PAPER

Intuitions about gravity and solidity in great apes: the tubes task

Trix Cacchione[1] and Josep Call[2]

1. Department of Psychology, Cognitive and Developmental Psychology, University of Zürich, Switzerland
2. Max-Planck-Institute for Evolutionary Anthropology, Developmental and Comparative Psychology, Leipzig, Germany

Abstract

We investigated whether great apes, like human infants, monkeys and dogs, are subject to a strong gravity bias when tested with the tubes task, and – in case of mastery – what the source of competence on the tubes task is. We presented 22 apes with three versions of the tubes task, in which an object is dropped down a tube connected to one of three potential hiding places and the subject is required to locate the object. In two versions, apes were confronted with a causal tube that varied in the amount of perceptual information it provided (i.e. presence or absence of acoustic cues). The third version was a non-causal adaptation of the task in which a painted line 'connected' dropping and hiding places. Results indicate that apes neither have a reliable gravity bias when tested with the tubes, nor understand the causal function of the tube. Even though there is evidence that they can integrate tube-related causal information to localize the object, they seem to depend mainly on non-causal inferences when searching for an invisibly displaced object.

Introduction

A vertically falling object represents a special case of object displacement for humans as well as for non-human primates. In recent years, several studies have investigated the understanding of invisible displacements involving gravity from a comparative perspective (Hauser *et al.*, 2001; Hauser, Williams, Kralik & Moskovitz, 2001; Hood, Hauser, Anderson & Santos, 1999; Osthaus, Slater & Lea, 2003; Santos & Hauser, 2002; Tomonaga, Imura, Mizuno & Tanaka, 2007). The adoption of paradigms that are used in human infant research offers a direct comparability of the data collected with different species. Human infants, chimpanzees (*Pan troglodytes*), cotton-top tamarins (*Saguinus oedipus oedipus*), rhesus monkeys (*Macaca mulatta*) and dogs (*Canis lupus familiaris*) are all reported to make a 'gravity error': they seem to expect unsupported objects to fall vertically, even if solid obstacles impede their trajectory.

Hood (1995) presented 2- to 4-year old children with a task in which an object is dropped down an opaque tube into one of three potential goal boxes. Two-year-olds repeatedly searched for the object in the box directly beneath the dropping point, even though this box was not connected to the tube. Hood (1995, 1998, 2006) proposed that the error reflects the inability to suppress the naïve theory that all unsupported objects fall in a straight vertical line. This hypothesis is based on the idea that children's vast experience with falling events leads them to acquire a naïve gravity concept (straight down belief) that is later over-generalized to cases in which it cannot be applied appropriately. Hood and colleagues did not find the gravity bias when upward or horizontal motion was presented (Hood, 1998; Hood, Santos & Fieselman, 2000), which suggests that the search error is associated with naïve gravity concepts. In fact, straight down beliefs have been reported in older children and – depending on the task context – even in adults (Bliss, Ogborn & Whitelock, 1989; Kaiser, Profitt & McCloskey, 1985; Krist 2000).

When tested with the same task, monkeys and dogs also show a gravity-biased search. Hood, Hauser, Anderson and Santos (1999) presented adult cotton-top tamarins with a version of the original tube task. Tamarins, like human infants, had a preference to search directly underneath the dropping place, in the goal box specified by the vertical falling line. The tendency to persistently select the gravity box was more marked in tamarins than in human infants. Despite extensive training and cost incentives for correct choices (only one choice was allowed), tamarins continued selecting the wrong gravity box. According to Hood (1995, 1998), the immunity to counter-evidence is a crucial aspect, suggesting that naïve concepts cause the search errors. To achieve sufficient coherence, naïve (and scientific) theories must possess a certain degree of resistance and should not be overturned easily by counter-evidence. That the error is connected to gravity (and does not, for example, reflect a mere alignment

strategy) is further corroborated by the finding that monkeys (as human children) are not biased when tested with a horizontal version (Hauser *et al.*, 2001; but see Southgate & Gómez, 2006 for an alternative account). Osthaus, Slater and Stephen (2003) also found a gravity error in dogs, even though it was less pronounced in dogs than in monkeys. Furthermore, in contrast to the tamarins, dogs eventually overcame the bias and learned where the object was located. However, dogs' success was not based on an understanding of the tubes' mechanism but on remembering where the reward had been found in the previous trial (i.e. they adopted the strategy to perseverate to a previously rewarded location). In any case, these data illustrate that the strength of the gravity error varies across species.

Little is known about how great apes reason about falling objects. Cacchione and Krist (2004) found some evidence that apes are sensitive to the effects of gravity on physical objects, namely, they seem (at least in some situations) to expect that unsupported objects fall. Recently, Tomonaga, Imura, Mizuno and Tanaka (2007) confronted young and adult chimpanzees with a modified tubes task in which subjects had to predict the landing location of objects before they were dropped into two either parallel or crossed tubes. The authors found that the chimpanzees' predictions were biased in the gravity direction. Neither previous experience with tubes nor learning over multiple trials improved their performance. However, this task is not directly comparable with the original tubes task on three counts. First, it was a prediction task: the subjects had to put their hand at the bottom of one of the tubes *before* the object was dropped in order to catch it. Seeing the food reward held out above one of the tubes might have been a strong appeal for the subjects to place their hand at the location directly underneath it. Second, the task of Tomonaga *et al.* (2007) was in some ways more complex, because it involved two tubes whereas the original Level I condition of Hood (1995) (with which human infants and the other non-human species were tested) consisted of only one tube. Third, in the Tomonaga *et al.* task, subjects' choices were more constrained as only two potential goal locations were available, whereas in the original task three goal locations were at choice. On these grounds it is difficult to compare the performance of chimpanzees with that of children, monkeys and dogs.

