutans, $U = 4, n = 7, p = .31$). Similarly, the order of presentation had no significant effect on the painted condition for the apes (Mann–Whitney tests [one-tailed]: chimpanzees, $U = 15.5, n = 12, p = .36$; orangutans, $U = 4, n = 7, p = .31$).

Children’s performance in the real condition was also unaffected by the order of presentation (2-year-olds: $U = 20, n = 16, p = .15$; 3-year-olds: $U = 12, n = 12, p = .19$; 4-year-olds: $U = 13, n = 12, p = .24$). In contrast, 2- and 4-year-olds, but not 3-year-olds, performed better in the painted condition when they received it after the real condition (2-year-olds: $U = 8, n = 16, p < .01$; 3-year-olds: $U = 8.5, n = 12, p = .07$; 4-year-olds: $U = 4.5, n = 12, p = .02$). To guard against this order effect, subsequent analyses of the children’s data are based only on the first condition that they received.

Real versus painted conditions

Fig. 3 presents the overall percentages of correct responses in each test condition for all groups (for raw frequencies, see Table S1 in the online supplementary material). Apes performed significantly
better in the real condition than in the painted condition (Wilcoxon tests: chimpanzees, $T = 66$, $n = 11$, $p < .01$; orangutans, $T = 28$, $n = 7$, $p = .02$). Apes were above chance in the real condition (Wilcoxon tests: chimpanzees, $T = 3$, $n = 12$, $p < .01$; orangutans, $T = 0$, $n = 7$, $p = .02$) and at chance levels in the painted condition (Wilcoxon tests: chimpanzees, $T = 42$, $n = 10$, $p = .16$; orangutans, $T = 4$, $n = 4$, $p = 1.00$). These results were similar for those apes that were raised by their mothers and for those that were hand-raised (Mann–Whitney tests: real, $U = 23.5$, $n = 19$, $p = .12$; painted: $U = 26.5$, $n = 19$, $p = .19$). Therefore, we do not consider rearing history in further analyses.

Like the apes, the children in all age groups performed significantly better in the real condition than in the painted condition (Mann–Whitney tests: 2-year-olds, $U = 7.5$, $n = 24$, $p < .01$; 3-year-olds, $U = 19$, $n = 24$, $p < .01$; 4-year-olds, $U = 37$, $n = 24$, $p = .04$). Furthermore, they performed above chance in the real condition (2-year-olds: $T = 0$, $n = 12$, $p < .01$; 3-year-olds: $T = 0$, $n = 12$, $p < .01$; 4-year-olds: $T = 0$, $n = 12$, $p < .01$) but at chance levels in the painted condition, except for the 4-year-olds, who were also above chance (Wilcoxon tests: 2-year-olds, $T = 17$, $n = 9$, $p = .56$; 3-year-olds, $T = 19.5$, $n = 11$, $p = .24$; 4-year-olds, $T = 0$, $n = 12$, $p < .01$).

**First trial performance and learning over trials**

Table 5 presents first trial responses in the first condition received for all groups. Binomial tests showed that chimpanzees and children performed at chance levels on their first trial in the first condition received, including both the real and painted conditions ($p > .15$). Two exceptions were 2-year-old children, who picked the wrong option (i.e., the discontinuous strip) above chance on their first trial in the painted condition ($n = 12$, $p < .05$), and 3-year-old children, who picked the correct option (i.e., the continuous strip) above chance on their first trial in the real condition ($n = 12$, $p < .05$). First trial performance could not be tested in orangutans because their small sample sizes ($n < 6$) made it impossible to reject the null hypothesis and, therefore, rendered the tests uninformative (Siegel & Castellan, 1988).

