ORIGINAL ARTICLE

Inferences by exclusion in the great apes: the effect of age and species

Josep Call

Abstract This study investigated the ability of chimpanzees, gorillas, orangutans, and bonobos to make inferences by exclusion using the procedure pioneered by Premack and Premack (Cognition 50:347–362, 1994) with chimpanzees. Thirty apes were presented with two different food items (banana vs. grape) on a platform and covered with identical containers. One of the items was removed from the container and placed between the two containers so that subjects could see it. After discarding this item, subjects could select between the two containers. In Experiment 1, apes preferentially selected the container that held the item that the experimenter had not discarded, especially if subjects saw the experimenter remove the item from the container (but without seeing the container empty). Experiment 3 in which the food was removed from one of the containers behind a barrier confirmed these results. In contrast, subjects performed at chance levels when a stimulus (colored plastic chip: Exp. 1; food item: Exp. 2 and Exp. 3) designated the item that had been removed. These results indicated that apes made inferences, not just learned to use a discriminative cue to avoid the empty container. Apes perceived and treated the item discarded by the experimenter as if it were the very one that had been hidden under the container. Results suggested a positive relationship between age and inferential ability independent of memory ability but no species differences.

J. Call (✉)
Max Planck Institute for Evolutionary Anthropology,
Deutscher Platz 6,
04103 Leipzig, Germany
e-mail: call@eva.mpg.de

One of the most important functions of cognition is to deal with the problem of inconsistent or incomplete information that animals often encounter in their environment. Inferential reasoning, defined as associating a visible and an imagined event (Premack 1995), is one of the processes that can be used to deal with fragmentary information. One of the most interesting types of inference is the so-called inference by exclusion, which consists of selecting the correct alternative by logically excluding other potential alternatives. Several studies have shown that chimpanzees, sea lions, and dolphins can solve inferences by exclusion using different paradigms (Herman et al. 1984; Schusterman et al. 1993; Schusterman and Krieger 1984; Tomonaga 1993, see also Watanabe and Huber 2006). For instance, in the fetching paradigm, subjects are presented with a set of familiar items and one novel item. Each of the familiar items has been associated with a label which, depending on the studies, can be either a visual or an auditory stimulus. Then subjects are requested to fetch the 'X' (i.e., the unknown label that designates the novel object). Results show that subjects select the novel item when requested the unfamiliar label, thus making the inference that the novel label refers to the novel unfamiliar item.

Analogous results have been obtained with a matching to sample paradigm in which subjects had to select an unfamiliar alternative after witnessing a novel sample (Hashiya and Kojima 2001; Schusterman et al. 1993; Tomonaga 1993). Hashiya and Kojima (2001) showed positive results for a chimpanzee in a cross-modal matching to sample paradigm. They presented the subject with two pictures of people that she knew and the voice of one of them. The chimpanzee's successfully matched the voice with the correct picture. Next,

the authors presented the chimpanzee with two pictures (one of someone she knew and the other of a stranger) and an unfamiliar voice. The chimpanzee correctly matched the unfamiliar voice to the unfamiliar picture.

Although subjects use the new label to select the new item, it is unclear whether animals also learn to associate the new label with the novel object. Schusterman et al. (1993) argued that animals used labels to discriminate between the alternatives but they did not learn to use those labels to refer to particular objects, which is precisely what studies on language acquisition in children have shown. Children can learn a new label for a new object (when presented with familiar alternatives) and later use that label to refer to the novel object. Recently, this skill has also been found in a 9-year-old dog (Border collie) that can fetch more than 200 different objects by their verbal label (Kaminski et al. 2004).

Going beyond the information provided by perceptual inputs not only enables subjects to acquire new associations, or at least distinguish between alternatives, it can also enable them to make efficient foraging decisions when searching for food. Numerous studies have documented how several species can infer the position of food items, which they have not seen directly based on the trajectories that they may have followed (see Doré and Dumas 1987; Tomasello and Call 1997 for reviews, but see Collier-Baker et al. 2004, 2006, for a critical analysis of some of those studies). A recent study reported that apes can find the location of food without directly perceiving the food or the displacements that it may have followed, but use indirect information to infer its location (Call 2004). Apes were presented with two opaque containers with one of them baited. Then, the experimenter shook the empty container and lifted the baited one. Subjects selected the baited container above chance levels even though there was no auditory cue emanating from any of the containers. Control conditions showed that subjects did not solve the problem by using inadvertently given cues from the experimenter or the food (e.g., smell or noise produced during baiting). This means that apes were able to infer the location of reward by the noise it would have made if it had been in a given location. Bräuer et al. (2006) found no evidence suggesting that dogs made such inference when presented with the same problem as the apes.

Premack and Premack (1994) tackled the question of inferential reasoning using a different paradigm. They presented four 3- to 4-year-old chimpanzees with two boxes and two types of fruit (banana and apple). Chimpanzees were allowed to witness the experimenter deposit each fruit in one of the boxes so that both boxes were baited. Later, subjects saw the experimenter eating one of the fruits (e.g., banana) without seeing the experimenter removing the food from the box. The question was whether given the opportunity to select one of the boxes, they would avoid the box in which the

experimenter had deposited the food that he was currently eating, presumably because it was now empty. Premack and Premack (1994) found that one of the chimpanzees, the oldest one, solved this problem from the first trial, suggesting that she was able to infer that if the experimenter was eating the banana, the box where the banana was deposited would be empty. Two other chimpanzees failed the first two and four trials, respectively, before responding correctly in a consistent manner for the remaining trials. One of the chimpanzees failed the problem because she always selected the box that had contained the food that the experimenter ate. Thus, these results indicated that there was some evidence of inference by exclusion in at least one of the chimpanzees. The alternative to this inferential strategy would be to learn to use the sight of the banana in the experimenter's possession as a discriminative sign for choosing the box with the other reward. Although this alternative cannot easily explain the performance of the chimpanzee that selected the correct container above chance from the beginning, it may explain the performance of the two chimpanzees that initially failed the test. Furthermore, the last chimpanzee did not solve the problem, and therefore, it is unclear how easily can chimpanzees use this learning strategy.