Findings from our own studies on the table task (Cacchione, Call & Zingg, in press) question the presence of a strong perseverating gravity bias in great apes. The table paradigm was originally implemented using a looking measure (Spelke, Breinlinger, Macomber & Jacobson, 1992) and later adapted for action measures in various species (e.g. Hauser, 2001; Hood, Carey & Prasada, 2000). Participants are confronted with a table-like apparatus and two cups, one placed on the table top and the other below it. Then an object is dropped behind a screen, the screen is raised, and the participant is allowed to search for the cup in which he/she expects the

object to be. If participants understand object solidity and expect that an object cannot move through a solid table they should select the upper cup. Great apes, unlike monkeys, did not show a pervasive gravity bias in this task (Hauser *et al.*, 2001). Even though the apes were guided by naïve gravity concepts (i.e. search failures were related to gravity), they did not show a strong tendency to select the gravity location, suggesting that they may inhibit unwarranted gravity responses. Hauser (2003) suggested that monkeys show a strong gravity bias because they never acquire sufficient inhibitory control. In support of this idea, chimpanzees, bonobos and orangutans generally outperform macaques on inhibitory control tasks (Amici, Aureli & Call, 2008; Rumbaugh & Pate, 1984). This difference, however, does not mark a monkey–ape divide because spider monkeys performed at the same level as chimpanzees, bonobos and orangutans, and better than gorillas and macaques (Amici *et al.*, 2008). However, Southgate and Gomez (2006) suggested an alternative explanation for the macaques' errors based on a preference for grabbing food located in enclosed spaces (i.e. under as opposed to on the table). Although this explanation may apply to macaques, it does not apply to apes because their responses were clearly related to gravity (Cacchione *et al.*, in press).

Admittedly, the table task is less demanding than the tubes task, as evidenced by the data on 2-year-old human infants, who show a gravity bias in the latter but not in the former (Hood *et al.*, 2000a). The reasons for this difference may involve the more complex object trajectories and the greater number of potential hiding places in the tubes task compared with the table task. Because Tomonaga *et al.*'s (2007) version of the tubes task is not directly comparable to the original task, it seems timely to relate great apes' performance to that of human infants, dogs and monkeys in the original tubes task.

The present studies investigate how great apes perform when tested with versions of the original tubes task. The first goal is to determine apes' behaviour when they are faced with the tubes apparatus. Would they, like younger children, monkeys and dogs, fail on this rather complex task, even though they showed no reliable bias in the table task? If so, error analyses may help to define possible causes of failure. It might be that apes simply do not know where the object is; in this case apes' selection of locations would be purely at random. Alternatively, they might prefer specific goal locations – for example, like children, monkeys and dogs, the goal box specified by gravity – suggesting that errors might result from a naïve folk theory of gravity. The second goal is to find out whether apes' behaviour is flexible over a series of trials. Would apes, like monkeys, be immune to evidence about the true location of the object and consistently continue searching for it in the gravity box? Or would they eventually overcome the bias and learn to locate the object (as has been reported for dogs)? Finally, the third goal of the present study is to investigate whether there

are species differences between gorillas (*Gorilla gorilla*), orangutans (*Pongo pygmaeus*), bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*).

In a series of experiments we attempted to find out if great apes appreciate that a causal connection exists between the tube and the future location of an object dropped into it. The first two experiments confronted them with three versions of the tubes task (Experiment 1: acoustic tube condition; Experiment 2: silent and painted tube conditions). The acoustic and silent tube conditions correspond to different levels of abstraction in which the degree of mental reconstruction that is needed to infer the future location of the invisible displaced object is varied. On the lower abstraction level, an opaque tube connects the dropping and goal locations, and the object travelling through the tube is accompanied by a moving sound (Experiment 1: acoustic tube condition). The acoustic cue signals that the object is still travelling inside the tube after its disappearance from view. The participant can track the object's movement acoustically, or, if acoustic cues do not support tracking, infer that an object moving inside the tube comes to rest in the goal box connected to it. (In the latter case, the acoustic cue may serve as a reminder that the object is still moving and that the movement takes place inside the tube; this may reduce the cognitive load to infer the correct location.) The participant need not understand the causal relation of tube and goal location of the object in order to be successful, at least not if he/she manages to track the object acoustically. On the higher abstraction level, participants are confronted with a 'silent' tube, that is, with a situation in which an opaque tube connecting dropping and goal locations is the only information perceptually accessible (Experiment 2: silent tube condition). The participants must mentally reconstruct the trajectory and infer that the tube determines the object's future motion. In the painted tube condition of Experiment 2, apes are confronted with a painted two-dimensional 'tube'. The three-dimensionality of the tube is an important visual cue that signals that the tube can *contain* other objects and is therefore basic to the assumption of a causal relation of tube and object movement. If participants really appreciate the causal function of the tube, they should be aware of the fundamental importance of three-dimensionality as a prerequisite to causal functioning and react accordingly. Finally, Experiments 3 and 4 were run as controls to verify if successful apes causally infer the object location or if they merely adopt non-causal strategies (i.e. by perseverating to a previously rewarded location or by selecting the box with the tube attached).

Experiment 1: Acoustic condition

Experiment 1 investigates the apes' reactions to a scaled-down version of the tubes task (Level I type). An acoustic cue is provided to facilitate the notion that, first, the object (albeit invisible) is actually in motion, and,

second, the movement occurs inside the tube. Can apes successfully locate the invisibly displaced object using the acoustic information? If not, we are interested to see: (a) if more gravity than non-gravity errors occur, and (b) if performance eventually improves. Finally, we examine if all ape species perform in a similar manner.

Method

Participants

Six gorillas (*Gorilla gorilla*), five orangutans (*Pongo pygmaeus*) and five bonobos (*Pan paniscus*) participated in Experiment 1. All apes were housed at the Wolfgang Köhler Primate Research Center in Leipzig Zoo, Germany. All subjects except one gorilla had prior experience with various experiments investigating physical cognition. Apes were tested alone either in an indoor observation room or in their sleeping room. Mothers with children younger than three years of age were tested in the company of their offspring.

