Great apes' capacities to recognize relational similarity

Daniel B.M. Haun [a,b,*], Josep Call [a]

[a] Max Planck Institute for Evolutionary Anthropology, Research Group for Comparative Cognitive Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[b] University of Portsmouth, Department of Psychology, King Henry Building, King Henry I Street, P01 2DY Portsmouth, UK

ARTICLE INFO

ABSTRACT

Recognizing relational similarity relies on the ability to understand that defining object properties might not lie in the objects individually, but in the relations of the properties of various object to each other. This aptitude is highly relevant for many important human skills such as language, reasoning, categorization and understanding analogy and metaphor. In the current study, we investigated the ability to recognize relational similarities by testing five species of great apes, including human children in a spatial task. We found that all species performed better if related elements are connected by logico-causal as opposed to non-causal relations. Further, we find that only children above 4 years of age, bonobos and chimpanzees, unlike younger children, gorillas and orangutans display some mastery of reasoning by non-causal relational similarity. We conclude that recognizing relational similarity is not in its entirety unique to the human species. The lack of a capability for language does not prohibit recognition of simple relational similarities. The data are discussed in the light of the phylogenetic tree of relatedness of the great apes.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Without much deliberation humans appreciate structural similarities between a treasure map and an island or between a marriage and an industrial joint venture. The underlying cognitive process is the recognition of relational similarity. One way to think about the recognition of relational similarity is as the ability to understand that defining object properties might not lie in the objects individually, but in the relations of the properties of various object to each other, and that furthermore, objects with common relations amongst their respective surroundings are similar to each other. This aptitude, in various disguises, is at the basis of many important human skills such as for example understanding propositional structure and predication (Tomasello, 2003), inductive inference (Holland, Holyoak, Nisbett, & Thagard, 1986), categorization

(Ramscar & Pain, 1996) and understanding analogy and metaphor (Gentner, 2003; Tomasello, 2003). Despite its relevance to human thought, children acquire proficiency in relational similarity tasks rather late. They start out relying on overall (mostly perceptual) commonalities to judge similarity between objects, but only shift to appreciating relational similarity, depending on the particular task, as late as 4 years (Blades & Cooke, 1994; Gentner, 2003; Loewenstein & Gentner, 2001, 2005; Rattermann & Gentner, 1998; Smith, 1984) or even 9–10 years of age (Gentner & Toupin, 1986). The apparent difficulty of acquiring relational concepts does not only show in cognitive development, but also in language acquisition. Children as old as 3;6 years have great difficulty grasping the meaning of relational nouns denoting concepts like "passenger" (Hall & Waxman, 1993) or spatial relational constructions such as "the ball is left of the tree" (Brown & Levinson, 2000). Further it is worth noting that in this latter example and many other circumstances the *cognitive relational shift* (Bowdle & Gentner, 2005; Quine, 1960) is often accompanied by the acquisition of the relevant linguistic vocabulary (Gentner, 2003). Further, it has been reported that preschool children were better able to carry out a

* Corresponding author. Address: Max Planck Institute for Evolutionary Anthropology, Research Group for Comparative Cognitive Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. Tel.: +49 0 341 3550815; fax: +49 0 341 3550444.
E-mail address: haun@eva.mpg.de (D.B.M. Haun).

challenging spatial relational similarity task when the experimenter used spatial relational terms (such as *top middle bottom*) to describe parts of the scene, even when retested days later, without any reinstatement of the spatial terms (Loewenstein & Gentner, 2005). The tightly interweaved developmental trajectory of relational language and the ability to recognize relational similarity and the beneficial effect of spatial relational language on children's performance in spatial relational tasks has led some researchers to believe relational language and the ability to recognize relational similarity to be correlated and most likely even causally related. (Gentner, 2003; Loewenstein & Gentner, 2005; Rattermann & Gentner, 1998). Furthermore, cross-linguistic differences in spatial relational language and parallel differences in processing-preferences for spatial relations have served as further indicators for the interaction between relational language and thought (Bowerman & Choi, 2001; Haun, Rapold, Call, Janzen, & Levinson, 2006; Majid, Bowerman, Kita, Haun, & Levinson, 2004). None of these authors has stated that the ability for language was necessary for recognizing relational similarity, however, the late acquisition in children, the interrelation with language and the central function in so many human cognitive abilities has led researchers to propose that the ability to recognize relational similarity is extraordinarily pronounced in humans (Oden, Thompson, & Premack, 2001; Penn, Holyoak, & Povinelli, 2008; Premack, 1983; Thompson & Oden,1995, 2000) if not even the one thing that "makes us smart" (Gentner, 2003).

The most insightful results about which cognitive abilities are unique (or better derived) in humans is to situate human cognition in its phylogenetic context by comparison with other animal species, especially closely related ones such as non-human primates. The question if non-human primates have the ability to recognize relational similarities has been controversial (Penn et al., 2008; Tomasello & Call, 1997). Most of the research involved with this discussion is based on so-called match-to-sample paradigms (MTS) in which a sample stimulus is presented either simultaneously or successively with two comparison stimuli, a correct and an incorrect match. Subjects have to choose between the two alternatives. This paradigm can be used to investigate either simple first order relations such as perceptual similarity or more complex second order relations in which the sample are for example two identical shapes and the match are two other but also identical shapes. The latter is an example of matching by relational similarity. Based on variations of this paradigm, some researchers have suggested that symbol-training can enhance the ability to recognize relational similarities in apes (Oden, Thompson, & Premack, 1988; Premack, 1983; Thompson & Oden, 1995, 2000). Indeed, untrained chimpanzees displayed evidence of understanding similarities and differences between relations in the looking and handling of objects despite failing the previous MTS task (Oden, Thompson, & Premack, 1990). Other authors documented the understanding of object-similarity based on function, a simple form of relational similarity, in chimpanzees (Spinozzi, 1993; Tanaka, 1996) as well as olive baboons (Bovet & Vauclair, 2001). Fagot, Wasserman, and Young (2001) even reported full-blown understanding of

relational similarity in baboons. In this study, animals matched either two pictures displaying arrays of identical objects or two pictures displaying arrays of different objects. However the similarities and differences between the stimuli were partly perceptual (all-same vs. all different) and baboons required thousands of training trials to reach criterion on the training set, hence baboons do not seem to spontaneously perceive such relations and it is still unclear if they form concepts about them beyond perception. Recently, Vonk (2003) reported that gorillas and orangutans succeeded spontaneously in a computerized relational MTS task. These animals were previously trained to solve identity match-to-sample problems on a computer screen and transferred successfully to a relational problem.

In summary there are two sets of problems with the current state of affairs concerning the issue of relational similarity in non-human primates: an excessive reliance on a single experimental paradigm and a lack of a comparative focus. Regarding the first problem, almost all studies are based on the visual MTS paradigm and they produced contradictory results through the years. Moreover, visual MTS tasks are always slightly confounded by perceptual similarity in the sense that pairs or arrays of identical things always have a different perceptual texture (or relative entropy) than pairs or arrays of different things. Even though in smaller arrays of objects, similarities in relative entropy between arrays are less prominent than in large arrays, they are not absent (Penn et al., 2008). Even if researchers attempt to control for this problem by increasing the perceptual similarity between the sample and the non-match (e.g. Vonk, 2003) the perceptual similarity between the sample and the match persists irrespectively and might serve as a basis for the animals' decision making. Hence, with any visual MTS paradigm we can never be certain to actually test relational similarity proper. Any outcome will heavily depend on the exact visual features of the stimuli, the reliance of the animals on visual cues and the perceptual contingencies and patterns in the (necessary) training phase. Therefore, the MTS paradigm might be one of the reasons why results have not produced a clear picture up to this point. Further, MTS requires extensive training in all non-human animals and hence involves a lot of effort and motivation on side of the animal and the researcher. As a result, MTS studies are hard to replicate which is reflected in the relatively small amount of studies on the topic, often repeatedly conducted with the same few animals.