The current study had three main objectives. First, it investigated whether a successful performance in this task could be explained as a result of conditional discrimination. Instead of inferring the solution, subjects may have learned to associate the presence of a certain food type with the appropriate response without understanding that the food that the experimenter was eating was the same food that was inside the box. For instance, if the experimenter was eating the banana, the subjects could have learned to select the grape box and vice versa. Although Premack and Premack (1994) acknowledged that two of the chimpanzees that they tested may have solved the problem in this way, the authors did not directly test this possibility. Therefore, the current study included tests aimed at finding out whether subjects would be able to learn to associate the presence of certain cues with producing particular responses instead of using inferential abilities. Two critical indicators that would suggest that inference rather than learning a conditional discrimination was involved would be a faster acquisition in the inferential compared to the association conditions, and no evidence of improvement during the test.

Second, this study also investigated two possible causes of the relative low performance of some of the chimpanzees tested by Premack and Premack (1994), given that chimpanzees (and other apes) have shown evidence of inference by exclusion in other paradigms (e.g., Call 2004; Hashiya and Kojima 2001; Tomonaga 1993). One obvious reason for this result could be that the subjects tested by Premack and Premack (1994) were too young (3–4 years of age). Therefore, the current study investigated the effect of age

on inferential reasoning by including apes of various ages. Another reason for the observed low performance may have been that there was an extended period of time between baiting the boxes and allowing subjects to select one of them. Indeed, the time elapsed since subjects last saw the reward has been postulated as one of the factors that may contribute to the difficulty of invisible object displacements compared to visible displacements (e.g., de Blois et al. 1999; Doré et al. 1996). Therefore, there were two types of inference task in which the timing when subjects could witness the reward removal from the box was varied. Subjects were expected to perform better if they were allowed to see how one of the food pieces was removed from the box compared to finding the reward after all manipulations had been completed.

Third, most of the research on ape cognition is based on chimpanzees. It is unclear how other apes would perform in this task. Although two previous studies have found no important differences in inferential reasoning among the great apes species (e.g., Barth and Call 2006; Call 2004), these results need to be confirmed. Therefore, all four great apes species were included in this study.

Experiment 1: Displacements included

This experiment tested apes in three different tasks corresponding to three conditions. There were two types of inference task that differed in the timing when the object that was removed from the box was shown to the subject. In one task, the object was shown as soon as it was removed from the container while in the other task, subjects had to wait until all manipulations had been completed. The third task assessed whether subjects would be able to solve the task by learning a conditional discrimination, in which the color of a plastic chip indicated the food that remained intact after the manipulation had been completed.

Methods

Subjects

Seven orangutans, seven chimpanzees, five gorillas, and five bonobos housed at the Wolfgang Köhler Research Center, Leipzig Zoo (Germany) participated in this study (see Table 1). There were 15 females and 9 males, with age ranging from 4 to 32 years. All male bonobos and all adult

Table 1 Name, species, age, sex, rearing history, and the experiments in which subjects participated

Name	Species	Age (years)	Sex	Rearing history	Experiment participation
Annett	Chimpanzee	6	F	Nursery	3
Alex	Chimpanzee	4	M	Nursery	3
Alexandra	Chimpanzee	6	F	Nursery	3
Brent	Chimpanzee	4	M	Mother	1 and 3
Dorien	Chimpanzee	25	F	Nursery	1–3
Fraukje	Chimpanzee	29	F	Nursery	1–3
Pia	Chimpanzee	6	F	Mother	1 and 3
Jahaga	Chimpanzee	12	F	Mother	3
Fifi	Chimpanzee	12	F	Mother	3
Sandra	Chimpanzee	12	F	Mother	1–3
Gertrudia	Chimpanzee	12	F	Mother	3
Frodo	Chimpanzee	12	M	Mother	1–3
Patrick	Chimpanzee	8	M	Mother	1–3
Gorgo	Gorilla	24	M	Nursery	1–3
Bebe	Gorilla	26	F	Unknown	1–3
Ndiki	Gorilla	26	F	Unknown	1–3
Viringika	Gorilla	10	F	Mother	1–3
Ruby	Gorilla	8	F	Mother	1–3
Joey	Bonobo	23	M	Nursery	1–3
Ulindi	Bonobo	12	F	Mother	1–3
Limbuko	Bonobo	10	M	Nursery	1–3
Kuno	Bonobo	9	M	Nursery	1–3
Yasa	Bonobo	8	F	Mother	1–3
Dunja	Orangutan	32	F	Unknown	1–3
Bimbo	Orangutan	25	M	Mother	1–3
Pini	Orangutan	17	F	Mother	1–3
Walter	Orangutan	16	M	Mother	1 and 2
Dokana	Orangutan	16	F	Mother	1–3
Toba	Orangutan	11	F	Mother	1 and 2
Padana	Orangutan	8	F	Mother	1–3

Fig. 1 Setup for Experiment 1 (inference condition)

chimpanzees were nursery-reared whereas all other subjects were mother-reared. All subjects lived in social groups of various sizes, with access to indoor and outdoor areas. Subjects were individually tested in their indoor cages and were not deprived of food or water.

Materials

Two blue opaque bins (25 cm × 13 cm × 13 cm) were placed on a plastic sliding platform about 40 cm apart (see Fig. 1). A small barrier (11 cm × 8.5 cm) with its top and sides covered by plastic pieces (to prevent subjects from seeing if a reward was behind it) was placed between the two bins forming a straight line. Depending on the conditions, this barrier was made of gray or clear plastic, rendering it either opaque or transparent, respectively. Grapes and pieces of banana were used as rewards. A green and a blue plastic chip (3 cm × 2 cm) served as discriminative stimuli for the association condition. The test materials were presented on a sliding platform situated behind a plexiglass partition that separated the subject from the experimenter. This plexiglass partition had two circular holes cut on its bottom that allowed subjects to touch one of the blue bins.