Apparatus

The apparatus was modelled after Hood's (1995) original apparatus (see Figure 1). It consisted of a grey plastic frame and a transparent rear panel made of plexiglas (height 42.5 cm × length 68.2 cm × depth 11.3 cm). In the upper section, three opaque plastic chimneys were attached. In the lower section, three opaque goal boxes (height 9.8 cm × length 9.1 cm × depth 9.4 cm) were placed directly below the chimneys. Chimneys and goal boxes could be connected by opaque tubes (length 74 cm × diameter 3 cm). The tubes were made of pieces of a flexible hose. Hazelnuts were used as food reward to drop down the tubes. The travelling of the hazelnut through the tubes was accompanied by a moving sound. A
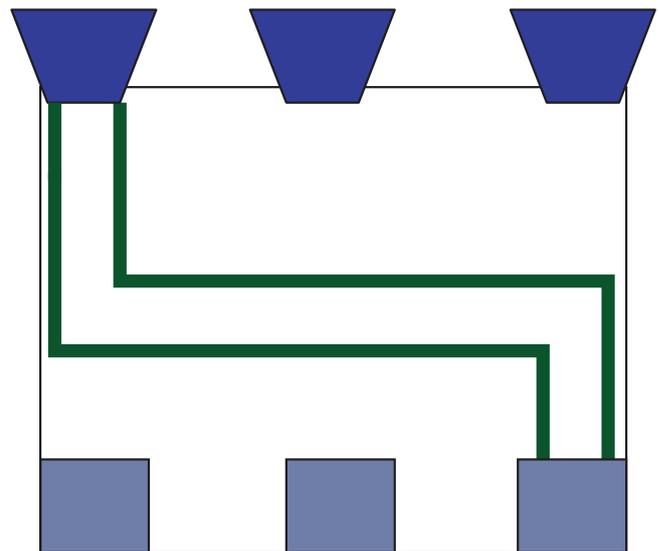

Figure 1 *Schematic depiction of the tubes apparatus.*

deformable cotton mat absorbed the sound of the reward arriving in the goal box.

Design and procedure

There were two experimental phases: baseline and testing (consisting of two baseline conditions and one test condition). All conditions were run by an experimenter who presented the stimuli. Both baseline and test conditions were presented in 9-trial sessions. In both baseline conditions the sessions were repeated until the ape was correct in eight out of nine trials. The apes then proceeded to the acoustic tube condition. The acoustic test condition included three sessions consisting of nine trials each. If an ape was correct in seven out of nine trials in one of the three sessions it was judged to have passed the acoustic test condition. All apes were presented with only one baseline or test session per day.

The baseline conditions served to familiarize the apes with the apparatus, to exclude baseline preferences for a specific goal box, and to make sure that apes can direct search to a new goal box after having repeatedly found the food in another location. In the first baseline condition the reward was placed successively in each goal box in a random fashion. In the second baseline condition the reward was placed in the same goal box for three consecutive trials, then three times in a second goal box and finally three times in the remaining goal box. Again the order of the boxes was at random.

The procedure in the first baseline condition was as follows. The ape sat behind a plexiglas panel. A testing surface (slide table) was fixed by a metal frame directly underneath the panel and the apparatus placed on top of it. The plexiglas panel had three holes through which the ape could point at each of the goal boxes. The experimenter sat in front of the slide table. The slide table was constructed such that its surface could be moved back and forth. The experimenter pulled the table back and placed a hazelnut in full view of the ape in one of the goal boxes and closed it. Then she pushed the apparatus against the panel. The ape could now point to the goal box in which he/she expected the food to be hidden. The experimenter pulled the table back again. If the ape had chosen correctly she opened the box and gave the reward to the ape. If he/she had pointed to the wrong goal box the experimenter pushed the table back against the panel to allow for another choice until the ape found the reward. If the ape was correct in eight out of nine trials he/she proceeded to the second baseline condition. If not, the first baseline condition was repeated. The procedure of the second baseline was identical to that of the first, except that the hazelnut was now placed in the same goal box on three consecutive trials and then changed location. If the ape was correct in eight out of nine trials he/she proceeded to the acoustic tube test. If not, the second baseline condition was repeated.

In the acoustic condition, half of the participants were presented with a single tube connecting the upper left chimney with the lower right goal box, whereas the other half were tested with the opposite tube configuration. The experimenter pulled the table back and waved the reward above the chimney. If the ape looked at it, the experimenter released the nut. Then the apparatus was pushed to the panel and the apes were allowed to indicate their choice by pointing to a goal box. The experimenter now pulled back the apparatus. If the ape had selected correctly, the experimenter took out the reward and gave it to him/her. If the wrong box was selected the experimenter pushed the apparatus back to the panel and the ape was allowed to choose again until he/she found the nut or abandoned the search.

Results

Figure 2 presents the frequency of each of the responses in the first trial for each session. Five of 16 apes chose the correct box on the first attempt in the first trial (binomial test: $p = .55$). However, there was no significant bias towards particular responses in the first trial of the first ($\chi^2 = 2.38$, $df = 2$, $p = .31$) or third ($\chi^2 = 4.63$, $df = 2$, $p = .099$) sessions. In contrast, subjects produced significantly fewer middle errors in the second session ($\chi^2 = 6.13$, $df = 2$, $p < .05$). This means that subjects did not show a gravity bias, although they were not above chance either. There was no significant improvement between the first and the third session in the frequency of correct first-trial responses (McNemar test: $p = .34$, $n = 16$). Focusing solely on the errors collapsed across sessions revealed that subjects made significantly more gravity than middle errors (Wilcoxon test: $z = 3.21$, $p < .001$).

Figure 3 presents the percentage of correct and gravity-biased responses for all the trials (not just the first trial) in each of the three sessions. In none of the three sessions did apes have an above-chance preference for the gravity box. Instead, they reliably preferred the
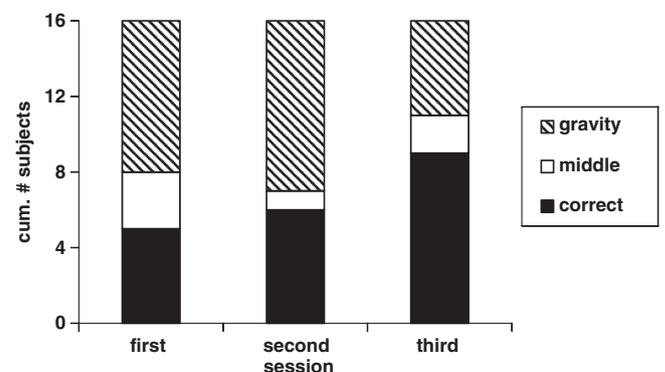
Figure 2 *Frequency of each of the responses in the first trial for each session in Experiment 1.*
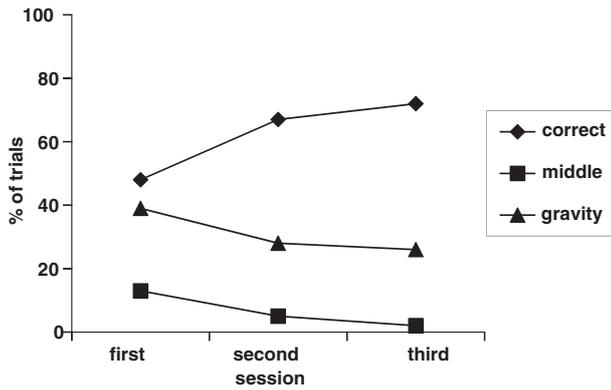
Figure 3 *Percentage of correct and incorrect responses for all nine trials in each of the three sessions in Experiment 1.*