The second problem is that most previous so-called comparative studies on relational similarity are not truly comparative, since they only investigate a single species and humans have not been directly compared to non-human primates using the same methods. Arguably, a single skilled species is potentially enough to prove a cognitive skill not to be uniquely human but without a detailed comparison based on multiple paradigms and multiple species it is unclear to what extent human and non-human animals share the ability to recognize relational similarity. This 'double' comparative approach can produce the answer to questions regarding the kinds of relational similarity problems that are human specialties or what has been the evolutionary history of these abilities in primates.

Studies of single species cannot produce answers to these questions.

In the current study, we attempted to solve these two sets of problems. With regard to the question of proper species comparisons, a recent set of publications outlines a powerful approach, which compares behaviour of modern primates in conjunction with an accurate phylogenetic tree of relatedness to chart early evolutionary history of cognition and the unique features of human cognition amongst primates (Byrne, 1995, 2000; Haun, Call, Janzen, & Levinson, 2006; Haun, Rapold et al., 2006; Nunn & Barton, 2001).

Using this comparative approach [also sometimes called 'cognitive cladistics', (Haun, Rapold et al., 2006)], we may find some cognitive skills shared across all members of a phylogenetic family, in our case the great apes (Hominidae: orangutans, gorillas, bonobos, chimpanzees and humans), while others will only occur in a subset of these species. Where patterns of skill distribution match the known phylogenetic tree of relatedness, we will be able to trace the phylogenetic inheritance of these traits, thus gaining access to the likely evolutionary history of the relevant abilities. For example, features shared only between humans and chimpanzees are likely to be approximately 6 million years old, those shared between chimpanzees, humans and gorillas, approximately 10 million years old, and so forth. Finally, if all species of a phylogenetic family exhibit the same behavioural tendencies or cognitive biases, this suggests inheritance from the common ancestor shared by all species. This information will be extremely valuable for understanding the gradual accumulation of cognitive traits in our phylogenetic family. Thus, in the current study we compared human children between 3;6 and 4;6 years of age and the other four great ape species.

With regard to the question of the experimental paradigm, we adopted a non-linguistic spatial relational similarity paradigm that has previously been used to investigate human children (Loewenstein & Gentner, 2005), diverse human cultures and non-human great apes (Haun, Rapold et al., 2006). Understanding the basic structure of our setup relies on logico-causal cognition at which all tested great apes species have been found to be highly proficient (Tomasello & Call, 1997; Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007). All great apes including humans show similar ontogenetic development across the Piagetian stages of object permanence up to stage 6 (Antinucci, 1989) and can cope with a similar variety of object displacements (Beran, Beran, & Menzel, 2005; Call, 2004; Barth & Call, 2006). Hence paradigms based on object displacement do not require a complicated setup and can most likely be run with non-human primates with little to no previous training. Our task consisted of two sets of three identical cups arranged in two similar constellations, in a way that every cup in its spatial position relative to the other objects within one constellation has a corresponding cup with the same relative position in the other constellation. The relations are now spatial relations and these might be similar or different between pairs of cups. Since all cups are identical in all conditions, perceptual similarity of objects cannot confound the data. The arrangement of the cups allowed us to distinguish spatial

relational similarity from other strategies such as alignment or proximity that subjects could use to solve a spatial mapping problem (see below).

To further investigate the effect of additional cues on the subjects' performance, we superimposed on the basic spatial setup an additional set of stimuli that physically linked the correct pairs of cups across sets. In one task the cups occupying the same relative position were linked by a tube (TUBES task) while in the other condition they were linked by a strip of tape (LINES task) equal in width to tubes. Although both links were perceptually very similar (see Fig. 1), they differed critically in their logico-causal properties. While the tubes could enable the unseen passage of the reward from one cup to the other, the strip of tape did not afford this possibility. The TUBES task therefore presents an invisible displacement problem the LINES task poses a highly comparable relational problem. Hence our main focus in the present study will lie on the LINES task, while the TUBES task functions as a control condition.

Little is yet known about great apes' performance on displacements through opaque tubes. Tomonaga, Imura, Mizuno, and Tanaka (2007) showed that chimpanzees failed to predict the trajectory of an object displacement along an s-shaped vertical opaque tube. However, moving the displacement from the vertical to a horizontal plane increased performance in cotton top tamarins (Saguinus oedipus) (Hauser, Williams, Kralik, & Moskovitz, 2001). To our knowledge, the TUBES task is the first investigation into non-human great apes' mastery of invisible displacements through opaque tubes in a horizontal plane.
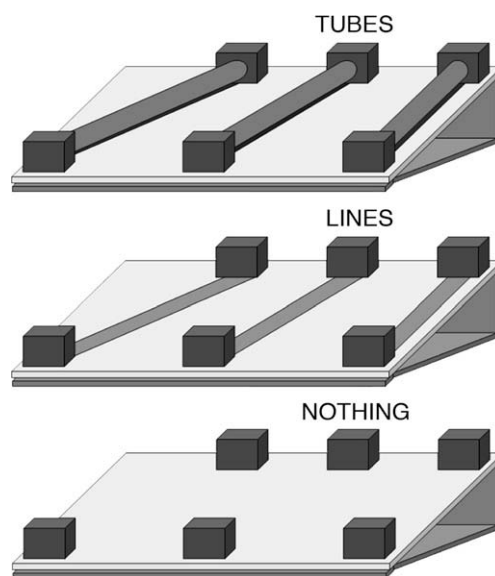

Fig. 1. Experimental setup 3D: An inclined table was placed between the participant and the experimenter. Six identical cups where placed on the table in two rows of three. In three alternative conditions, the spatially corresponding cups in the two rows (left–left; middle–middle; right–right) were connected following one of three ways: Either with grey opaque PVC tubes (TUBES task) or grey lines with a width equal to the tubes' diameter (LINES task) or nothing (NOTHING task, study 3 and 4 only).

Call (2006a) has proposed that great apes solve problems based on logico-causal relations more easily than problems based on non-causal relations. Thus, an additional question raised in this study was whether subjects responded differentially to the type of link and whether the presence of those links was necessary to solve the task in the first place.

Of the two rows of cups placed on an inclined experimental table, one row was placed along the top edge of the incline, the other along the bottom edge. While the cups forming the bottom row were distributed along the full length of the edge, the top row only spread across approximately two thirds of that distance. The two rows of cups were aligned either flush right or flush left (counterbalanced across participants). As a result one pair of cups was spatially aligned while the other pairs were spatially misaligned (Fig. 2). This setup, especially in the LINES task where the tubes did not provide causal connections, created three qualitatively different pairings of cups which revealed their connections following three different combinations of strategies:

GPR: In the first pair the two cups are spatially aligned on an incline; hence the line of gravity (G) can be used as an indicator of pairing. Further, the two cups are closer to each other in proximity (P) than to any other potential partner in the respective other constellation. Finally the two cups have similar spatial relations (R) to the other cups within their respective constellation. PR: The second pair of cups is misaligned on the incline, hence G is not a viable strategy to connect the pair. However, P and R still provide valid strategies for pairing. R: In the third and final pair neither G nor P work as connection strategies. R is not just the only viable strategy left, but set up in conflict to P, which in turn suggest a connection between for e.g. the left-most cup in the top row and the central cup in the bottom row of Fig. 2. We argue that any participant consistently solving the R condition correctly in the LINES task (or NOTHING task), recognizes the relational similarities between the corresponding cups in the two presented rows. Hence, throughout all experiments presented here,

the R condition in the LINES task (or NOTHING task) provides the crucial test case for the participants' ability to recognize relational similarities in this spatial task.

2. Study 1: children

2.1. Method

2.1.1. Participants

Forty-eight four-year-old children (mean age = 47.6 SD = 3.4, range 42–53 months) participated in this study. There was an equal number of boys and girls. All children were recruited from local kindergartens, were native German speakers of normal ability range and came from mixed socio-economic backgrounds. All caretakers gave their informed consent. Participants that did not pass one of the pre-tests (see below) were excluded from the final analysis (1 excluded). Children were free to choose not to participate at all times (0 quit). The samples reported above are the final numbers after exclusions.