Procedure and design

There were three phases: pre-test, test, and post-test. The pre- and post-test phases were identical and assessed the preference for one of the two types of food. Such information was needed in case a strong preference for some types of reward may have interfered with test performance. The experimenter (E) sat facing the subject behind the platform separated by a plexiglass partition and placed a banana slice on one side of the platform and a grape on the other side, closest to the subject but still outside of reach. In half of the trials the grape was on the left side and the banana on the right side and vice

versa for the other half trials. After the subject had witnessed the reward placement, the E covered each reward with a blue bin and pushed the sliding platform forward against the plexiglass partition so that the subjects could touch one of the bins located in front of the holes. The first container touched by the subject was scored as his/her choice. Subjects received 12 trials in the pre- and post-test phases.

During the test phase, the E also placed two rewards on the platform and covered each with a blue bin. Then he placed the barrier in the center of the platform between the blue bins. The E executed the following sequence of movements in all conditions aimed at removing one of the rewards from one of the bins: E reached inside the left bin, closed his hand, pulled his hand out, and moved it to the center barrier. There he opened his hand, pulled it out from the barrier and showed it empty to the subject. Finally, he reached inside the second bin, again closing his hand and bringing it behind the barrier. In the perceptual condition, the E extracted the reward and moved to the barrier while visibly holding the reward whereas in the other two conditions he kept his hand closed so that subjects were not able to see the reward. After this sequence was completed, the E removed the barrier from the platform so that the subject was able to see the object that the experimenter had left behind it. After the subject had seen the reward, the E discarded the reward by throwing it into a bucket, and pushed the sliding platform forward so that the subjects could touch one of the bins located in front of the holes. The first container touched by the subject was scored as his/her choice.

There were three conditions that differed depending on the type of barrier that the experimenter used (clear or opaque) and the type of object that subjects saw after removing the barrier:

Perceptual: The barrier was clear and subjects witnessed one of the two rewards on the platform before being discarded.
Inference: The barrier was opaque and subjects witnessed one of the two rewards on the platform before being discarded.
Association: The barrier was opaque and upon its removal subjects witnessed one of the two plastic chips that indicated the reward that remained under one of the two bins. The color associated with each reward was counterbalanced between subjects, this means that for some subjects the blue chip signaled the presence of grapes under of the bins while for others it signaled the presence of the banana slice. Since establishing an arbitrary association between the colored chip and the presence of one of the rewards is a demanding task, half of the subjects were tested with a simplified version of the task that consisted of leaving the plastic token on the platform while subjects made a choice.

Each subject received two 12-trial sessions per condition (24 trials total). The order of the conditions was counterbalanced across subjects so that the six possible orders were presented to the same number of subjects. The two sessions corresponding to a given condition were tested sequentially, with no other session between them. The position of the type of reward (left vs. right) was randomly determined with the only two restrictions that it appeared the same number of times on each side and could not appear more than two times in succession on the same side.

Data analysis

The main analyses investigated whether the condition, species and age affected performance. The three conditions were also compared against chance both at the group and at the individual level. I also investigated whether there was any evidence of improvement during testing. I did not assess inter-observer reliability because subjects' choices could be determined without uncertainty. I used parametric statistics except for those analyses on the first trial (one sample chi-square, sign test) and individual performance (binomial test). All tests were two-tailed.

Results

Subjects showed no significant preference for grapes over banana pieces during the pre-test, $t_{23} = 1.21$, $p = 0.24$, or the post-test, $t_{23} = 0.52$, $p = 0.61$. Moreover, preferences were quite consistent between the pre-test and the post-test, $r = 0.57$, $p = 0.004$, $n = 24$. Finally, there was no significant relation between grape preference (calculated as the joint preference for grapes over banana in the pre-test and post-test) and any of the conditions (perception: $r = -0.17$, $p = 0.42$; association: $r = 0.16$, $p = 0.45$; inference: $r = 0.20$, $p = 0.34$; $n = 24$ in all cases). Thus, there was no evidence that food preference affected the subjects' performance.

Figure 2 presents the mean percent of correct trials in each condition. A 3×4 ANOVA with condition as within-subject factor and species as between-subject factor revealed a significant effect for condition, $F_{2,40} = 33.85$, $p < 0.001$, and no significant effect for species, $F_{3,20} = 2.31$, $p = 0.11$ or condition $\times$ species, $F_{6,40} = 0.73$, $p = 0.63$. Bonferroni–Holm post hoc tests (Holm 1979) revealed that subjects performed significantly better in the perception compared to the inference and association conditions ($p < 0.001$ in both cases). Subjects also performed significantly better in the inference compared to the association condition ($p = 0.014$). Moreover, subjects as a group were above chance in the perception, $t_{23} = 9.12$, $p < 0.001$, and inference conditions, $t_{23} = 2.64$, $p = 0.015$, but not in the association condition, $t_{23} = 0.54$, $p = 0.59$. Such results were already evident in the first trial. Subjects selected the correct container above
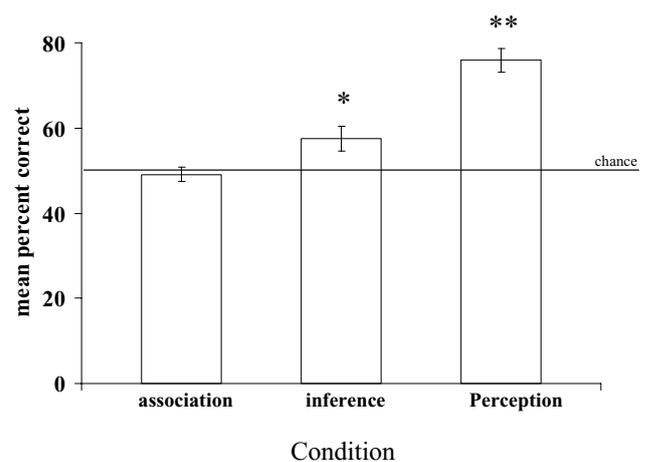