correct box in all three test sessions (Wilcoxon tests: S1: $z = 2.17$, $p < .05$; S2: $z = 3.15$, $p < .01$; S3: $z = 3.19$, $p < .01$). There was a significant improvement in the percentage of correct responses across sessions (Friedman test: $\chi^2 = 7.56$, $df = 2$, $p < .05$, $n = 16$) but no significant reduction in the percentage of gravity-biased responses (Friedman test: $\chi^2 = 4.75$, $df = 2$, $p = .093$, $n = 16$). Although initially only 18.8% of the apes reached criterion (7 out of 9 correct) in the first session, later on 87.5% of the apes reached criterion in at least one of the three sessions ($\chi^2 = 9.0$, $p < .01$). Only two gorillas never reached criterion during the three test sessions of the acoustic condition.

In general there were no species differences, except that in the second session gorillas produced significantly fewer correct responses (Kruskal–Wallis test: $\chi^2 = 6.12$, $p < .05$) and more gravity errors (Kruskal–Wallis test: $\chi^2 = 6.29$, $p < .05$) than other ape species.

Discussion

Apes were initially unable to locate a nut dropped down an opaque tube even though acoustic cues accompanied its movement through the tube. Even though apes selected the correct box more often than expected by chance, on average fewer than 50% of the choices were correct during test session one. Only a small proportion of the sample reached criterion in the first test session, and the first trial performance was at random. This suggests that, at least initially, apes do not understand the function of the tube. However, unlike children,

[1] The two gorillas that did not reach criterion in the acoustic test condition were run on a transparent tube test. The procedure in the transparent condition was the same as in the acoustic condition except that a transparent (not an opaque) tube connected the chimney and the goal box, so that the nut moving through the tube was now – in addition to having an acoustic cue due to movement sound – fully visible. Both selected the correct box in the first trial of transparent test. Furthermore, both reached criterion in the transparent condition with an average of 94.5% of correct choices.

monkeys, dogs, and the chimpanzees in the Tomonaga *et al.* (2007) task, apes in the current study did not show a significant preference for the gravity box. Nevertheless, apes selected the gravity box in the first trial more often than both other goal boxes and incurred more gravity than non-gravity errors. Over trials the performance reliably improved, and finally the majority of the apes chose the correct goal box[1]. Likewise, the number of erroneous choices declined. This suggests that, even though apes are vulnerable to the gravity error, they were able to inhibit perseverative gravity responses to select the correct goal box.

It is unclear what type of information apes used to solve the task. One possibility is that acoustic information provided by the falling reward was responsible for the successful performance. Even though the apes' performance was initially low, it might have been even lower in the absence of acoustic cues. In fact, it is possible that the apes that chose correctly in the first session were relying on acoustic information. Another possibility is that participants learned to select a particular box (i.e. by adopting the strategy to perseverate to a previously rewarded location). Finally, the possibility that apes eventually realized the causal function of the tube cannot be completely ruled out. In the next experiment we investigated the contribution of each type of information to success.

Experiment 2: Silent and painted conditions

Successful localization of the object in Experiment 1 could have been due to: (a) the use of acoustic information, (b) perseveration to a previously rewarded location, (c) learning to search the location with the tube attached, or (d) an understanding of the tubes' causal function. In Experiment 2 apes were confronted with two conditions to investigate the possible causes of correct performance: first, a 'silent' tube condition in which object movement is both invisible and non-audible; and second, a two-dimensional painted tube condition in which no physical (causal) connection exists between the tube shape and the goal box. Note that in contrast to the silent condition the painted condition has no solution, at least not if search behaviour is based on logical inferences. Contrasting these two conditions generated the following predictions. If correct performance was due to the use of acoustic information, apes should (at least initially) fail to localize the object in the silent condition (as well as in the painted condition). If perseveration to a previously rewarded location was the reason for correct performance, apes should locate the object equally well in the real and in the painted condition. Finally, if correct performance was due to searching the box with the tube attached to it, or even associated with a causal understanding of the tube, apes should find the object in the silent test condition, but not (at least not initially) in the painted condition.

Method

Participants

The same 16 apes that participated in Experiment 1 were included in Experiment 2. In addition, six chimpanzees (*Pan troglodytes*) housed at the Wolfgang Köhler Research Center (Leipzig Zoo) in Germany were also tested. All apes had prior experience with various experiments investigating physical cognition.

Apparatus

In the silent tube condition, the same apparatus as in Experiment 1 was used with one alteration: the upper part of the tube was blocked by a barrier so that the nut inserted in the chimney rested invisibly inside the tube and did not travel through it. In the painted condition, the same apparatus as in the silent condition was used but instead of the tube a plexiglas panel was inserted flush with the rear side panel. On the panel, a two-dimensional tube was painted with black gloss paint. Once it was inserted, the painted panel appeared to 'connect' the upper left chimney with the lower right goal box (or vice versa). An oblong screen (height 83.5 cm × length 68.7 cm) could be mounted in front of the apparatus so that the hazelnut could be removed between trials.

Procedure

Before testing, chimpanzees were first presented with the same two baseline conditions as described in Experiment 1. Then all apes received the silent condition and the painted condition. Half of the apes were tested with the silent condition first, and the other half were first run on the painted condition. Both conditions were composed of three 9-trial sessions. If the apes were correct in 7 out of 9 trials in one of the three sessions the test condition was judged as being passed. All apes were presented with only one test (or baseline) session per day. Half of the apes were presented with an upper-left–lower-right configuration, and the other half with the opposite configuration. The apes that had already been tested in Experiment 1 were presented with the same configuration as in Experiment 1.