2.1.2. Materials

A table (80×50 cm) with a 15 degree incline was placed between the participant and the experimenter. Six identical cups where placed on the table in two rows of three (see Fig. 1). One row was placed along the top edge of the incline, the other along the bottom edge. While the cups forming the bottom row were distributed along the full length of the edge, the top row only spread 55 cm (see Fig. 2). The two rows of cups were aligned either flush right or flush left (counterbalanced across participants). As a result one pair of cups was spatially aligned while the other pairs were spatially misaligned. As described above, this setup created three qualitatively different pairings of cups: GPR, PR and R. The cups were fixed to the table with Velcro to allow moving and removing during a session. The spatially corresponding cups in the two rows (left–left; middle–middle; right–right) could be flexibly connected with PVC tubes (2.5 cm diameter) or lines (2.5 cm wide). Participants were seated on the bottom end of the incline while the experimenter was seated at the top end.

2.1.3. Procedure

There were two phases: pre-test and test.

Pre-test: Before the first session of every condition (TUBES, LINES), participants were subjected to one of two pre-tests. Before the first session of the TUBES task a PVC tube was placed in an empty inclined table. The head-end of the tube was always placed at the center of the table. The tail end was moved between a left-middle and right position from trial to trial. After placing the tube, the experimenter holds a target object right in front of the head-end of the tube, waiting for participant to place their hand where they expect the bait to fall when dropped. Placing the hand at the tail end of the tube was counted as a correct response. Participants passed the tests if they made three correct choices in a row. Before the first session of the LINES task (and also the NOTHING task in studies 3 and 4) three identical cups where placed equidistant in a row at the bottom of the incline, right in front of
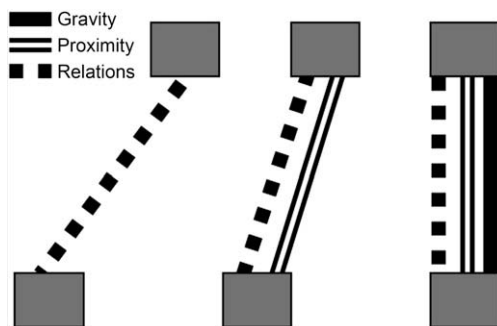

Fig. 2. Experimental setup birds-eye: Three different spatial strategies in different combinations could be used by the participants to code the relationship between corresponding cups: G: The two cups are spatially aligned on an incline; hence the line of gravity can be used as an indicator of pairing. P: The two cups are closer to each other in proximity than to any other potential partner in the respective other constellation. R: The two cups have similar spatial relations to the other cups within their respective constellation.

the participant. The experimenter opened all the cups, placed bait in one of them and closed all cups from left to right. Then the experimenter elicited a choice response (for slight variations see individual experiments). After the choice, all cups in the participant's row are opened and in case of a correct response the participant receives the bait as reward. After an incorrect response, the bait is placed back in a holding container. Participants passed the tests if they made three correct choices in a row.

Test: In Study 1 the spatially corresponding cups in the two rows (left–left; middle–middle; right–right) were connected following one of two ways: Either with grey opaque PVC tubes (TUBES task) or grey lines (LINES task) with a width equal to the tubes' diameter. At the beginning of each trial an occluder is raised to prevent participants from observing the first half of the baiting procedure. At this point, all cups are open. Behind the occluder, with a target object in the right hand, the experimenter visits all cups in the participant-row from right to left, dropping the target in one of them. Then, still behind the occluder the experimenter closes the lids of all cups in the participant's row from left to right. Then he removes the occluder, holds up an identical target object to catch the participants' attention and then places the bait in one of the cups in the experimenter-row. Then the experimenter shakes the table slightly and then looks at the participant to elicit a choice response. After the choice, all cups in the participant-row are opened and in case of a correct response the participant receives the bait as reward. After an incorrect response, the bait is placed back in a holding container. Each child received one 18-trial session. Half of the trials corresponded to the TUBES task and the other half to the LINES task. Each task was presented as a 9-trial block with the order of task presentation counterbalanced across subjects.

2.1.4. Data scoring and analysis

We videotaped all trials. We scored the container selected by subjects based on videotapes and/or in situ notes. Our dependent variable was the percent of correct trials. Our independent variables were the type of task (TUBES, LINES), condition (GPR, PR, R) as well as order of task administration and age group (young = 42–48 months; old = 48–54 months). We investigated the effect of all of these factors on the dependent variable using non-parametric statistics because the data did not meet the supposition of normality. We also assessed whether subjects performed above chance ($p = 0.33$) in each condition for each task. All tests were two-tailed.

2.2. Results

Overall children performed significantly better with the TUBES than the LINES task (Wilcoxon test: $z = 3.90$, $p < 0.001$, $N = 40$; TUBES: mean = 81.5, SEM = 3.9; LINES: mean = 60.9, SEM = 4.2). Therefore we decided to analyze the TUBES and LINES tasks separately.

2.2.1. TUBES task

Fig. 3 presents the percentage of correct trials in each condition of the TUBES task. Overall, there were significant
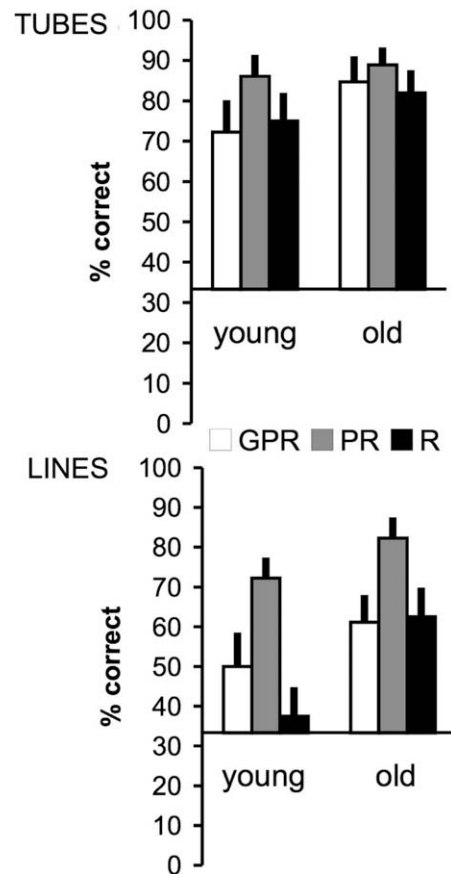

Fig. 3. Study 1: Young (3;6–3;11) and old (4;0–4;6) children's performance displayed as mean percent correct (+SE) for the TUBES and LINES tasks in all three conditions (GPR, PR, R).

differences across conditions (Friedman test: $\chi^2 = 8.00$, $p = 0.018$, df = 2, $N = 48$). Post-hoc tests revealed that children performed better in the PR condition compared to both the GPR (Wilcoxon test: $z = 2.50$, $p = 0.012$, $N = 15$) and R conditions (Wilcoxon test: $z = 2.29$, $p = 0.022$, $N = 19$).

To test for age differences we divided the children into two age groups (young: 3;6–3;11 and old: 4;0–4;6 years of age). There were no significant differences between age groups in any of the conditions (Mann–Whitney tests: GPR: $z = 1.26$, $p = 0.21$; PR: $z = 0.32$, $p = 0.75$; R: $z = 0.67$, $p = 0.50$; $N = 48$ in all cases). There was no significant effect of order of administration for any of the conditions (Mann–Whitney tests: $z < 1.01$, $p > 0.31$, $N = 48$ in all cases, Fig. 3). Children performed above chance in all conditions of the TUBES task regardless of whether they received the TUBES or the LINES task first (Wilcoxon tests: $z > 3.97$, $p < 0.001$, $N = 24$ in all cases). Amongst the young children 17/24 passed the TUBES task significantly above chance (Binomial test). Amongst the old children 19/24.

2.2.2. LINES task

Fig. 3 presents the percentage of correct trials in each condition for the LINES task. Overall, there were significant

differences across conditions (Friedman test: χ^2 = 30.06, p < 0.001, df = 2, N = 48). Post-hoc tests revealed that children performed better in the PR condition compared to both the GPR (Wilcoxon test: z = 3.60, p < 0.001, N = 28) and R conditions (Wilcoxon test: z = 4.58, p < 0.001, N = 32).