Fig. 2 Mean percent of correct trials ($\pm$ SEM) in each of the three conditions of Experiment 1 ($^*p < 0.05$, $^{**}p < 0.01$)

chance in the perception ($\chi^2 = 16.67$, df $= 1$, $p < 0.001$) and inference conditions ($\chi^2 = 4.17$, df $= 1$, $p = 0.041$) but not in the association condition ($\chi^2 = 0.17$, df $= 1$, $p = 0.68$).

There was no evidence of an improvement in performance throughout testing either when comparing the first and the last trial in each condition (sign test: perception, $p = 0.5$; inference, $p = 0.39$; association, $p = 0.29$) or when comparing the first and the second session for each condition (perception: $t_{23} = 0.11$, $p = 0.91$; inference: $t_{23} = -2.35$, $p = 0.028$; association: $t_{23} = 0.0$, $p = 1.0$). In fact, the performance of subjects in the inference condition significantly worsened.

Individual analyses showed that 12, 5, and 0 subjects were above chance (Binomial test: $p < 0.05$) in the perceptual, inference, and association conditions, respectively. There was no significant relation between age and any of the conditions (perception: $r = 0.01$, $p = 0.96$; association: $r = -0.07$, $p = 0.76$; inference: $r = 0.32$, $p = 0.13$).

Discussion

Subjects performed above chance (and from the start) in the perception and inference conditions, but not in the association condition. Moreover, subjects performed significantly better in the perception compared to the inference condition. There were no significant effects of species or age on the percent of correct responses and no evidence of improvement over trials in any of the conditions. In fact, subjects' performance worsened in the inference condition, perhaps due to the high attentional demands of the task.

Subjects performed above chance in those conditions in which they saw which food item was being discarded. This means that subjects understood that the bin that had contained that food item was now empty – an inference that they were able to make even when they did not directly see the food being pulled out from the bin but found it later once the opaque barrier was removed. Nevertheless, being able to see

the reward being pulled out of the bin (without seeing the bin empty) substantially improved the subjects' performance, as indicated by the clear difference between the perception and inference conditions. It is very likely that a reason for such improvement was that subjects did not have to remember the location of the reward that had been pulled out. Such a difference between conditions in the current study is reminiscent of the data on visible and invisible displacements found in the literature (e.g., Call 2001; Pepperberg and Funk 1990; Mendes and Huber 2004). Subjects perform better when they are allowed to see (as opposed to not see) whether the reward is still being displaced between containers.

It is precisely the absence of the food on the platform after the opaque barrier was removed that may have increased the difficulty of the association condition, thus compromising the comparison across conditions. Recall that the discriminative stimulus in the association condition (i.e., colored chip) was not as attractive as the one used for the inference or perception conditions (i.e., food reward). This means that subjects may have paid less attention to the plastic chip than to the food item on the platform. It is possible that if the discriminative cue had been a food item, subjects would have paid more attention. However, the absence of the food item at the end of the manipulation may not be the only factor that contributed to the low performance in the association condition because there was a decrease in performance over trials in the inference condition. This may indicate that the current procedure in which the experimenter removed one of the rewards may have taken too long to complete and subjects got distracted during its implementation.

There were no species or age differences in the inference or perception conditions. This fits with previous results in which no major differences were found between great ape species (Call 2004). One possible reason for this result, especially regarding the lack of an age influence, is that I was unable to control for memory differences independently from inferential ability. If some adults had worse memory than that of some youngsters, this would have influenced their performance in the inference test and masked any possible age effects.

The next two experiments introduced several variations to solve the potential problems raised in the current experiment regarding the high attention demands (produced by a long experimental manipulation and an unattractive discriminative cue in the case of the association condition) and the lack of a memory control test.

Experiment 2: Increasing the salience of the discriminative cue

This experiment addressed the issue of providing an attractive discriminative cue that subjects could use to make a correct decision. The colored plastic chips of Experiment 1 were replaced by apple pieces or peanuts to indicate the presence (or absence) of the rewards at the time of choice.

Methods

Subjects

There were 22 apes in this experiment. I tested all apes included in the previous study except the chimpanzees Brent and Pia, which were not available at the time of testing.

Materials

I used the same bins and opaque barrier as in Experiment 1 but replaced the two colored plastic chips with two food types (peanut and apple) that served as discriminative stimuli.

Procedure

The procedure was the same as in the Association condition of Experiment 1 except that one of two food types instead of one of two colored chips signaled the permanence of one or the other reward under the bins. Such discriminative cue was removed before subjects were allowed to select one of the bins. Although the type of discriminative stimuli was initially counterbalanced across subjects, the failure of several subjects to participate made the groups unbalanced. In all, 9 and 13 subjects received the peanut as an indication that the grape and the banana remained, respectively. The position of the reward (left–right) was counterbalanced within subjects across trials. Subjects received two 12-trial sessions with the same constraints as in the original test regarding the position of the reward. Data were analyzed in the same way as in Experiment 1.

Results

Subjects did not find the reward above chance, $t_{21} = 0.28$, $p = 0.78$. Individual analyses confirmed this result as nobody performed above chance (Binomial test: $p > 0.05$). The type of discriminative cue had no effect on the percent of correct trials, $t_{20} = 0.62$, $p = 0.54$. There was no evidence of improvement throughout testing either when comparing the first and the last trial (sign test: $p = 0.34$) or the first and the second sessions, $t_{21} = 1.20$, $p = 0.24$.