Figs. 4 and 5 show the performances across trials in the real and painted conditions for apes and children. Table S2 in the supplementary material shows the means and variances of each data point for all groups. Orangutans, but not chimpanzees, showed clear evidence of learning in the real condition (Spearman $r$: chimpanzees—real, $r = .29$, $p = .09$; painted, $r = -.22$, $p = .21$; orangutans—real, $r = .50$, $p < .01$; painted, $r = .02$, $p = .92$). We found evidence of learning in the real condition but not in the painted condition in all child age groups (Spearman $r$: 2-year-olds—real, $r = .56$, $p < .01$; painted, $r = .13$, $p = .44$; 3-year-olds—real, $r = .48$, $p < .01$; painted, $r = .20$, $p = .24$; 4-year-olds—real, $r = .45$, $p < .01$; painted, $r = .11$, $p = .54$).
Group comparisons

Focusing on the first condition received by subjects, we found group differences both in the real condition (Kruskal–Wallis test: $\chi^2(4, n = 46) = 25.96, p < .01$) and in the painted condition (Kruskal–Wallis test: $\chi^2(4, n = 45) = 20.05, p < .01$). Pairwise comparisons with the Bonferroni–Holm correction (Holm, 1979) revealed that in the real condition chimpanzees performed worse than 3- and 4-year-old children (Mann–Whitney tests: 3-year-olds, $U = 1, n = 18, p < .01$; 4-year-olds, $U = 0, n = 18, p < .01$), as did orangutans ($U = 0, n = 16, p < .01$ in both cases) and 2-year-old children (3-year-olds: $U = 28.5, n = 24, p < .01$; 4-year-olds, $U = 23, n = 24, p < .01$). All other comparisons were nonsignificant.

### Table 5

Number of subjects of each group who chose the correct and incorrect options (i.e., continuous and broken strips, respectively) on their first trial in the first condition received.

<table>
<thead>
<tr>
<th>Group</th>
<th>Real condition</th>
<th>Painted condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous strip</td>
<td>Broken strip</td>
</tr>
<tr>
<td>Chimpanzees</td>
<td>12 4 2 6</td>
<td>4 2 6</td>
</tr>
<tr>
<td>Orangutans</td>
<td>7 2 2 4</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Children (2 years)</td>
<td>24 7 5 12</td>
<td>2 10 12</td>
</tr>
<tr>
<td>Children (3 years)</td>
<td>24 10 2 12</td>
<td>4 8 12</td>
</tr>
<tr>
<td>Children (4 years)</td>
<td>24 9 3 12</td>
<td>5 7 12</td>
</tr>
</tbody>
</table>

Note. Significant results are marked with bold numbers (binomial test, $p < .05$) (see text).

![Fig. 4.](image)

Mean percentages of correct responses for chimpanzees (top) and orangutans (bottom) in the real and painted conditions as a function of trial number. Both the first and second conditions received by each ape are included.

**Group comparisons**

Focusing on the first condition received by subjects, we found group differences both in the real condition (Kruskal–Wallis test: $\chi^2(4, n = 46) = 25.96, p < .01$) and in the painted condition (Kruskal–Wallis test: $\chi^2(4, n = 45) = 20.05, p < .01$). Pairwise comparisons with the Bonferroni–Holm correction (Holm, 1979) revealed that in the real condition chimpanzees performed worse than 3- and 4-year-old children (Mann–Whitney tests: 3-year-olds, $U = 1, n = 18, p < .01$; 4-year-olds, $U = 0, n = 18, p < .01$), as did orangutans ($U = 0, n = 16, p < .01$ in both cases) and 2-year-old children (3-year-olds: $U = 28.5, n = 24, p < .01$; 4-year-olds, $U = 23, n = 24, p < .01$). All other comparisons were nonsignificant.
(\(p > .11\) in all cases except the comparison between chimpanzees and 2-year-olds where \(p = .02\), but after the Bonferroni–Holm correction it was nonsignificant). In the painted condition, 4-year-olds outperformed all of the other groups except 3-year-olds (chimpanzees: \(U = 0.5, n = 18, p < .01\); orangutans: \(U = 0, n = 15, p < .10\); 2-year-olds: \(U = 4, n = 24, p < .01\); 3-year-olds: \(U = 28, n = 24, p < .01\)). We found no other differences between groups (\(p > .24\) in all cases).