The basic testing procedure was the same as in the testing conditions of Experiment 1, with some exceptions as noted below. In the silent condition, the experimenter presented the tube and inserted it inside the apparatus in full view of the ape, and followed the same steps as in the acoustic condition of Experiment 1. The procedure in the painted condition was identical to that in the silent condition, except that there was no tube connecting a chimney and a box. Instead, the plexiglas with the painted pattern was held in front of the ape and rotated to make sure that he/she appreciated that it was a two-dimensional pattern. Then it was inserted in full view of the ape into the apparatus and the test began. Note that, even though the painted condition had no correct solution, subjects were rewarded for selecting the box at the end of the painted line. This was done to make the functionality of the tube (not the reinforcement regime) the only difference between the test conditions. This allowed us to attribute all performance differences to the difference in tube functionality. (For the sake of simplicity, the box at the end of the painted line is below referred to as the 'correct box', and responses to this box as 'correct responses'.) The search behaviour was recorded on a check-sheet, and all sessions were videotaped.
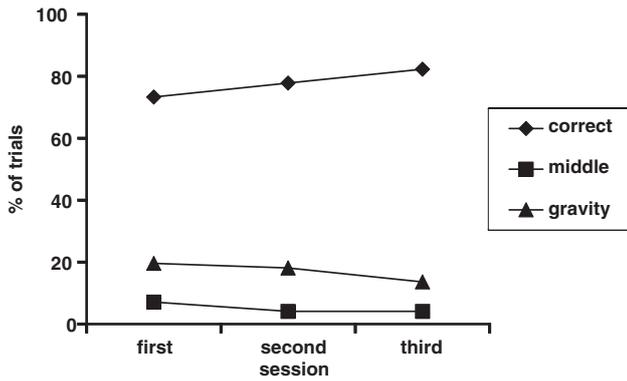
Results

Silent condition

Figure 4 presents the frequency of each of the responses in the first trial for each session. Fourteen of the 22 apes chose the correct box on the first attempt in the first trial (binomial test: $p < .01$). Subjects also selected the correct box above chance in the second and third sessions (binomial test: $p < .05$ in both cases). There was no significant improvement between the first and the third session in the frequency of correct responses (McNemar test: $p = .69$, $n = 22$). Focusing solely on the errors collapsed across sessions revealed that subjects made significantly more gravity than middle errors (Wilcoxon test: $z = 2.58$, $p < .01$).

Figure 5 presents the percentage of correct and incorrect responses for all the trials in each of the three sessions. In none of the three sessions did apes have an above-chance preference for the gravity box. Instead they reliably preferred the correct box in all three test sessions (Wilcoxon tests: S1: $z = 3.76$, $p < .001$; S2: $z = 3.94$, $p < .001$; S3: $z = 4.04$, $p < .001$). There was no significant improvement in the percentage of correct responses across sessions (Friedman test: $\chi^2 = 4.51$, $df = 2$, $p = .11$, $n = 22$) and no significant reduction in the percentage of gravity errors (Friedman test: $\chi^2 = 2.39$, $df = 2$, $p = .30$, $n = 22$). Sixty per cent of the apes reached criterion (7 out of 9 correct) in the first session, whereas 81.8% of the apes reached criterion in at least one of the three sessions
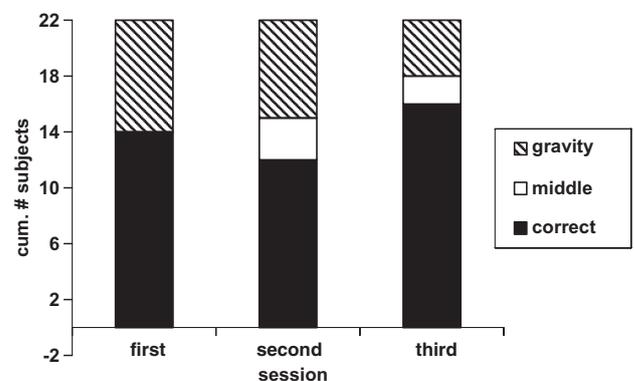
Figure 4 *Frequency of each of the responses in the first trial for each session in the silent condition of Experiment 2.*
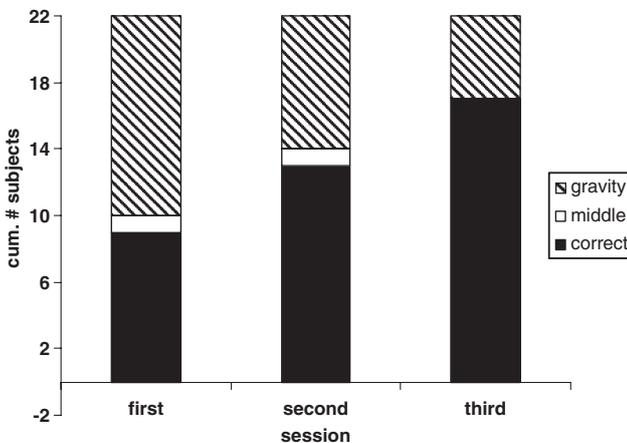
Figure 5 *Percentage of correct and incorrect responses for all the trials in each of the three sessions in the silent condition of Experiment 2.*

(χ^2 = 8.909, p < .01). In general, gorillas and chimpanzees tended to perform worse than orangutans and bonobos in session one (Kruskal–Wallis test: χ^2 = 7.442, p = .059). No significant differences were detected in the other two sessions (Kruskal–Wallis tests: χ^2 < 6.59, p >.085 in both cases).