Discussion

Subjects performed at chance levels with no evidence of improvement over trials. This means that the use of neutral stimuli in the last experiment cannot alone explain those negative results. The current experiment also highlights that

learning a conditional discrimination in which one type of food indicated which reward was available is hard. Indeed, Nissen et al. (1948) found that it took chimpanzees hundreds or even thousands of trials to master conditional discrimination. These negative results contrast with the positive results of the inference test in the previous experiment, and it could suggest that mastering a conditional discrimination is harder than is solving the problem by inferential reasoning.

One could argue that the difference between tests did not reside in the type of cognitive process involved (conditional discrimination vs. inferential reasoning) but in the type of stimuli used. Note that in the current experiment the presence of one type of food (e.g., peanut) between the two bins indicated the absence of a different type of food (e.g., banana) from under one of the bins. In contrast, in the previous inference test, the type of food present between the bins also indicated its absence from one of the bins.

I addressed this problem in the next experiment by making the food type that act as the same discriminative cue as the food type that would be removed from under one of the bins. This manipulation also made the association and inference tests more comparable to each other because the discriminative cue in both tests was identical. Additionally, I simplified the bin manipulations, thus reducing the presentation time and added a condition to assess memory retention.

Experiment 3: Memory test and simplified procedure

This experiment addressed two outstanding issues that made the interpretation of the two previous experiments problematic. First, subjects received a memory test to screen out those subjects that may have failed the inference test because they experienced difficulties remembering critical information regarding the food locations. Second, the procedure was simplified by reducing the time devoted to manipulating the stimuli and implementing an association test that used the same edible discriminative stimuli as those used in the inference test.

Methods

Subjects

There were 28 apes included in this experiment. I tested all apes included in Experiment 1 except Walter and Toba (which were not available at the time of testing) and six additional chimpanzees (one male, five females) ranging in age from 4 to 12 years of age that had not participated in any of the previous experiments. These additional chimpanzees belonged to a second group of chimpanzees housed at Wolfgang Köhler Research Center, Leipzig Zoo. Their housing

conditions were comparable to the apes tested in previous experiments. Although these chimpanzees were also experienced in cognitive testing, they had not received Experiments 1 and 2. The inclusion of these subjects allowed us to assess whether previous experience on this kind of task facilitated the performance in the current experiment.

Materials

Two white opaque cups (9.5 cm × 7 cm) placed upside down on a sliding platform were used. An opaque plastic screen was used to block the subjects' visual access to some experimental manipulations. Grapes, pieces of banana, or monkey chow were used as rewards.

Procedure and design

The procedure was similar to that used in previous experiments. Namely, the experimenter placed two different rewards on the platform covered with cups and removed one of them before letting the subject choose one of the cups. Subjects received four conditions depending on the number of rewards placed initially on the platform and the timing in removing them from it:

Inference (I): The E placed a banana piece and a grape on opposite sides of the platform in full view of the subject and covered them with the white cups. Then, he interposed the opaque screen between the subject and the cups, lifted the left cup with one hand while moving the other hand toward the uncovered reward first and then to the center of the platform while at the same time replacing the left cup in its original position. This procedure was repeated with the cup on the right side. In half of the trials, the experimenter removed the reward from the left side while in the other half of the trials he removed the reward from the right side. Then, the experimenter removed the screen revealing the removed reward in the center of the platform that was discarded after the subject had seen it.

Association (A): The E placed one banana slice and a grape on opposite sides of the platform in full view of the subject and covered them with the white cups. Additionally, he placed a third reward on the center of the platform that corresponded to the reward that the experimenter would later removed from under one of the cups. Thus, if the E was intending to remove the banana, he placed another banana on the center. This manipulation was aimed at investigating whether subjects would use that as a signal for the reward that would be missing after the E's manipulation. Then, he interposed the screen and carried out the same food removal procedure mentioned in the previous condition except that before removing the screen, the E took out one of the two rewards that were now on the center

of the platform (recall that one reward had been placed there before the screen was raised and the other resulted from removing it from under one of the cups). So when the E removed the screen, there was only one reward on the center of the platform that he discarded after the subject had seen it.

Memory (M): The E placed one banana and one grape on opposite sides of the platform in full view of the subject and covered them with the white cups. Then, he lifted one of the cups, removed the reward (in full view of the subject) and discarded it. The E then interposed the screen and waited for the same amount of time that it took him to complete the food removal manipulation in the previous conditions. After this period had elapsed, the E lowered the screen and allowed the subject to select one of the two cups on the platform. The aim of this condition was to assess whether subjects were able to remember the reward that had been removed. More importantly, this condition allowed us to screen out those subjects that may have failed the previous two conditions due to inattention or memory failure.

Control: The E placed two cups on the platform, interposed the screen, and showed the subject a food reward (either a grape or a piece of chow). Then, he placed the reward under one of the two cups behind the screen. The baiting procedure consisted of lifting each cup in succession and depositing the reward under one of then. After completing this procedure, the E removed the screen and allowed the subject to select one of the cups. This condition assessed whether subjects may be solving previous conditions by using inadvertently given cues either from the experimenter, the food, or the baiting procedure.

Each subject received one session of 20 trials. The first 16 trials were devoted to the inference (four trials), association (four trials), and memory (eight trials) conditions. Each condition was administered in four-trial blocks. Thus, in a given session there was one inference and one association block and two memory blocks. Inference and association trials never followed each other but were always preceded or followed by memory trials. In particular, there were four testing orders (I-M-A-M, M-I-M-A, A-M-I-M, and M-A-M-I) counterbalanced across subjects so that the four possible orders were presented to the same number of subjects. The last four trials of the testing session (trials 17–20) were devoted to the control condition because if subjects had learned to use inadvertently given cues, the last few trials were deemed the best trials to discover it. The position of the type of reward (left vs. right) was randomly determined with the only two restrictions that it appeared the same number of times on each side and could not appear more than two times in succession on the same side.