Table 6 shows the numbers of subjects in each group that solved each condition at above chance levels (binomial test, \(p < .05\)) both overall and on the first 12 trials (equivalent to the first session in nonhuman apes and to two thirds of the first session in children). Some apes and most children solved

---

**Fig. 5.** Mean percentages of correct responses for 2-, 3-, and 4-year-old children in the real and painted conditions as a function of trial number. Only the first condition received by each child is included.
the real condition, and members of all groups did so from the first trials, with 3- and 4-year-old children clearly outperforming the rest. In contrast, none of the apes solved the painted condition. Among the children, one 2-year-old child did, but not from the beginning, four 3-year-olds solved it as well, three of them from the beginning, and most 4-year-olds mastered the task, also from the beginning.

Discussion

In the current study, we compared the cognitive skills of three hominoid species—chimpanzees, orangutans, and humans—using exactly the same paradigm: the broken cloth problem. All of them preferentially pulled the continuous strip, thereby obtaining a reward in most trials. This confirms previous results with apes and provides evidence that young children can also master this task. Because the only difference between the continuous and discontinuous strips was the presence or absence of a gap along it, it is highly probable that subjects attended to this specific, functionally relevant feature in order to make their choices. By comparing a real version with a painted version of the problem, we further explored whether the information used was purely perceptual or rather abstract.

Within the perceptual knowledge hypothesis (Cohen & Cashon, 2006; Hauser & Santos, 2007; Penn & Povinelli, 2007; Seed & Call, 2009), one potential explanation to the solution of the broken cloth problem would be that subjects had a predisposition to avoid the (image of a) gap. The fact that most chimpanzees, orangutans, and 2-year-olds did not show any initial preference for the continuous strip, however, makes this possibility unlikely (in fact, 2-year-olds seemed to prefer the discontinuous strip on their first trial in the painted condition). Another explanation would be that subjects learned associatively to avoid the gap due to the reinforcement regimes of the task. Although this could certainly account for the lack of immediate success and the improvement across trials in apes and children in the real condition, it would not explain the poor results obtained by these same groups (except for the older children) in the painted condition, especially when this was the first condition encountered by subjects.

Overall, subjects performed better in the real condition than in the painted condition despite the fact that both conditions looked almost identical and that similar contingencies were applied in both cases (i.e., picking the continuous strip provided the reward resting on top). When the second condition was removed from the analyses (to avoid order effects), all groups solved the real condition and no group (except for the 4-year-olds) solved the painted condition. In fact, whereas most of our subjects solved the former, no chimpanzee or orangutan, and only one 2-year-old and four 3-year-olds (4-year-olds aside), solved the latter, suggesting that the perceptual cues provided were not sufficient for them to learn in the amount of trials given. This is further supported by the fact that although all groups improved their performance throughout trials in the real condition (for the chimpanzees this was nonsignificant, possibly due to the need for more trials), none did so in the painted condition. This general failure in the painted condition is surprising if we consider that there were only two options to choose from (the continuous and discontinuous strips) and always the same choice was rewarded (the continuous strip). Moreover, when the apes went from the real condition to the painted condition, they did not transfer their good performance to the new task. This means that in the real condition

<table>
<thead>
<tr>
<th>Species</th>
<th>Real</th>
<th>Subjects</th>
<th>Painted</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Subjects</td>
<td>n</td>
<td>Subjects</td>
</tr>
<tr>
<td>Chimpanzees</td>
<td>12</td>
<td>6 (1)</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Orangutans</td>
<td>7</td>
<td>5 (2)</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Children (2 years)</td>
<td>12</td>
<td>11 (4)</td>
<td>12</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Children (3 years)</td>
<td>12</td>
<td>12 (12)</td>
<td>12</td>
<td>4 (3)</td>
</tr>
<tr>
<td>Children (4 years)</td>
<td>12</td>
<td>12 (12)</td>
<td>12</td>
<td>11 (10)</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>46 (31)</td>
<td>55</td>
<td>16 (13)</td>
</tr>
</tbody>
</table>

Note. Numbers in parentheses indicate those subjects who solved the task within the first 12 trials (equivalent to the first session for apes and the first two thirds of the first session for children).
they had not learned just to avoid the image of a gap or to select that of a continuous strip. Otherwise, they could have continued to do the same and, as a consequence, would have solved the painted condition as well. Instead, they seemed to have attended to the presence or absence of a gap only when this feature was causally meaningful. In other words, although apes are demonstrably capable of learning arbitrary associations, in this task with almost identical perceptual features, learning was significantly faster when the discriminative cues had functional relevance to the outcome than when there was no such relevance.