There were no significant differences between age groups in the GPR (Mann–Whitney test: z = 0.92, p = 0.36, N = 48) and PR conditions (Mann–Whitney test: z = 1.56, p = 0.12, N = 48). In contrast, older children outperformed younger ones in the R condition (Mann–Whitney test: z = 2.30, p = 0.021, N = 48). However, there was a significant effect of order of task administration because children who received the TUBES task prior to the LINES task performed better in the LINES task than those children who received the LINES task first (Mann–Whitney tests: z > 2.53, p < 0.02, N = 48 in all cases). The order effect was particularly evident for older children. Amongst the 12 (out of 24) older children who passed the LINES task (Binomial test: p < 0.05), 11 had previously received the TUBES task (Fisher's exact test: p < 0.001). Of the 7 (out of 24) younger children who passed the LINES task, 5 had previously received the TUBES task (Fisher's exact test: p = 0.37). Table 1 presents the median percent correct as a function of age, order of task administration and condition. When subjects received the LINES task first, they only responded above chance in the PR condition. In contrast when they received the LINES task second, they responded above chance in all conditions except the younger children in the R condition. Collapsing the data across ages (indicated by the subtotals in Table 1), children performed above chance in all conditions (Wilcoxon tests: z > 3.73, p < 0.001, N = 24 in all cases) except in the GPR (Wilcoxon test: z = 1.90, p = 0.057, N = 24) and R conditions (Wilcoxon tests: z = 1.21, p = 0.23, N = 24) when they received the LINES task first.

2.2.3. Box preferences

We also analyzed whether children showed a preference for one of the three boxes (left, middle, right) by comparing the percent of trials in which they selected each box independently of their success and the task. Children showed a strong response bias (Friedman test: χ^2 = 38.69, p < 0.001, df = 2, N = 48). Post-hoc tests indicated that children selected the middle box significantly more often than the left- and the right-side box (Wilcoxon tests: z > 5.01, p < 0.001, N = 42 in both cases).

Table 1
Median percent correct in the LINES task as a function of condition, age and order of task presentation in Study 1. Asterisks indicate above chance performance (Wilcoxon test: p < 0.05 against 33% expected).

First task	Age	Condition		
		GPR	PR	R
Lines	Young	33	67[*]	17
	Old	33	67[*]	33
	Subtotal	33	67[*]	33
Tubes	Young	50[*]	83[*]	33
	Old	83[*]	100[*]	100[*]
	Subtotal	67[*]	100[*]	67[*]

2.3. Discussion

Children found the TUBES task much easier than the LINES task. Spatial relations per se cannot explain the difference between conditions because both the TUBES and LINES task instantiated identical spatial relations. Therefore, we argue, the difference lies on the type of stimuli that connect the two sets of boxes. The reason for the discrepancy between tasks may lay in the causal or arbitrary nature of the stimuli connecting both sets of boxes. Tubes, but not lines, can enable the invisible transfer of a target object from the top to the bottom row of boxes.

Children performed better in the PR condition compared to the other two conditions, especially in the lines task. Unlike, the difference between the TUBES and LINES task to which we alluded previously, we suspect that the difference between conditions may be an artifact because children, in general, showed a greater predisposition for selecting the middle box – a predisposition also observed in at least one previous study (Call, 2001). Thus, the high performance in the PR condition may be partly seen as an artifact of the children's intrinsic preference for the middle box, not as a genuine difference. Box preferences, however, cannot explain their success in the two other conditions. Further, in the LINES task children who received the TUBES task prior to the LINES task performed better than those children who received the LINES task first. This suggests that some understanding of the relational structure of the setup gained via the causal connection presented in the TUBES task can carry over into the less intuitive LINES task. Finally, and most importantly, especially after receiving the TUBES task first, the older age group outperformed the younger age group in the R condition where gravity and proximity were ruled out as potential strategies, but not in any of the other conditions. Since, as described above, the R condition in the LINES task is the crucial test case, only the older children showed reliable evidence of recognition of the relational similarity between cups. This difference seems largely due the older children's ability to transfer knowledge from the TUBES task. Next, we investigated the performance of the non-human great apes in the same task.

3. Study 2: non-human great apes

3.1. Method

3.1.1. Participants

Five bonobos (*Pan paniscus*), 10 chimpanzees (*Pan troglodytes*), 2 gorillas (*Gorilla gorilla*) and 4 orangutans (*Pongo pygmaeus*) participated in this experiment. There were 7 males and 14 females ranging from 6 to 31 years of age (M = 16 years 8 months; SD = 8 years 11 months). All subjects were housed at the Wolfgang Köhler Primate Research Center at Zoo Leipzig (Germany). They lived in social groups with conspecifics and had access to indoor and outdoor areas designed appropriate to the species. During testing, the apes were fed according to their daily routine four times a day on a diet of fruit, vegetables and monkey chow; water was at their disposal at all times. Par-

ticipants that did not pass one of the pre-tests were excluded from the final analysis (2 excluded). Animals were free to choose not to participate at all times (3 quit). The samples reported above are the final numbers after exclusions.

3.1.2. Materials
We used the same apparatus as in study 1.

3.1.3. Procedure
The pre-test and test were identical to those in Study 1 with two exceptions. First, we increased the number of sessions. Each subjects received four 18-trial sessions instead of just one. Two sessions were devoted to the TUBES task and two sessions to the LINES task. Thus, subjects received a total of 36 trials per task. Second, to initiate a response, instead of shaking the table as in study 1 the experimenter slid the table towards the subjects to allow them to select one of the three boxes. Furthermore grapes were used as target objects instead of toys.

3.1.4. Data scoring and analyses
We used the same scoring procedure and analyses as in Study 1 with the exception of investigating the effect of species as the dependent variable instead of age.

3.2. Results

Overall apes performed significantly better with the TUBES than the LINES task (Wilcoxon test: $z = 3.92$, $p < 0.001$, $N = 20$; TUBES: mean = 87.0, SEM = 3.3; LINES: mean = 67.2, SEM = 3.8). Therefore we decided to analyze the TUBES and LINES tasks separately.

3.2.1. TUBES task
Fig. 4 presents the percentage of correct trials in each condition for the TUBES task as a function of species. There were no overall differences across conditions (Friedman test: $\chi^2 = 0.14$, $p = 0.93$, df = 2, $N = 21$). Additionally, there were no differences between species in any of the three conditions (Kruskal–Wallis tests: GPR: $\chi^2 = 4.86$, $p = 0.18$, df = 3, $N = 21$; PR: $\chi^2 = 7.17$, $p = 0.067$, df = 3, $N = 21$; R: $\chi^2 = 1.67$, $p = 0.64$, df = 3, $N = 21$). Overall, apes performed above chance ($p = 0.33$) in all conditions (Wilcoxon test: $z > 4.03$, $p < 0.001$, $N = 21$). Amongst apes 18/21 passed the TUBES task significantly above chance (9/10 chimpanzees; 5/5 bonobos; 2/2 gorillas; 2/4 orangutans) (Binomial test).