Painted condition

Figure 6 presents the frequency of each of the responses in the first trial for each session. Nine of 22 apes chose the correct box on the first attempt in the first trial of the painted condition and 12 chose the gravity box (binomial test, p < .05). In sessions two and three of the painted condition, apes chose reliably correctly in the first trial (binomial test, p < .01 and p < .001). There was a significant improvement between the first and the third sessions in the frequency of correct responses (McNemar test: p < .05, n = 22). Focusing solely on the errors collapsed across sessions revealed that subjects made significantly more gravity than middle errors (Wilcoxon test: z = –3.099, p < .01).
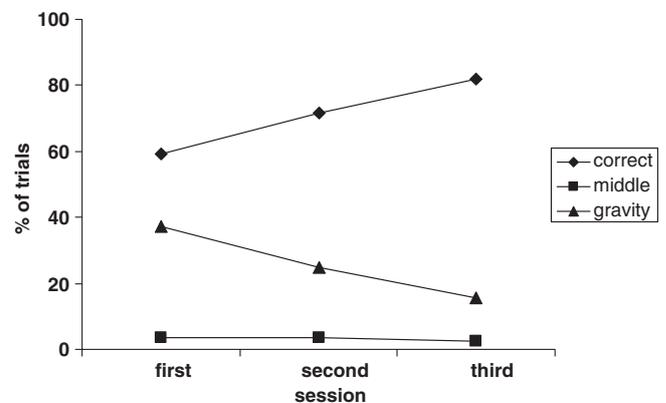
Figure 7 presents the percentage of correct and incorrect responses for all the trials in each of the three sessions. The initial above-chance preference for the gravity box was only apparent within the first three trials of session 1 (Wilcoxon tests: z = 3.19, p < .01). In none of the three sessions did apes prefer the gravity box over all nine trials. Instead they reliably preferred the correct box in all three test sessions (Wilcoxon tests: S1: z = 2.92, p < .01; S2: z = 3.72, p < .001; S3: z = 4.11, p < .001). There was a significant improvement in the percentage of correct responses across sessions (Friedman test: χ^2 = 17.73, df = 2, p < .001, n = 22) and a significant reduction in the percentage of gravity errors (Friedman test: χ^2 = 15.75, df = 2, p < .001, n = 22). Forty-one per cent of the apes reached criterion (7 out of 9 correct) in the first session, whereas 68.2% of the apes reached criterion in at least one of the three sessions. There were no significant species differences in any of the three sessions (Kruskal–Wallis tests: χ^2 < 5.52, p >.135 in all cases).

Acoustic versus silent condition

For the apes that participated in both Experiment 1 and Experiment 2 (orangutans, gorillas and bonobos), we compared the percentage of correct choices (all trials) of the third session of the acoustic test and the first session of the silent test to evaluate if performance declined when the acoustic cue was no longer available. This comparison revealed no significant differences (Wilcoxon test: z = .904, p = .366, n = 16).

Discussion

Apes succeeded in the silent condition and failed to choose the prior rewarded box in the painted condition. This result supports the hypothesis that subjects selected the box to which the tube was attached and rules out that apes used acoustic cues to solve the task. Apes' performance in the last acoustic and the first silent test

Figure 6 *Frequency of each of the responses in the first trial for each session in the painted-tube condition of Experiment 2.*


Figure 7 *Percentage of correct and incorrect responses for all the trials in each of the three sessions in the painted-tube condition of Experiment 2.*

session did not differ. So, whatever apes learned during Experiment 1, it was not linked to the presence of acoustic cues. When confronted with the silent tube, two-thirds of the apes reached criterion in the first test session, and the great majority of them within three sessions. Subjects reliably chose the correct box from the first trial, and, when mistaken, they produced many more gravity than non-gravity responses. There were some species differences, with chimpanzees and gorillas performing worse than the other species. However, chimpanzees' worse performance can be explained by their lack of prior experience with Experiment 1. As testing progressed their performance mirrored that of the rest of the sample in Experiment 1; that is, initially they showed no preference between boxes, made more gravity than non-gravity errors, and learned quite quickly to locate the object. The cause of the poor performance of gorillas is, however, less clear. Apparently they generally have more difficulty with the tubes task (even in Experiment 1, gorillas' performance was worse than that of other species) and profited less from experience with the task.

Apes clearly distinguished between the silent and the painted condition. They chose correctly more often in the silent than in the painted condition and made more gravity errors in the painted than in the silent condition. This is remarkable, because apes showed no such gravity bias either in the acoustic condition of Experiment 1 or in the silent condition in Experiment 2, and it suggests that the gravity response occurs as a default when no other information about the location of the object is available (this issue is addressed further in the general discussion). In order to assess whether the order of presentation of the conditions (silent–painted versus painted–silent) affected performance, we compared the first trial of those subjects who received the silent condition first with the first trial of those subjects who received the painted condition first. Seven apes in the silent condition chose the correct box and four the gravity box, whereas three chose the correct box in the painted condition, seven the gravity and one the middle box. Moreover, chimpanzees, who by virtue of not being included in Experiment 1 had no prior experience with the acoustic tube, performed in similar ways to the other apes in the silent and painted conditions (see Figure 8), thus further reinforcing the idea that acoustic cues had little effect on performance and that the painted condition produced more gravity-biased responses than the silent condition.

However, performance improved significantly across all three sessions of the painted condition. Unlike the results of the silent condition, there were no significant differences between species in the painted condition. Although apes responded differently to the 3D than to the 2D tube it is still unclear whether apes made this distinction based on non-causal information (such as the pattern) or on causal information regarding the functionality of the tube. The following three experiments
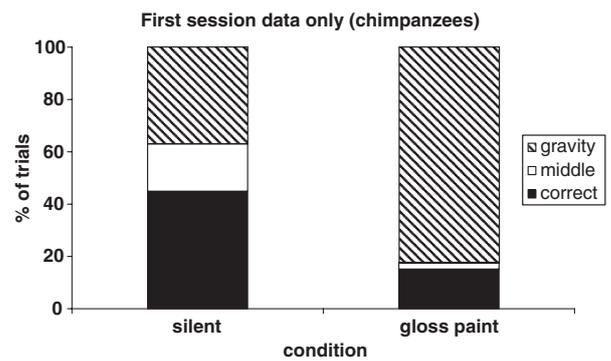
Figure 8 *Frequency of each of the responses in all trials of the first session comparing silent- and painted-tube conditions (chimpanzees only).*

further investigated if apes prefer to select the box (a) previously reinforced, (b) with the tube attached to it, or (c) specified by the tubes' causal function.

Experiment 3: Reversed condition

In this experiment apes are presented with the opposite tube configuration to that of Experiments 1 and 2 (reverse condition). If location learning is the cause of correct choices, apes should fail to locate the object. Successful location, on the other hand, indicates either the use of a more elaborate strategy based on selecting the box with the tube attached but without understanding its functionality, or a true understanding of the function of the tube. To differentiate between gravity biases and the strategy to perseverate to a previously rewarded location apes were presented with an upper-middle-box–lower-outside-box configuration. The lower outside box corresponded to the one they had not received anything from in Experiments 1 and 2.

Method

Participants

All 19 apes that passed the silent condition of Experiment 2 participated in the current experiment. This meant that three gorillas were excluded.