Data analysis

The memory condition was used to screen out subjects that may not be paying attention to the experimenter's manipulations or were not able to remember critical events accurately. Subjects that scored less than six trials correctly (out of eight) were excluded from further analyses. After the initial screening was completed, it was investigated whether condition, species, or age had an effect on performance. Finally, I also analyzed whether there was any evidence of improvement during testing, including the assessment of first trial performance.

Results

Overall, subjects selected the correct container in 84.8% (SEM = 3.2) of the trials in the memory condition, which is well above chance, $t_{27} = 10.97$, $p < 0.001$. However, there were six subjects (two gorillas, two bonobos, one orangutan, and one chimpanzee, all of which had participated in the previous experiments) that scored less then six correct responses (out of eight) in this condition and were dropped from subsequent analyses. Figure 3 presents the mean percent of correct trials in each condition. A 3×4 mixed ANOVA investigated the effect of species and condition on the percentage of correct trials. Since sphericity could not be assumed for the condition factor (Mauchly's $W = 0.543$, df $= 2$, $p = 0.006$), the degrees of freedom were adjusted using the Huynh–Feldt index. There were significant differences across conditions, $F_{2,31} = 8.20$, $p = 0.002$, but no effect of species, $F_{3,38} = 1.23$, $p = 0.33$, or species $\times$ condition, $F_{5,31} = 0.23$, $p = 0.95$. Bonferroni–Holm post hoc tests revealed that subjects performed significantly better in the inference compared to the association ($p = 0.015$) and control conditions ($p = 0.016$). There were no significant differences between the association and the control condition ($p = 0.39$). Moreover,
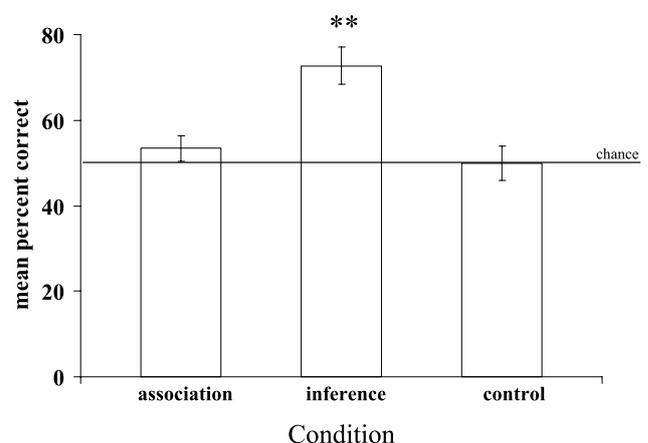

Fig. 3 Mean percent of correct trials ($\pm$ SEM) in each of the three conditions of Experiment 3 (*$p < 0.05$, **$p < 0.01$)
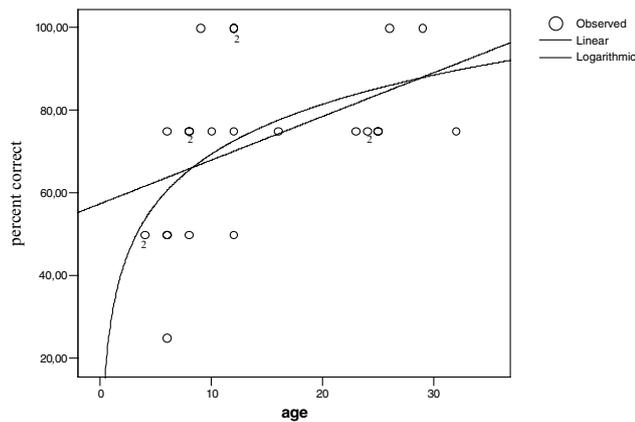
Fig. 4 Percent of correct trials in the inference condition of Experiment 3 as a function of age. Numerals denote those data points that represents two subjects instead of just one

subjects were above chance in the inference condition, $t_{21} = 5.26$, $p < 0.001$, but not in the association, $t_{21} = 1.14$, $p = 0.27$, or in the control condition, $t_{21} = 0.0$, $p = 1.0$.

These results were already evident in the first trial. Subjects performed significantly better in the inference compared to the association condition (sign test: $p = 0.022$, $n = 13$). Although 15 of the 22 subjects selected the correct alternative on the first trial of the inference condition, this was not statistically significant ($\chi^2 = 2.91$, df $= 1$, $p = 0.088$). There was, however, no significant improvement over trials when comparing the first with the last trial (sign test: $p = 0.29$, $n = 22$). I also investigated the effect that previous experience with this task may have on performance in the inference condition by comparing those chimpanzees that had previously participated in Experiment 1 with those that had only participated in the current experiment. There were no significant differences between those two groups, $t_{10} = 1.60$, $p = 0.14$, especially after matching the groups by age by removing the two adults (Fraukje and Dorien) that had only participated in Experiment 1, $t_8 = 1.08$, $p = 0.31$.

Figure 4 presents the number of correct responses in the inference condition as a function of age. Due to the non-linear distribution of the variables depicted in Fig. 4, the variable age was log transformed and a significant correlation between the number of correct responses and age emerged, $r_{21} = 0.53$, $p = 0.011$. This correlation was still significant after controlling for the performance in the memory condition, $r_{19} = 0.56$, $p = 0.008$. No significant relation was found between age and the score in the association condition with, $r_{19} = 0.21$, $p = 0.37$, or without controlling for the memory score, $r_{21} = 0.20$, $p = 0.37$.