Apes that received the TUBES task prior to the LINES task performed better in the PR and R conditions of the TUBES task (but not in the GPR) than those apes that received the LINES task first (Mann–Whitney tests: GPR:
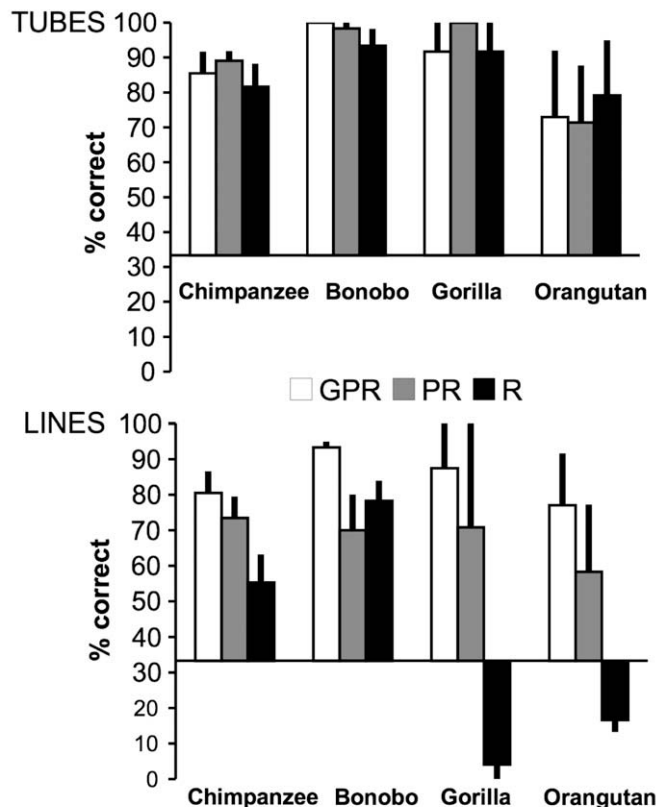

Fig. 4. Study 2: Mean percent correct (+SE) for the TUBES and LINES tasks in all three conditions (GPR, PR, R) for all four extant species of non-human great apes: Chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*).

$z = 1.84$, $p = 0.066$; PR: $z = 2.17$, $p = 0.03$; R: $z = 2.53$, $p = 0.012$, $N = 21$ in all cases). Apes performed above chance in all conditions of the TUBES task regardless of the order of administration of the TUBES and LINES tasks (Wilcoxon tests: $z > 2.61$, $p < 0.01$, $N = 11$ in all cases).

3.2.2. LINES task

Fig. 4 presents the percentage of correct trials in each condition for the LINES task as a function of species. Overall, there were significant differences between conditions (Friedman test: $\chi^2 = 22.14$, $p < 0.001$, df = 2, $N = 21$). Post-hoc tests revealed that apes performed better in the GPR condition compared to both the PR (Wilcoxon test: $z = 2.04$, $p = 0.041$, $N = 17$) and R conditions (Wilcoxon test: $z = 3.83$, $p < 0.001$, $N = 19$). Likewise, subjects performed better in the PR compared to the R condition (Wilcoxon test: $z = 2.54$, $p = 0.011$, $N = 18$).

Amongst apes 10/21 passed the LINES task significantly above chance (5/10 chimpanzees; 5/5 bonobos; 0/2 gorillas; 0/4 orangutans) (Binomial test). There were no significant differences between species in the GPR (Kruskal–Wallis tests: $\chi^2 = 1.25$, $p = 0.74$, df = 3, $N = 21$) and PR conditions (Kruskal–Wallis tests: $\chi^2 = 0.37$, $p = 0.95$, df = 3, $N = 21$). In contrast, there were significant differences between species in the R condition (Kruskal–Wallis test: $\chi^2 = 13.46$, $p = 0.004$, df = 3, $N = 21$). Post-hoc tests revealed that chimpanzees outperformed both gorillas (Mann–Whitney test: $z = 2.15$, $p = 0.031$, $N = 12$) and orangutans (Mann–Whitney test: $z = 2.63$, $p = 0.009$, $N = 14$) but not bonobos (Mann–Whitney test: $z = 1.67$, $p = 0.094$, $N = 15$). Similarly, bonobos outperformed both gorillas (Mann–Whitney test: $z = 2.01$, $p = 0.044$, $N = 7$) and orangutans (Mann–Whitney test: $z = 2.50$, $p = 0.012$, $N = 9$). There were no significant differences between gorillas and orangutans (Mann–Whitney test: $z = 1.67$, $p = 0.095$, $N = 6$). A comparison of the genus Pan against gorillas and orangutans pooled together (*non-Pan*) indicated that the former significantly outperformed the latter (Mann–Whitney test: $z = 3.41$, $p = 0.001$, $N = 21$), thus confirming the previous results.

Chimpanzees and bonobos performed above chance in all conditions (Wilcoxon test: $z > 3.07$, $p < 0.01$, $N = 15$ in all cases). Gorillas and orangutans performed above chance in the GPR condition (Wilcoxon test: $z = 2.23$, $p = 0.026$, $N = 6$), at chance in the PR condition (Wilcoxon test: $z = 1.57$, $p = 0.12$, $N = 6$) and significantly below chance in the R condition (Wilcoxon test: $z = 2.21$, $p = 0.027$, $N = 6$).

We also assessed the effect of order of administration of the subjects' performance in the LINES task. Apes that received the TUBES task prior to the LINES task performed better in the GPR and PR conditions of the LINES task (but not in R) than those apes that received the LINES task first (Mann–Whitney tests: GPR: $z = 2.64$, $p = 0.008$; PR: $z = 2.72$, $p = 0.007$; R: $z = 1.13$, $p = 0.26$, $N = 21$ in all cases, see Fig. 4). Apes performed above chance in all conditions of the LINES task regardless of the order (Wilcoxon tests: $z > 2.13$, $p < 0.05$, $N = 10$ in all cases) except in the R condition (Wilcoxon test: $z = 0.66$, $p = 0.51$, $N = 10$) when they received this condition before to the TUBES task (see Table 2).

Table 2
Median percent correct in the LINES task as a function of condition, genus and order of task presentation in Study 2. Asterisks indicate above chance performance (Wilcoxon test: $p < 0.05$ against 33% expected).

First task	Age	Condition		
		GPR	PR	R
Lines	*Pan* (n = 7)	67[*]	60[*]	50
	non-Pan (n = 3)	92[a]	42[a]	17[a]
	Subtotal	67[*]	55[*]	29
Tubes	*Pan* (n = 8)	100[*]	83[*]	75[*]
	non-Pan (n = 3)	92[a]	83[a]	8[a]
	Subtotal	92[*]	83[*]	67[*]

[a] The small sample size of the genus *non-Pan* (gorillas and orangutans) prevented us from testing the observed values against chance.

3.2.3. Comparing children and apes

For purposes of comparability between children and the other great ape species we only considered the first nine trials that each subjects received. There were significant differences between groups only in the GPR condition (Kruskal–Wallis test: $\chi^2 = 7.42$, $p = 0.025$, df = 2, $N = 73$). Post-hoc tests revealed that apes outperformed young (Mann–Whitney test: $z = 2.33$, $p = 0.02$, $N = 49$) and old children (Mann–Whitney test: $z = 2.47$, $p = 0.014$, $N = 49$). In contrast, there were no significant differences between children (Mann–Whitney test: $z = 0.20$, $p = 0.84$, $N = 48$). Comparing the pattern of results presented in Tables 1 and 2 indicated that members of the genus Pan performed in a way similar to older children, that is, they failed the R condition when they received the LINES task prior to the TUBES task but they passed it when they received the LINES task after the TUBES task. Order of presentation did not determine whether subjects passed or failed the other conditions.

3.3. Discussion

All tested non-human species successfully solved the TUBES task. Like children, they found the TUBES task easier than the LINES task. However, there were striking differences between the TUBES and the LINES task across species. Whereas all apes performed at similar levels, and above chance in all conditions of the TUBES task, only chimpanzees and bonobos solved all the conditions of the LINES task. Gorillas and orangutans failed to reach above chance performance in the PR and were below chance in the R condition.

Pooled across tasks, non-human apes outperformed human children in the GPR condition. In the other conditions, both groups performed equally well. Unlike children, apes did not show a preference for the middle box. In fact, apes performed in all conditions at a similar level with the exception of the orangutans and gorillas noted above on the PR and R conditions of the LINES task. Since, as described above, the R condition in the LINES task is the crucial test case, only chimpanzees and bonobos showed reliable evidence of recognition of the relational similarity between cups. One question that remains unanswered is whether chimpanzees (and bonobos) would also be able to solve the task without the presence of stimuli (tubes

or lines) connecting the two sets of boxes. Additionally, it would be important to confirm the differences detected here between apes of the genus Pan (chimpanzees and bonobos) and the other great ape species. We addressed these two in the next two experiments.

