

Task Constraints Mask Great Apes' Ability to Solve the Trap-Table Task

Antje Girndt, T. Meier, and J. Call
Max Planck Institute for Evolutionary Anthropology

Researchers have investigated animals' causal knowledge with a task requiring subjects to use a tool to bring a reward within reach whilst avoiding a trap. Previous studies have suggested limitations in the ability of several species to avoid traps in tubes or tables. However, certain features may have inflated task difficulty. We tested 20 chimpanzees (*Pan troglodytes*), 7 orangutans (*Pongo pygmaeus*), 5 bonobos (*Pan paniscus*), and 5 gorillas (*Gorilla gorilla*) in the trap-table—a task in which subjects have to pull one of two rakes prepositioned behind two rewards on a flat surface. One of the rewards is in front of a trap into which it will fall. We investigated the effect of trap type, tool type, the number of available tools, and reinforcement regime on performance. We replicated previous findings showing that apes failed to choose the correct rake above chance. However, when they could instead choose where to insert a single tool, around 80% of the apes solved the trap-table task in the first trial, revealing an important effect of task constraints on their performance.

Keywords: tool-use, problem solving, inhibition, primates, causal knowledge

Research on problem solving involving tool-use has historically contributed (and continues to contribute) in significant ways to the development of the area of causal knowledge in primates (e.g., Köhler, 1925; Natale, 1989; Povinelli, 2000; Visalberghi & Limongelli, 1994). Recently, this interest has extended to certain avian species such as parrots (Pepperberg, 2004), woodpecker finches (Tebbich & Bshary, 2004), and corvids (Chappell & Kacelnik, 2002; Seed, Tebbich, Emery, & Clayton, 2006). Typically, the tasks used to probe causal knowledge in these species involve presenting a reward that cannot be directly accessed due to the presence of a spatial or physical barrier. Simpler versions of these tasks entail displacing the reward with a tool until it is within reach (e.g., Natale, 1989) whereas harder versions also require negotiating obstacles (Köhler, 1925) or avoiding traps during the reward displacement (Visalberghi & Limongelli, 1994).

One task that has been successfully used with a variety of species is the trap-tube task (Visalberghi & Limongelli, 1994), in which individuals have to extract a reward from inside a Plexiglas tube without dropping the reward inside a trap situated in one of the sides of the tube. Capuchin monkeys, chimpanzees, and woodpecker finches find this task quite difficult (Limongelli, Boysen, & Visalberghi, 1995; Povinelli, 2000; Tebbich & Bshary, 2004; Visalberghi & Limongelli, 1994). Only one capuchin (out of four), three chimpanzees (out of 11) and one woodpecker finch (out of six) have consistently solved this task. However, a consistent successful performance only appeared after dozens of trials. Moreover, only the woodpecker finch provided evidence that it was not solely using the position of the trap (regardless of its functionality) to solve the problem. It ignored the trap after the tube was rotated 180 degrees so that the trap was on top and the reward could not fall inside it.

It is conceivable that part of the difficulty of the trap-tube may not be related to the subjects' lack of understanding of its critical features, but to the intervention of other factors that may mask their knowledge. Following this reasoning, there have been several attempts at simplifying the task while keeping it functionally equivalent to the original trap-tube task. Povinelli (2000) reasoned that the central position of the trap (and the consequent off-center position of the reward) used in the original trap-tube study may have biased subjects toward using a rule based on the distance of the reward in relation to the tube openings without reference to the presence of the trap. In other words, subjects may have learned that inserting the tool on the side that was farthest away from the reward invariably produced the reward. Shifting the location of the reward to the center was also adopted by Tebbich and Bshary (2004) in their study with woodpecker finches. The results, however, were not that different from those obtained with the original task, as the majority of subjects still failed that task.

Mulcahy and Call (2006) argued that part of the difficulty of this task resided in the motor response required to get the reward out of the tube. In particular, the diameter of the tube and tool prevented subjects from raking the reward out, and instead forced them to push the reward out. Pushing the reward out may be particularly hard for primates if its initial effect is to increase the distance between the reward and the subject (Guillaume & Meyerson, 1930; Köhler, 1925). Note that this criticism does not apply to the woodpecker finch study because these birds could rake the reward out, which is the response that they spontaneously use to get prey from crevices (Tebbich & Bshary, 2004). Thus, Mulcahy and Call (2006) modified the original task so that subjects could choose between raking or pushing the reward out. The results indicated that subjects preferred to rake the reward in, therefore potentially making the problem simpler. Yet, only 3 out of 10 apes (two orangutans and one chimpanzee) solved the problem above chance levels although they did so faster than any other animals tested until then. Moreover, like the successful woodpecker finch, they also disregarded the trap when the tube was inverted but continued

Correspondence concerning this article should be addressed to Antje Girndt, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103, Leipzig, Germany. E-mail: girndt@eva.mpg.de

to avoid it when it was functional again. However, unlike the woodpecker finch, apes rarely inserted the tube on opposite sides multiple times. The authors concluded that these results made it unlikely (at least for these three subjects) that subjects had used a procedural rule to solve the problem instead of considering the effect of the trap - an explanation that might have been sufficient in earlier studies.