A different picture was found in the children, who at least partially transferred information from the real condition to the painted condition (as shown by an order effect in 2- and 4-year-olds), and especially in the 4-year-olds, who performed in the painted condition as correctly as in the real condition, in both cases from the first trials. Yet, the worse performance of the younger children in the painted condition suggests that, although leading to similar results, the cognitive mechanisms used by the 4-year-olds to solve the real and painted conditions were not the same. As proposed for the apes, the solution to the real condition in children could have come from subjects acquiring some abstract knowledge (i.e., detached from purely perceptual cues) about the physical properties of strips and baits, about their relations to each other, and about their potential reactions under different types of force. This way, unlike any other perceptual feature of the task, the gap would become relevant due to its causal meaning, thereby allowing subjects to make predictions about the outcomes of their own actions on the paper strips. With regard to the 4-year-olds’ success in the painted condition, one possible explanation could be that associative learning skills improve ostensibly in humans at around this age. This would explain why the 4-year-olds mastered the painted condition after a very small number of trials, whereas only a few 3-year-olds solved it, after more trials, and nearly all 2-year-olds failed, even after 36 trials. However, there is an alternative explanation.

Our results resemble those of a previous study where 2½- to 5½-year-old children proved to be able to obtain a reward from a transparent puzzle box—with its mechanism visible—whereas only 6½-year-olds solved an opaque version of this task, where stickers depicted the barriers and traps contained in the box (Seed & Call, 2014). The authors proposed that the older children might have interpreted the arbitrary perceptual cues of the task as symbolic cues, which would have allowed them to master the problem. The 4-year-olds in our study might also have succeeded in the painted condition due to symbolic thought. In fact, unlike the apes, children who had previous experience with paper strips performed better with painted strips (at least the 2- and 4-year-olds did so), which means that they transferred information from the real condition to the painted condition (not the other way around). One possibility is that they took the painted strips as representations of paper strips. However, those 4-year-olds who did not have experience with paper strips also performed well in the painted task. It might be that they gave the meaning of “wrong” and “correct” to the discontinuous and continuous strips, respectively. Indeed, it is at around 4 or 5 years of age that children start understanding the communicative dimension of artifacts (DeLoache, 2000; Tomasello, 1999). Even though in the current study children were not given verbal instructions, it could be that 4-year-olds were able to infer the implicit meaning of the painted strips from the rewarding regimes applied by the experimenter. We believe that this explanation is currently the best way to account for the difference between 4-year-olds, on the one hand, and younger children and apes, on the other. The growing competence with symbols could emerge either because of a change in underlying cognitive structures over development or because of greater experience with symbols and communication. Future work is needed to disambiguate these alternatives. One possible avenue would be to look at the performance of symbol-trained apes.

In sum, all of the species tested solved the broken cloth problem and showed some causal knowledge of the task. This contrasts with the interpretation of Povinelli (2000) that we outlined in the Introduction from the negative results he reported in a similar task with chimpanzees. Basically, he used a very small number of trials (4, whereas we used 36 trials) that, in addition, were intermixed with spacer trials where subjects did not need to consider the functionality of the cloth. Many of Povinelli’s transfer tasks made use of subtle differences between the right and wrong alternatives, which may have placed higher demands on attention and memory (Seed, Seddon, Greene, & Call, 2012). Sample effects, including individual differences, age differences, and different rearing conditions, might also explain the discrepancy in our results.
Nevertheless, in our study we found overall differences in the performances of apes and children in the broken cloth task. Although we tried to make the methods as comparable as possible for all groups, in terms of both motivation elicited and information provided, we cannot rule out the possibility that the use of different types of reward for apes and children (food and toys, respectively) may have affected our results (see Boysen & Berntson, 1995, for an example of how the use of food rewards hindered chimpanzees’ performance in a numerical task). In addition, unintentional cues given by the experimenter may have affected each species differently. Future studies should consider these questions. Some researchers have questioned the practice of comparing apes with children due to their different rearing conditions (e.g., Boesch, 2007, and the response by Tomasello & Call, 2008). Although rearing conditions can affect apes’ performance in cognitive tests (e.g., Bering, 2004; Call & Tomasello, 1996), we found no significant differences in performance between mother-reared apes and human-reared apes in the current study. This result fits well with Russell and colleagues (2011), who found that enculturated and nonenculturated apes performed similarly in support tasks, although they differed in other tasks. It is an open question how children reared without cultural scaffolding or wild chimpanzees would perform in support tasks.