4. Study 3: chimpanzees without tubes or lines

4.1. Method

4.1.1. Participants

Four chimpanzees (one 8-year-old male, three 13-year-old females) that had not participated in Study 2 were included in this experiment. Housing and testing conditions were identical to Study 2. None of the participants were excluded from analysis.

4.1.2. Materials

We used the same apparatus as in Study 2 except that there were no tubes or lines connecting the two sets of boxes (Fig. 1).

4.1.3. Procedure

The procedure for the pre-test and test were identical to those used in Study 2 except that we tested a single task in which there were no tubes or lines connecting the two sets of boxes (NOTHING task). Each subjects received two 18-trial sessions.

4.1.4. Data scoring and analyses

We used the same scoring procedure as in the previous studies. We compared the subjects in the current study to those chimpanzees that in Study 2 had received either the TUBES task or the LINES task in the first place. This means that unlike Study 2, the current study is based on a between-subject design.

4.2. Results

There were significant differences between tasks (Kruskal–Wallis test: $\chi^2 = 6.09$, $p = 0.048$, df = 2, $N = 14$). Post-hoc tests confirmed that chimpanzees performed better in the TUBES compared to the LINES task (Mann–Whitney test: $z = 2.61$, $p = 0.009$, $N = 10$). In contrast, there were no significant differences between the NOTHING and the TUBES (Mann–Whitney test: $z = 1.35$, $p = 0.18$, $N = 9$) or LINES tasks (Mann–Whitney test: $z = 0.25$, $p = 0.86$, $N = 9$). Focusing on the NOTHING task, chimpanzees appeared to perform equally well in all conditions (Friedman test: $\chi^2 = 1.27$, $p = 0.53$, df = 2, $N = 4$, Fig. 5). Amongst chimpanzees 5/5 passed the TUBES task, 3/5 passed the LINES task and 2/4 passed the NOTHING task significantly above chance (Binomial test).

Our limited sample size prevented us from testing whether subjects were above chance in the NOTHING task. However, pooling together that data from the LINES and NOTHING tasks indicated that subjects selected the correct cup above chance levels (Wilcoxon test: $z = 2.67$, $p = 0.008$, $N = 9$). This value represents a conservative estimate of the apes' performance on the NOTHING task because subjects
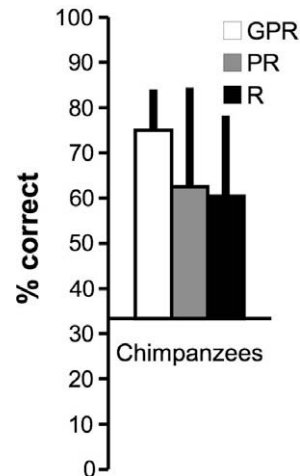

Fig. 5. Study 3: Mean percent correct (+SE) for the NOTHING task across all three conditions for chimpanzees (*Pan troglodytes*).

performed worse in the LINES task than in the NOTHING task.

4.3. Discussion

Chimpanzees' performance in the NOTHING task fell in between that of the TUBES and LINES tasks. Although there were no significant differences between the NOTHING task and the other two tasks, it is very likely that this was due to our small sample size. Nevertheless, since subjects' performance in the NOTHING task was higher than in the LINES task and subjects were above chance in the LINES task in Experiment 2, this suggests that chimpanzees were also able to solve the task without the aid of lines. Our conservative assessment of above chance performance (based on pooling together the LINES and the NOTHING task) reinforced this conclusion. Again, chimpanzees showed evidence of recognition of the relational similarity between cups even when there was no connector (tube or line) between corresponding cups. Next, we sought to confirm the results with orangutans in a different population than the one investigated in Study 2.

5. Study 4: confirmatory study with orangutans

5.1. Method

5.1.1. Participants

Twelve orangutans (*P. pygmaeus*) participated in this study. There were 5 males and 7 females ranging from approximately 5 to 7 years 6 months of age ($M \approx 6$ years 3 months; SD ≈ 9 months). All animals were housed at the Orangutan Care Centre and Quarantine, Pasir Panjang, Indonesia. They lived in same-sex peer groups with conspecifics and had access to 80 ha of forest area around the clinic every other day. During testing, the apes were fed according to their daily routine. Animals were free to choose not to participate at all times (8 quit). All participants passed the pre-tests (see Study 2). The samples reported above are the final numbers after exclusions.

5.1.2. Materials
We used the same apparatus as in Study 2 and 3.

5.1.3. Procedure
The procedure for the pre-test and test were identical to those used in Study 2 except that we tested three different conditions in which there were either tubes or lines or nothing connecting the two sets of boxes. Each subject received one 18-trial session in one of the three conditions.

5.1.4. Data scoring and analyses
We used the same scoring procedure as in the previous studies.

5.2. Results

Fig. 6 presents the percentage of correct trials in each task. There were significant differences between tasks (Kruskal–Wallis test: $\chi^2 = 8.78$, $p = 0.012$, df = 2, $N = 12$). Post-hoc tests confirmed that orangutans performed better in the TUBES task compared the LINES task (Mann–Whitney test: $z = 2.18$, $p = 0.029$, $N = 8$) and the NOTHING task (Mann–Whitney test: $z = 2.32$, $p = 0.02$, $N = 8$). In contrast, there was no significant difference between the LINES task and the NOTHING task (Mann–Whitney test: $z = 1.90$, $p = 0.058$, $N = 8$) although unlike chimpanzees, orangutans tended to perform worse in the NOTHING than in the LINES task. Amongst orangutans 4/4 passed the TUBES task, 2/4 passed the LINES task and 0/4 passed the NOTHING task significantly above chance (Binomial test).

Focusing on the NOTHING task, orangutans' performance tended to decrease directly as function of the number of strategies available to the subjects (see Fig. 6), although this difference was not significant (Friedman test: $\chi^2 = 1.29$, $p = 0.53$, df = 2, $N = 4$). Our limited sample size prevented us from testing whether subjects were above chance in the NOTHING task.

5.3. Discussion

Orangutans performed better in the TUBES compared to the LINES task, thus confirming the results of
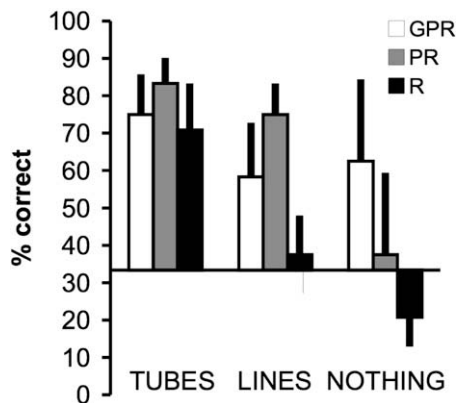

Fig. 6. Study 4: Mean percent correct (+SE) for the TUBES, LINES and NOTHING tasks across all three conditions for orangutans (*Pongo pygmaeus*).

Study 2. Unlike chimpanzees, however, the orangutans' performance on the NOTHING was lower than in the LINES task. This result paired with the marked decrease in performance as a function of the number of strategies available (they performed worse when relational similarity (R) was the only strategy available) suggests that chimpanzees and orangutans differ in their ability to recognize relational similarities in a spatial task. This again appears to confirm the differences detected between members of the genus Pan and other non-human great apes.

6. General discussion

There were three main findings in this study. First, all species including 3- to 4-year-old children performed better in the TUBES than the LINES task. Second, although all species performed above chance in the TUBES task, only children, bonobos and chimpanzees, unlike gorillas and orangutans, performed above chance in the LINES task. Third, only older children, bonobos and chimpanzees, but not younger children gorillas and orangutans, solved the condition in which a spatial relational similarity was the only viable strategy to pair corresponding cups. Next we discuss these findings in more detail.