Apparatus

The apparatus was the same as that used in the silent condition of Experiment 2.

Procedure

The procedure was the same as in the silent condition of Experiment 2, except that apes were presented with an upper-middle-box–lower-outside-box configuration. Those who were run on Experiments 1 and 2 with an upper-left–lower-right configuration were now presented

with an upper-middle–lower-left configuration and vice versa. Apes received one 9-trial session. If the ape was correct in 7 out of 9 trials it was judged to have passed the test.

Results

Six out of 19 apes chose the correct box on the first attempt in the first trial (not different from the 6.3 correct responses expected by chance). Twelve selected the previously rewarded location (binomial test, $p < .01$), and in one case the middle (now gravity) box was selected. There were no species differences in this respect. On average, 68.53% of the choices were correct throughout trials 1 to 9. Again there was a significant improvement in the percentage of correct choices between trials 1–3 and trials 7–9 (Wilcoxon test: $z = -3.355$, $p < .001$), but no significant reduction in the percentage of gravity-biased responses (Wilcoxon test: $z = -.577$, $p = .56$). In Experiment 3, 57.9% of the apes reached criterion.

Discussion

The majority of apes selected the previously rewarded location in the first trial. Thus, the majority of apes used the strategy to perseverate to a previously rewarded location, not tube connectivity, as the main cue to locate the object. However, the results of the painted condition of Experiment 2 cast some doubt on the idea that the strategy to perseverate to a previously rewarded location is the only factor determining the subjects' responses. If apes merely adopted this strategy they would not differ in their reactions to the real and the painted tube. Those findings further suggest that apes may have at least some intuitive sensitivity to the functional properties of the tubes' mechanism and are able to integrate this information to make inferences about the future location of an invisibly displaced object. This cannot be a very elaborate understanding and has only a minor influence on box selection, as apes are initially unable to find the object. Finding the reward repeatedly in the same box may have exerted a powerful influence on the subjects' choices, although subjects easily abandoned the strategy to perseverate to a previously rewarded location when it became unrewarded. Experiment 4 further investigates this issue by confronting the apes with a situation in which subjects can choose a box connected to a painted or a real tube.

Experiment 4: Real versus painted tube

In this experiment apes that passed Experiment 3 are confronted with both a real and a painted tube simultaneously. To find the object apes must differentiate between a causal and a non-causal tube and cannot rely on the strategy to perseverate to a previously rewarded location.

Method

Participants

Only the 11 apes (2 gorillas, 3 orangutans, 3 bonobos and 3 chimpanzees) that passed Experiment 3 were tested in Experiment 4.

Apparatus

The apparatus was the same as in Experiment 3 except that a tube painted on plexiglas was inserted into the apparatus. Thus, there were two tubes (one painted and one real) connecting the upper central chimney with the two extreme bottom locations.

Procedure

The procedure was identical to that of Experiment 3 except that the side (left or right) to which the real and painted tubes were connected was switched after each trial. All apes started with an arrangement in which the real tube was inserted as in Experiment 3 (e.g. upper middle–lower left) and the painted tube the opposite way (e.g. upper-middle–lower-right). Subjects received one 9-trial session. Apes that searched correctly in 7 out of 9 trials were judged to have passed the test.

Results and discussion

Three of the 11 apes chose the correct box in the first trial (this does not differ from the 3.6 correct responses expected by chance). Seven chose the box that was previously rewarded in Experiments 1 and 2 (binominal test, $p < .05$), and only one ape chose the middle (now gravity) box. On average, 48.36% of the choices were correct throughout trials 1 to 9. Performance improved over trials. Again there was a significant improvement in the percentage of correct choices between trials 1–3 and trials 7–9 (Wilcoxon test: $z = -2.013$, $p < .05$) but no significant reduction in the percentage of gravity-biased responses (Wilcoxon test: $z = -1.633$, $p = .10$). In Experiment 4, 27.3% of the apes reached criterion.

The majority of the apes did not take into account the type of tube (real or painted) that connected the top with the bottom portion of the apparatus. This is exactly what can be expected if apes continue to select the previously rewarded location. The middle (gravity) box was only rarely selected. However, the number of correct choices significantly increased, indicating that some apes changed their search strategy throughout the test session. Finally, three apes passed criterion for Experiment 4. They reliably chose the box that was connected to the real tube.

General discussion

Great apes did not display the strong gravity bias in the tubes task that has been reported in 2-year-old human

infants, monkeys and dogs (Hood, 1995; Hood *et al.*, 1999; Osthaus *et al.*, 2003). The current results also differ markedly from those in which chimpanzees were tested with a modified version of the tubes task (Tomonaga *et al.*, 2007). Even though apes selected the gravity box more often than the correct and the middle box in the first trial, this preference did not reach statistical significance and was abandoned in a relatively short time. Similar to dogs but unlike human infants and cotton-top tamarins, apes were able to modify their choices depending on the reward's location. In fact, most apes succeeded in this task within three sessions if not earlier.

However, this does not mean that gravity responses are non-existent in apes. Most observed errors could be accounted for by a gravity bias. Moreover, when apes were confronted with a situation in which no real tube connected the top and bottom parts of the apparatus (and hence no logical solution could be inferred) they responded in ways consistent with a gravity bias, which in this case was possibly the only sensible reaction (other than using the painted pattern to infer the reward's location). This suggests that apes do hold a naïve belief about gravity, but it is dormant and reappears only under specific circumstances. This would corroborate the idea that when development progresses the gravity bias is not erased, but suppressed because the capacity to inhibit prepotent responses increases. These results fit in with our findings with the table task (Cacchione *et al.*, in press), in which apes performed generally well but gravity-related factors affected the probability for a reappearance of gravity-biased responses. Recently Hood, Wilson and Dyson (2006) showed that the gravity bias also reappears in 4-year-old children if the recruitment of inhibitory mechanisms is hindered by increasing the cognitive load that the task poses on infants. The findings of Tomonaga *et al.* (2007) can also be interpreted along these lines. It is conceivable that both the presence of two tubes and the holding of the reward over one of the tubes while subjects chose one of the alternatives made the task more demanding, which resulted in a greater prevalence of gravity-biased errors. Thus, we suggest that apes hold naïve gravity beliefs, but, in contrast to monkeys, are generally able to inhibit unwarranted gravity responses. However, gravity-biased responses may reappear depending on the complexity of the task.