Discussion

The results of the current experiment replicated those of Experiment 1. Subjects found the reward above chance in the

inference but not in the association or in control conditions. This means that subjects were not solving the task by learning to associate the sight of one reward with the selection of the other reward, or by using inadvertently given cues. Furthermore, the ability to establish inferences appeared to increase with age. Subjects below 8 years of age rarely scored above 75% correct in the inference test. This effect was independent of memory ability, which in the current study remained constant regardless of age (although this may have happened because only subjects above 3 years of age were included in the study). Finally, there was no evidence of improvement in the inference condition either within this experiment or across experiments by comparing naïve and experienced subjects in the inference task administered in Experiment 1.

A comparison between inference and association conditions is important. These two conditions were identical regarding what the subjects saw just prior to selecting one of the containers—a food item informed them of what they should avoid. The difference, of course, was created during the setting up of the experiment. In the inference condition, there were only two items, one under each cup, whereas in the association condition there was an additional item placed on the center of the platform that could have been used as discriminative cue. Subjects consistently performed better in the inference compared to that in the association condition, a result that was independent from memory ability and that was evident from the beginning of testing.

Subjects performed better in the inference condition of the current experiment compared to the inference condition of Experiment 1. In fact, the inference condition of Experiment 3 was equivalent to the perceptual condition of Experiment 1. Such an improvement cannot be attributed to a practice effect because the six naïve chimpanzees that only participated in Experiment 3 performed at the same level as the others. Moreover, recall that there was no evidence of improvement due to practice in Experiment 1. There are two other factors that may have contributed to the observed improvement. First, the task administration substantially reduced the attentional demands on the subjects because there were less containers and no displacements. Second, the memory condition in Experiment 3 allowed us to screen out those subjects that may have suffered from attention or memory deficits during the task. This may have boosted the scores, although recall that subjects performed at the same level in the inference condition of Experiment 3 and the perception condition of Experiment 1.

One could argue that the greater number of items in the association condition may have imposed greater cognitive demands than the inference condition, thus explaining the poorer performance in the former. Alternatively, the sight of the food item on the platform prior to raising the screen may have distracted the subjects, thus impairing their memory of the items' locations. However, recall that in the previous

experiments included in the current study, there were the same number of items in all conditions, and subjects still performed better in the inference compared to the association conditions.

General discussion

Subjects of all great ape species were capable of making inferences by exclusion involving two different food items. After one item was hidden in one of two bins, and they later saw one of those items being discarded, they selected the bin containing the non-discarded item. There was a positive relation between inferential ability and age, which was independent from species and memory ability. In contrast, apes were unable to solve the same problem using conditional discrimination with a variety of discriminative cues. Several implications can be derived from these results.

First, it appears that subjects found the food using an inferential process, not one based on conditional discrimination in which they learned to avoid the bin that contained the discarded food. In fact, there was no evidence of learning in any of the three experiments. Moreover, if very fast learning explains the performance in the inference conditions, it is unclear why subjects did not learn in any of the three association conditions in each of the three experiments. This lack of learning would fit Schusterman et al.'s (1993) proposal that subjects in exclusion paradigms do not learn to associate the novel label to refer to unfamiliar stimuli, but they used labels to discriminate between the alternatives available. Similarly, in the current task subjects did not learn to associate certain cues with certain outcomes, they were using the cue to decide between alternatives. Therefore, it is highly unlikely that conditional discrimination is responsible for the current findings.

One could postulate that subjects may have learned to associate the discriminative cue (i.e., the sight of a particular food item) with the appropriate response (i.e., selection of the container that held the other food item) in the past. However, this possibility is unlikely for several reasons. First, and most importantly, it is unclear why subjects did not also apply such a rule in the association condition of Experiment 3. Recall that in that experiment, subjects witnessed the same discriminative stimulus in both the inference and association conditions at the time of choice. Yet, they only performed above chance in the inference condition. One way to further rule out a previous history of conditional discrimination would be to present subjects with novel foods in this setup and see how they respond. Second, it is very unlikely that subjects had experienced a situation in which they witnessed a human (let alone another ape) hide two different types of food under two different containers and later they observed the human discard one of the food items without

seeing from where the food item had been extracted. This is simply something that apes do not typically experience in our facility and the current experiment was the first time that they received this kind of test. Moreover, our results confirm that conditional discrimination involving arbitrary stimuli, which is often invoked to explain a variety of phenomena, is extremely difficult to master for apes (e.g., Nissen et al. 1948).

Second, the current findings also have important implications for current discussions on object representation in animals. Indeed, using inference by exclusion in the current setup requires subjects to encode 'what is where' information because there were two locations and two types of food item. Moreover, it also requires subjects to be capable of object individuation based on object features (Santos et al. 2002; van de Walle et al. 2000; Xu and Carey 1996) because subjects have to understand that the object that the experimenter discarded was the same that was in the box. Results from the association condition in Experiment 3 further reinforce the notion that apes engage in object individuation. In fact, subjects' failure in this condition may have occurred precisely because they assumed that the piece left in the center of the platform during the baiting process and the one they found right before their choice was the same one. In this case, however, object individuation was based on spatio-temporal features, not on object features. Mendes, Rakoczy, and Call (unpublished data) have recently found that apes can use both spatio-temporal and feature object information in a search paradigm previously used with human infants and rhesus macaques (Santos et al. 2002; van de Walle et al. 2000).

Third, the current results suggested an improvement in inferential ability as a function of age. Compared to older subjects, subjects below 8 years of age rarely scored above 75% correct. Such change with age was not detected in the association or the memory conditions. The former is probably due to a floor effect (since subjects did not master the task), the latter probably due to a ceiling effect because even our younger subjects were capable of solving the memory task. The age before adolescence may have especial significance for ape cognitive development. Other studies have shown that this age marks a transition in several abilities such as double-checking in gaze following (Bräuer et al. 2005), mirror self-recognition (Povinelli et al. 1993), and second-order classification of objects (Spinozzi 1993). Such a marked change with age suggests that the type of inference investigated in the current study is not equivalent to the inference displayed in object permanence tasks such as invisible displacement, which is mastered much earlier during development. Since Barth and Call (2006) had tested invisible displacements in many of the same subjects included in the current study, it was possible to correlate the performance in both tasks. There was no significant

relation between the performance in the inference condition of Experiment 3 and single ($r = 0.39$, $N = 19$, $p = 0.10$) or double adjacent displacements ($r = -0.14$, $N = 19$, $p = 0.57$). Additional studies will be needed to confirm the positive relationship between inferential ability and age observed in the current study.