All groups performed above chance in the TUBES task and better than in the LINES task. The spatial disposition of the containers cannot explain this difference between tasks because it was identical. Similarly, it is difficult to make a convincing argument purely based on the perceptual configurations of the stimuli in each task. Both tasks presented a bottom row of containers connected to a top row of containers by grey lines of approximately the same thickness, the only difference being that in the tubes task the connectors were 3D while in the lines task they were 2D. Such similarity across tasks is important because it rules out the idea that all that subjects did in the TUBES task was to use a perceptual configuration without any insight on the causal relations between the reward, the containers and the lines that connected them. If that had been all that subjects did, we would have expected no differences between tasks.

Still one might suggest that subjects succeed in this task because they perceive two corresponding cups and their connector as a single object by means of perceptual grouping. At this point we know little about perceptual grouping in non-human great apes (Fagot & Tomonaga, 2001). However, we see two possible attempts to explain the present findings by means of perceptual grouping and find neither highly convincing: If subjects had perceived the two connected cups and their connector as a single grouped object irrespective of the shape of the connector (line or tube), no difference should occur between tasks or conditions. Alternatively, if 3D-connectors (tubes) facilitated grouping in comparison to 2D-connectors (lines), this might explain differences between tasks, but could still not account for the presented differences between conditions. Therefore, we have to hypothesise another reason for explaining the robust difference between TUBES and LINES.

Our hypothesis is grounded on the causal knowledge that subjects have about objects and their relation with other objects characteristic of the Piagetian late sensori-motor and representational period. We argue that to solve the TUBES task, subjects have to understand that a tube can enable a target object to move undetected from its starting box to the goal box – something that cannot occur in the LINES task. This requires a mastery of both invisible displacement and tertiary objects relations, two skills that have been previously documented in the great apes (e.g. Antinucci, 1990). In the LINES task no object could move from one container to the other undetected. Causal knowledge and invisible displacement are clearly not a viable explanation. While initially, the visual connection provided by the lines, even though maybe a weaker cue than the causal tubes, might be considered crucial for subjects to understand the relationships between cups, we find this unlikely for two reasons. First, many subject failed at least one of the conditions in the LINES task, despite the presence of a visual connection and second, chimpanzees, at least, also solved the task without lines being present at all. Based on our data we hypothesise that subjects solve the TUBES task by means of mastery of both invisible displacement and tertiary objects relations and the LINES task by applying non-causal, more abstract strategies such as proximity or relational similarity.

Crucially, gorillas and orangutans, which are species that are more distantly related to humans compared to chimpanzees and bonobos, displayed inferior relational abilities in this task. Not only did they fail the LINES task overall but more interestingly, they were also below chance in the condition in which spatial relational similarity was the only viable strategy to pair corresponding cups (R condition). In contrast they performed above chance in those conditions in which other strategies such as proximity could be used. This suggests that there may have been an increase in the capacity to recognize relational similarity in the great ape family (Hominidae) between 6 and 10 million years ago. This conclusion should remain tentative until a larger sample of apes of the four species, especially bonobos and gorillas, is tested. If our results should prevail we need to consider mechanistic and ultimate explanations for these differences across phenotypes. For example, as pointed out by a reviewer, the behavioural differences reported here parallel differences in the presence and density of a particular cell type (spindle or Von Economo neurons) across species as well as ages. (Nimchinsky et al., 1999; Allman, Hakeem, & Watson, 2002; Allman, Watson, Tetreault, & Hakeem, 2005). The density of this cell type is particularly high in the anterior cingulate cortex (ACC) relaying into anterior prefrontal cortex (aPFC), which has, among other functions, been associated with the manipulation of relational knowledge (Kroger et al., 2002; Ramnani & Owen, 2004). Hence differences in ACC and aPFC anatomy and function might account for the differing behavioural proficiencies reported here. However, in order to ultimately understand why advanced skill evolved in one line of great apes and not the others, we need to precisely isolate the application of this increased cognitive capacity in humans, bonobos and chimpanzees in order to deduct selection pressures favouring this phenotype.

Furthermore, the preference for proximity-based over relation-based strategies in some great apes species but not others raises questions about previous research reporting similarities in spatial behaviour across all great apes without controlling for alternative strategies as precisely as it is done here (e.g. Haun, Rapold et al., 2006). Manipulations similar to the ones presented in this manuscript will help to more closely dissect differences and similarities between great ape species' spatial cognition.

Although 4-year-old children solved the LINES task at the same level of chimpanzees and bonobos including the condition in which gravity and proximity were not viable strategies, we cannot conclude from these data alone that they used relational similarity because unlike chimpanzees, we did not run a condition in which there were no lines present. However, since previous studies had already established that children can solve spatial tasks using relational similarity (e.g. Loewenstein & Gentner, 2005), it is likely that children from around 4 years of age onwards, like chimpanzees, used the spatial relations between stimuli to solve the task.

One important conclusion of this study is that recognition of simple relational similarities is not language dependent. Chimpanzees and Bonobos were capable of solving the spatial relational task without language or any explicit training on symbolic devices. Note that this type of problem is also not much aided by having language given that chimps were as good as 4-year-old humans. Nevertheless, previous studies have shown that linguistic relational labels help children to solve spatial relational problems (Gentner, 2003; Gentner & Rattermann, 1991; Loewenstein & Gentner, 2005). In combination with the present study we argue that having language is not necessary or even not very helpful for simple spatial relational reasoning, but that using language while solving the task will improve performance (Gentner, 2003; Gentner & Rattermann, 1991; Loewenstein & Gentner, 2005). We would however also consider that the impact of having a language might increase as the complexity of the relational task increases. We do not mean to argue that language has no bearing on the recognition of simple relational similarities. Previous research has strongly suggested that preferred relational strategies align with particular features of language systems across human cultures (Haun, Rapold et al., 2006; Levinson, 2003; Majid et al., 2004; Mishra & Dasen, 2005; Neumann & Widlok, 1996; Pederson et al., 1998; Wassmann & Dasen, 1998). Hence language might very well change the shape of the ability to recognize simple relational similarity, but apparently not its existence.

Alternatively to our interpretation, one might argue that none of our participant groups used relational similarity to solve the LINES and NOTHING tasks but proximity-based strategies to solve the GPR and PR conditions and inference by exclusion to deduct the association between the cups in the R condition. This combination of strategies provides a non-relational interpretation not only of our data, but also previous studies in which all mappings could be solved using proximity and hence deductions by inference where not necessary (e.g. Loewenstein & Gentner, 2005). Based on our own data we find this alternative unlikely for several reasons. First, all great apes have been

reported to infer by exclusion (Call, 2004, 2006b). Hence this alternative does not easily fit with the species difference reported above. Second, we find this alternative less parsimonious since it relies on a larger combination of non-trivial cognitive demands, all previously reported to be difficult to match for non-human primates (Tomasello & Call, 1997): (a) The prior (or rapidly acquired) assumption on part of the participant that any cup can only ever have exactly one correspondent, (b) the switching between two strategies (proximity-exclusion) from one trial to the next, (c) including inhibition of a response that works 2 out of 3 times (proximity). Future studies will have to provide definite proof for one or the other explanation.

In conclusion, we argue that great apes including humans can infer the location of the reward if the stimuli relations afford the displacement of an invisible reward from one location to another. This result is consistent with previous work showing that apes perform better in tasks whose elements are connected by logico-causal as opposed to arbitrary relations (Call, 2006a). Additionally, all great apes use non-causal spatial strategies such as proximity to associate objects in space. However, only chimpanzees, bonobos and human children from around 4 years of age are also capable of engaging another mechanism based on perceiving relational similarity between containers to solve the problem. The presence of logico-causal and proximity-based spatial reasoning in all great apes suggests that the precursors of this skill may have evolved at least around 15 million years ago when all extant great apes shared a common ancestor. In contrast, relational reasoning may have been a more recent appearance, perhaps within the last 6–10 million years ago when chimpanzees, bonobos and humans shared a common ancestor. Future studies are required to confirm this finding and to explore the limits of this skill in our closest living relatives.