Although the general performance of children was better than that of apes, this boundary was not sharp given that apes did not differ from 2-year-olds in the real condition. In contrast, only the children were able to solve the task just from the visual cues provided. The fact that causal knowledge is shared by all three species and appears early in human ontogeny seems to make evolutionary sense given that highly manipulative, extractive foraging species must benefit from recognizing the causal properties of objects and the causal structure of events (Seed & Call, 2009). In contrast, the fact that 36 trials in a two-option choice task did not suffice for apes and many younger children to learn to solve the painted condition suggests that learning based on arbitrary cues might be harder for great apes than previously thought (Albiach-Serrano, Bugnyar, & Call, 2012; Call, 2006; Hanus & Call, 2008; Hanus & Call, 2011). Such ability seems to be relevant for humans because it allows us to find fast solutions to otherwise opaque problems such as pressing the “on” button of our computer when we want to start it without the need to understand how it is that this works. This could be achieved both by simply associating a feature, X, with an outcome, Y, or by learning or inferring that X causes Y, without needing to encode the precise underlying mechanism. In any case, the age of the only children who could easily solve the painted condition, 4 years, suggests that symbolic thought might be playing a role.

Finally, although few people would doubt that children understood something about the causal relation of strips and rewards in the current study, a skeptic might argue that differential learning in the real and painted conditions did not necessarily show causal knowledge in apes. Their better results in the real condition could alternatively be explained by differences in the perceptual saliency of the two types of strips (i.e., saliency effect; Albiach-Serrano et al., 2012). That is, even though real and painted strips looked very similar, subjects might have perceived real strips as more salient from the tactile point of view because they had a rougher texture, they could be moved, and pulling them transmitted some feeling of resistance, as opposed to painted strips. This could have resulted in the apes attending more in the real condition and, as a consequence, being better able to learn from visual cues in this task (note that at the time of choosing the only cues available are visual). This would not explain, however, why our apes did not transfer their knowledge from the real condition to the painted condition. Moreover, recent studies have shown that the saliency of paper strips or strings does not help chimpanzees to learn from visual cues when the task is stripped of the causal relation between actions and outcomes (Albiach-Serrano & Call, 2014; Mayer et al., 2014). Taken together, our data suggest that causal knowledge rather than perceptual knowledge might be behind nonhuman apes’ tool use behavior and that symbolic knowledge might rather be the key difference with humans. We believe that this is a question worthy of further research that would greatly benefit from a wider use of the comparative approach.

Acknowledgments

We thank La Navata, Pequecampus, San Gabriel, and Santa María de la Hispanidad (Madrid, Spain) for providing us with the opportunity to obtain participants for this study. In addition, we thank the
parents for allowing us to test their children. We are grateful to Esther García Medina, Mónica Terrazo-Felipe, and Natalia Hortigüela for helping in the collection of data with young children. C. S. ebastían-Enesco’s and F. Colmenares’s contribution was partly supported by Project Grants UCM-BSCH GR58/08, GR35/10-A-940813, and PSI2011-29016-C02-01 to F. Colmenares. Finally, we appreciate the work of the keepers from Pongoland (Leipzig Zoo, Germany) who provided help with the apes.

Appendix A. Supplementary data
Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jecp.2015.06.004.

References