The present study may offer some hints regarding apes' intuitions about the causal function of the tube, but they are not substantial enough to support a true understanding of the task. Apes located the object more often when they were presented with a 3D tube compared with a 2D painted one. This difference is especially remarkable because subjects received the painted test after they had already learned to select one of the boxes in the course of previous experiments. The two tube versions differ in causal and perceptual properties, but only the difference in causal properties can explain differential success in retrieving the object. Only the real tube can allow the reward to move undetected from the top to the bottom of the apparatus. It affords, unlike the painted tube, a causally plausible event. The painted condition confronts the ape with an event without a causally plausible solution, just an arbitrary one. Here apes can at best guess where the reward is located and eventually learn to find it by establishing the strategy to perseverate to a previously rewarded location or by selecting the box in contact with the painted line. If the apes experienced only a perceptual difference between the real and painted tubes, they would in neither situation have valuable information about the object's current location. The observed difference between the 3D and the 2D tube may suggest first that they are to some extent aware of the causal function that the tube exerts on the object's trajectory, and second that they are to some extent aware that the spatial extension is the fundamental property that supports the causal functioning. However, these intuitions seem to be of a very weak nature because they do not lead to consistent correct solutions in this task. This suggests that both naïve gravity concepts and intuitions about the causal function of the tube have only a weak impact on apes' behaviour in this task context, leaving room for the adoption of various non-causal strategies.

The main strategy employed by the apes in this study was the strategy to perseverate to a previously rewarded location, the same strategy as adopted by dogs in the Osthaus *et al.* (2003) study. This might be an artifact of the repeated reinforcement, because apes (just like dogs) received multiple trials with the same tube configuration and therefore were repeatedly rewarded to choose a specific goal box. That apes (and dogs) succeeded with this procedure is interesting per se, because other non-human primates did not, but instead perseverated in choosing the gravity location. Note, however, that developing the strategy to perseverate to a previously rewarded location did not result in a fixed pattern of responding because such a strategy was quite easily abandoned when it became ineffective (e.g. in Experiments 3 and 4). It is unclear whether subjects replaced the strategy to perseverate to a previously rewarded location with other strategies such as selecting the box with the tube attached or whether they simply learned another position bias. Future studies are required to answer this question.

Finally, we identified some between-species differences as well as substantial within-species variability in this task. In general, gorillas performed worse than the other species –they made more errors and they more often failed to reach criterion. This fits well with data on inhibition, in which gorillas perform worse than the other great apes, particularly chimpanzees (Amici *et al.*, 2008), and thus lends some credence to the idea that gravity errors result mainly from an inability to inhibit

prepotent responses (Hauser, 2003). However, this species difference was outweighed by the inter-individual variance – some apes were initially gravity-biased, whereas others chose correctly even in the first trial. Not all apes adopted the same strategy to locate the object, and not all were able to change strategy in a similarly flexible way.

In conclusion, the tubes task seems to be much more demanding for great apes than the table task, which they generally solved well in a previous study (Cacchione *et al.*, in press). However, despite their lack of understanding of the tubes, great apes did not show the strong gravity bias (including in the very first trial) that has been described in previous studies. Moreover, apes seem to exhibit some sensitivity to the causal relation between the tube, the object motion and the goal location. Apes were, however, unable to adopt their intuitions to locate an object dropped down the tube but mostly relied on the strategy to perseverate to a previously rewarded location. Further research should address why apes' causal sensitivity remains subliminal and under what circumstances it could be used more effectively .

References

Amici, F., Aureli, F., & Call, J. (2008). Fission–fusion dynamics, behavioural flexibility and inhibitory control in primates. *Current Biology*, **18**, 1415–1419.

Bliss, J., Ogborn, J., & Whitelock, D. (1989). Secondary school pupils' common sense theories of motion. *International Journal of Science Education*, **II**, 261–272.

Cacchione, T., & Krist, H. (2004). Recognizing impossible object relations: intuitions about support in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, **118**, 140–148.

Cacchione, T., Call, J., & Zingg, R. (in press). Gravity and solidity in great apes: the table task. *Journal of Comparative Psychology*.

Hauser, M.D., & Santos, L.R. (2002). A non-human primate's understanding of solidity: dissociations between seeing and acting. *Developmental Science*, **5**, F1–F5.

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 oedipus) *Journal of Comparative Psychology*, **115**, 140–151.

Hood, B.M. (1995). Gravity rules for 2-4-year-olds? *Cognitive Development*, **10**, 577–598.

Hood, B.M. (1998). Gravity does rule for falling events. *Developmental Science*, **1**, 59–64.

Hood, B.M., Hauser, M.D., Anderson, L., & Santos, L. (1999). Gravity biases in a nonhuman primate? *Developmental Science*, **2**, 35–41.

Hood, B.M., Carey, S., & Prasada, S. (2000a). Predicting the outcomes of physical events. *Child Development*, **71**, 1540–1554.

Hood, B.M., Santos, L., & Fieselman, S. (2000b). Two year-old's naïve predictions for horizontal trajectories. *Developmental Science*, **3**, 328–332.

Hood, B. M., Wilson, A., & Dyson, S. (2006). The effect of divided attention on inhibiting the gravity error. *Developmental Science*, **9**, 303–308.

Kaiser, M.K., Profitt, D.R., & McCloskey, M. (1985b). The development of beliefs about falling objects. *Perception & Psychophysics*, **38** (6), 533–539.

Krist, H. (2000). Development of naive beliefs about moving objects: the straight-down belief in action. *Cognitive Development*, **15**, 281–308.

Osthaus, B., Slater, A.M., & Lea, S.E.G. (2003). Can dogs defy gravity? A comparison with the human infant and a non-human primate *Developmental Science*, **6**, 489–497.

Rumbaugh, D.M., & Pate (1984). The evolution of cognition in primates: a comparative perspective. In H.L. Roitblat, T.G. Bever & H.S. Terrace (Eds.), *Animal cognition* (pp. 403–420). Hillsdale, NJ: Erlbaum.

Southgate, V., & Gómez, J. C. (2006). Searching beneath the shelf in macaque monkeys: evidence for a gravity bias or a foraging bias? *Journal of Comparative Psychology*, **120**, 314–321.

Spelke, E.S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, **99** (4), 605–632.

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

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