Acknowledgements

References

Allman, J. M., Hakeem, A., & Watson, K. (2002). Two phylogenetic specializations in the human brain. The Neuroscientist, 8(4), 335–346.
Allman, J. M., Watson, K. K., Tetreault, N. A., & Hakeem, A. Y. (2005). Intuition and autism: a possible role for Von Economo neurons. Trends in Cognitive Sciences, 9(8), 367–373.
Antinucci, F. (1989). Cognitive structure and development in nonhuman primates. Hillsdale: Erlbaum.
Antinucci, F. (1990). The comparative study of cognitive ontogeny in four primate species. In K. R. Gibson & S. T. Parker (Eds.), "Language" and intelligence in monkeys and apes: Comparative developmental perspectives (pp. 157–171). Cambridge: Cambridge University Press.
Barth, J., & Call, J. (2006). Tracking the displacement of objects: A series of tasks with Great apes and young children. Journal of Experimental Psychology: Animal Behavior Processes, 32, 239–252.
Beran, M. J., Beran, M. M., & Menzel, C. R. (2005). Spatial memory and monitoring of hidden items through spatial displacements by chimpanzees (Pan troglodytes). Journal of Comparative Psychology, 119, 14–22.
Blades, M., & Cooke, Z. (1994). Young children's ability to understand a model as a spatial representation. Journal of Genetic Psychology, 155(2), 201–218.
Bovet, D., & Vauclair, J. (2001). Judgment of conceptual identity in monkeys. Psychonomic Bulletin and Review, 8(3), 470–475.
Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. Psychological Review, 112(1), 193–216.
Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the learning of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), Language acquisition and conceptual development. Cambridge, UK: Cambridge University Press.
Brown, P., & Levinson, S. C. (2000). Frames of spatial reference and their acquisition in Tenejapan Tzeltal. In L. Nucci (Ed.), Culture, thought and development (pp. 167–197). Mahwah, NJ: Erlbaum.
Byrne, R. W. (1995). The thinking ape: Evolutionary origins of intelligence. New York: Oxford University Press.
Byrne, R. W. (2000). Evolution of primate cognition. Cognitive Science, 24(3), 543–570.
Call, J. (2001). Object permanence in orangutans (Pongo pygmaeus), chimpanzees (Pan troglodytes), and children (Homo sapiens). Journal of Comparative Psychology, 115(2), 159–171.
Call, J. (2004). Inferences about the location of food in the great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus). Journal of Comparative Psychology, 118(2), 232–241.
Call, J. (2006a). Descartes' two errors: reason and reflection in the great apes. In S. Hurley & M. Nudds (Eds.), Rational animals? (pp 219–234). Oxford: Oxford University Press.
Call, J. (2006b). Inferences by exclusion in the great apes: The effect of age and species. Animal Cognition, 9(4), 393–403.
Fagot, J., & Tomonaga, M. (2001). Effects of element separation on perceptual grouping by humans (Homo sapiens) and chimpanzees (Pan troglodytes): Perception of kanizsa illusory figures. Animal Cognition, 4(3–4), 171–177.
Fagot, J., Wasserman, E. A., & Young, M. E. (2001). Discriminating the relation between relations: The role of entropy in abstract conceptualization by baboons (Papio papio) and humans (Homo sapiens). Journal of Experimental Psychology: Animal Behavior Processes, 27(4), 316–328.
Gentner, D. (2003). Why we are so smart. In D. Gentner & S. Goldin-Meadow (Eds.), Language in mind. Cambridge, MA: MIT Press.
Gentner, D., & Rattermann, M. J. O. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), Perspectives on language and thought: Interrelations in development (pp. 225–260). Cambridge, UK: Cambridge University Press.
Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. Cognitive Science, 10(3), 277–300.
Hall, D., & Waxman, S. (1993). Assumptions about word meaning: Individuation and basic-level kinds. Child Development, 64(5), 1550–1570.
Haun, D. B. M., Call, J., Janzen, G., & Levinson, S. C. (2006). Evolutionary psychology of spatial representations in the hominidae. Current Biology, 16(17), 1736–1740.
Haun, D. B. M., Rapold, C. J., Call, J., Janzen, G., & Levinson, S. C. (2006). Cognitive cladistics and cultural override in Hominid spatial cognition. Proceedings of the National Academy of Sciences of the United States of America, 103(46), 17568–17573.
Hauser, M. D., Williams, T., Kralik, J. D., & Moskovitz, D. (2001). What guides a search for food that has disappeared? Experiments on cotton-top tamarins (Saguinus oedipus). Journal of Comparative Psychology, 115(2), 140–151.
Herrmann, E., Call, J., Lloreda, M., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. Science, 317, 1360–1366.
Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1986). Induction: Processes of inference, learning, and discovery. Cambridge, MA: The MIT Press.

Kroger, J. K., Saab, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex, 12,* 477–485.

Levinson, S. C. (2003). *Space in language and cognition.* Cambridge, UK: Cambridge University Press.

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far mappings. *Journal of Cognition and Development, 2*(2), 189–219.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*(4), 315–353.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*(3), 108–114.

Mishra, R. C., & Dasen, P. R. (2005). Spatial language and cognitive development in India: An urban/rural comparison. In B. Chakkarath & Schwarz (Eds.), *Culture and human development: The importance of cross-cultural research to the social sciences.* Hove, UK: Psychology Press.

Neumann, S., & Widlok, T. (1996). Rethinking some universals of spatial language using controlled comparison. In M. D. Puetz & Rene (Eds.), *The construal of space in language and thought* (pp. 345–373). Berlin, GER: Mouten de Gruyter.

Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., & Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proceedings of the National Academy of Sciences, 96,* 5268–5273.

Nunn, C. L., & Barton, R. A. (2001). Comparative methods for studying primate adaptation and allometry. *Evolutionary Anthropology, 10,* 81–98.

Oden, D. L., Thompson, K. R., & Premack, D. (2001). Can an ape reason analogically? Comprehension and production of analogical problems by Sarah, a chimpanzee. In Gentner, D. H., Kokinov, B. N. (Eds.) (pp. 471–498). Cambridge, MA: MIT Press.

Oden, D. L., Thompson, R. K. R., & Premack, D. (1988). Spontaneous transfer of matching by infant chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 14*(2), 140–145.

Oden, D. L., Thompson, R. K., & Premack, D. (1990). Infant chimpanzees spontaneously perceive both concrete and abstract same/different relations. *Child Development, 61*(3), 621–631.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language, 74*(3), 557–589.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109–130.

Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences, 6*(125), 67.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Ramnani, N., & Owen, A. M. (2004). The anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews: Neuroscience, 5,* 184–194.

Ramscar, M., & Pain, H. (1996). *Can a real distinction be made between cognitive theories of analogy and categorisation.* Paper presented at the proceedings of the eighteenth annual conference of the cognitive science society. Hillsdale, NJ.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development, 13*(4), 453–478.

Smith, L. (1984). Young children's understanding of attributes and dimensions: A comparison of conceptual and linguistic measures. *Child Development, 55*(2), 363–380.

Spinozzi, G. (1993). Development of spontaneous classificatory behavior in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 107*(2), 193–200.

Tanaka, M. (1996). Information integration about object–object relationships by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110*(4), 323–335.

Thompson, R. K. R., & Oden, D. L. (1995). A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioural Processes, 35*(1–3), 149–161.

Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: the paleological monkey and the analogical ape. *Cognitive Science, 24*(3), 363–396.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M., & Call, J. (1997). *Primate cognition.* New York, NY, USA: Oxford University Press.

Tomonaga, M., Imura, T., Mizuno, Y., & Tanaka, M. (2007). Gravity bias in young and adult chimpanzees (*Pan troglodytes*): Tests with modified opaque-tubes task. *Developmental Science, 10*(3), 411–421.

Vonk, J. (2003). Gorilla (*Gorilla gorilla* gorilla) and Orangutan (*Pongo abelii*) understanding of first- and second-order relations. *Animal Cognition, 6*(2), 77–86.

Wassmann, J., & Dasen, P. R. (1998). Balinese spatial orientation: Some empirical evidence for moderate linguistic relativity. *The Journal of the Royal Anthropological Institute, 4,* 689–711.