Finally, there was no evidence of species differences in inferential ability. This means that this ability can be traced at least as far back to the common ancestor of all extant great apes. Clearly, other non-human species such as marine mammals and one dog have shown evidence of inference by exclusion (e.g., Herman et al. 1984; Kaminski et al. 2004; Schusterman et al. 1993). However, all these studies involved an extensive period of training in which subject learned to associate certain labels with certain objects, and only then exclusion was tested. In contrast, the current study tested subjects without any prior training on stimuli relations. It is still an open question whether those species that can solve exclusion problems after extensive training on stimuli relations will also be able to solve the task used in the current study.

References

Barth J, Call J (2006) Tracking the displacement of objects: a series of tasks with great apes and young children. J Exp Psychol: Anim Behav Proc (in press)

Bräuer J, Call J, Tomasello M (2005) All great ape species follow gaze to distant locations and around barriers. J Comp Psychol 119:145–154

Bräuer J, Kaminski J, Riedel J, Call J, Tomasello (2006) Making inferences about the location of hidden food: social dog–causal ape. J Comp Psychol 120:38–47

Call J (2001) Object permanence in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and children (*Homo sapiens*). J Comp Psychol 115:159–171

Call J (2004) Inferences about the location of food in the Great Apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus*). J Comp Psychol 118:232–241

Collier-Baker E, Davis JM, Suddendorf T (2004) Do dogs (*Canis familiaris*) understand invisible displacement? J Comp Psychol 118:421–433

Collier-Baker E, Davis JM, Nielsen M, Suddendorf T (2006) Do chimpanzees (*Pan troglodytes*) understand single invisible displacement? Anim Cogn 9:55–61

de Blois ST, Novak MA, Bond M (1999) Can memory requirements account for species' differences in invisible displacement tasks? J Exp Psychol: Anim Behav Proc 25:168–176

Doré FY, Dumas C (1987) Psychology of animal cognition: Piagetian studies. Psychol Bull 102:219–233

Doré FY, Fiset S, Goulet S, Dumas MC, Gagnon S (1996) Search behavior in cats and dogs: interspecific differences in working memory and spatial cognition. Anim Learn Behav 24:142–149

Hashiya K, Kojima S (2001) Hearing and auditory–visual intermodal recognition in the chimpanzee. In: Matsuzawa T (ed) Primate origins of human cognition and behavior. Springer, Berlin Heidelberg New York, pp 155–189

Herman LM, Richards DG, Wolz JP (1984) Comprehension of sentences by bottlenosed dolphins. Cognition 16:129–219

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Kaminski J, Call J, Fischer M (2004) Word learning in a domestic dog: evidence for "fast mapping". Science 304:1682–1683

Mendes N, Huber L (2004) Object permanence in common marmosets (*Callithrix jacchus*). J Comp Psychol 118:103–112

Nissen HW, Blum JS, Blum RA (1948) Analysis of matching behavior in chimpanzees. J Comp Physiol Psychol 41:62–74

Pepperberg IM, Funk MS (1990) Object permanence in four species of psittacine birds: an African grey parrot (*Psittacus erithacus*), an Illiger mimi macaw (*Ara maracana*), a parakeet (*Melopsittacus undulatus*), and a cockatiel (*Nymphicus hollandicus*). Anim Learn Behav 18:97–108

Povinelli DJ, Rulf AB, Landau KR, Bierschwale DT (1993) Self-recognition in chimpanzees (*Pan troglodytes*): distribution, ontogeny, and patterns of emergence. J Comp Psychol 107:347–372

Premack D (1995) Cause/induced motion: intention/spontaneous motion. In: Changeux JP, Chavaillon J (eds) Origins of the human brain. Oxford University Press, Oxford, pp 286–308

Premack D, Premack AJ (1994) Levels of causal understanding in chimpanzees and children. Cognition 50:347–362

Santos LR, Sulkowski GM, Spaepen GM, Hauser MD (2002) Object individuation using property/kind information in rhesus macaques (*Macaca mulatta*). Cognition 83:241–264

Schusterman RJ, Krieger K (1984) California sea lions are capable of semantic comprehension. Psychol Rec 34:3–23

Schusterman RJ, Gisiner R, Grimm BK, Hanggi EB (1993) Behavior control by exclusion and attempts at establishing semanticity in marine mammals using match-to-sample paradigms. In: Roitblat HL, Herman LM, Nachtigall PE (eds) Language and communication: comparative perspectives. Lawrence Earlbaum Associates, Hillsdale, NJ, pp 249–274

Spinozzi G (1993) Development of spontaneous classificatory behavior in chimpanzees (Pan troglodytes). J Comp Psychol 107:193–200

Tomasello M, Call J (1997) Primate cognition. Oxford University Press, New York

Tomonaga M (1993) Tests for control by exclusion and negative stimulus relations of arbitrary matching to sample in a "symmetry-emergent" chimpanzee. J Exp Anal Behav 59:215–229

Van Der Walle GA, Carey S, Prevor M (2000) Bases for object individuation in infancy: evidence from manual search. J Cogn Dev 1:249–280

Watanabe S, Huber L (2006) Animal logics: decisions in the absence of human language. Anim Cogn 9. DOI 10.1007/s10071-006-0043-6

Xu F, Carey S (1996) Infants' metaphysics: the case of numerical identity. Cogn Psychol 30:111–153