Povinelli (2000) attempted to overcome some of the difficulties inherent to the trap-tube by designing another task that was functionally equivalent to the trap-tube but that used a completely different setup. Chimpanzees faced a table divided into two runs. One run had a hole cut in it (true trap) so that any object displaced over it would fall inside it and be lost while the other run had a fake trap that consisted of a painted area identical in dimensions and position to the true trap. Chimpanzees could see a reward placed behind each of the traps and they could use a rake placed behind the rewards to bring them within reach. Obviously, only the reward behind the fake trap could be obtained and the question was whether chimpanzees would prefer to pull the rake on that side. Chimpanzees were allowed to pull from either of the two rakes until they got the reward that was available. Thus, this task attempted to simplify the problem by presenting the two options readily available so subjects merely had to pull from one of the alternatives. Results indicated that chimpanzees showed no preference for pulling from the rake behind the fake trap. Povinelli (2000) concluded that chimpanzees had little understanding of the trap or the effect it had on the reward. Only one of the subjects managed to select the correct rake as her first choice but she did that during her first 10 trials. Povinelli (2000) suggested that her successful performance could also be explained by an initial preference for the blue surface. After training all subjects to successfully solve the trap table task, he conducted a study to control for this explanation. As a result, none of the tested chimpanzees exhibited a preference for the painted surface. The author concluded that the chimpanzees had, after their training, learned to use some of the relations that the task embodied.

Fujita, Kuroshima, and Asai. (2003) tested four capuchin monkeys on a version of the trap-table in which rakes were replaced by hoes. Capuchin monkeys did not avoid the trap above chance levels. Fujita et al. (2003) concluded that capuchins failed to understand the spatial relations between the tool, the reward, and the trap. Santos, Pearson, Spaepen, Tsao, and Hunter (2006) tested tamarins and vervet monkeys on a modified trap table task. When the trap was functionally relevant, they found that the vervet monkeys chose the tool on the no-trap surface over the tool on the trap surface whereas the tamarins performed at a chance level. The authors inferred that vervet monkeys did take into account the trap's impact on the food. Recently, Cunningham, Anderson, and Mootnick (2006) tested four hoolock gibbons (*Bunopithecus hoolock*) in this task. Two of the four subjects performed above chance, and their performance was high from the beginning of testing. In fact, one of the subjects made no mistakes in 50 trials. However, the authors concluded that rather than understanding the critical properties of the problem, subjects may have either learned to associate the continuous surface of the table with reinforcement or avoided the tool in front of the trap because the trap was perceived as an obstruction.

Although the trap-table task ruled out some of the problems that the original trap-tube may have had, it also raised new ones. The

tools used in this task, a pair of rakes (or hoes), were more complex than those used in the trap-tube task, a single straight stick. Additionally, presenting prepositioned alternatives may have encouraged subjects to grab the first thing that they saw without paying attention to the trap position. This may be especially true given that subjects came into the room for a single trial and left upon completing it. Strong predispositions paired with certain procedures may account for some of the failures of previous tasks. Moreover, because subjects were not differentially reinforced for selecting the correct alternative and the fact that pulling the first rake that came into sight worked by chance in half of the trials, subjects may have had little incentive to learn to select the correct alternative. Finally, the small sample size ($n = 7$) may have prevented the authors from finding individuals that could solve the task as has been seen in other studies.

The goal of this study was to investigate whether subjects would be able to solve the trap-table task and, if so, to investigate what were the task features controlling their responses. To that end, we contrasted the setup originally used by Povinelli (2000) with one in which we varied the form of tool presentation, the trap type, the tool type, and the reinforcement regime. In particular, we presented a single tool that subjects had to direct toward one of the alternatives, thus removing the potential problem of their knowledge being masked by interfering responses or their inability to inhibit a powerful response like grabbing one of the tools. We simplified the setup by making the trap more obvious and using a straight stick (not a rake) as a tool. Finally, we used differential reinforcement by letting each subject select only one alternative so that they maximized the opportunity to focus on the task. Besides the procedural modifications mentioned before, we tested a larger sample of chimpanzees ($n = 20$) and five representatives of each of the other three great ape species to explore whether there were significant differences between great ape species.

Experiment 1: One Table and Two Rakes Versus Two Tables and One Stick

The goal of this experiment was twofold. First, we attempted to replicate the results obtained by Povinelli (2000) with his original setup. Second, we tested subjects on a modified version of the original task in which we modified all four factors that we thought may have contributed to the chimpanzees' failure in the original study. Thus, subjects received a single tool that they had to use to rake a reward from one of two tables. One table was flush against the mesh while the other was placed 15 cm away from the mesh so that it created a gap. Subjects could make only one choice, because after subjects directed the tool toward one of the rewards the experimenter removed the other table's reward. We contrasted this modified version to the original task presented by Povinelli (2000), in which subjects continued to pull at the rakes until they pulled the one that did not have a trap in front of it and secured the reward.

Method

Subjects. Thirteen chimpanzees (*Pan troglodytes*), five bonobos (*Pan paniscus*), four gorillas (*Gorilla gorilla*), and five orangutans (*Pongo pygmaeus*) participated in this experiment. We tested 10 males and 17 females with an age range of 4 years to 32

years (see Table 1 for further information). Subjects were housed in social groups at the Wolfgang Köhler Primate Research Center in the Leipzig Zoo where they can access indoor and outdoor enclosures. All subjects are used to participate in various kinds of experiments as a part of their daily routine. Subjects were never deprived of food or water and received their regular food, which consisted mainly of vegetables, fruits, and monkey chow.

Materials. In the One-table condition, we used one table made of PVC material (43.5 cm × 64 cm), bordered on three of its sides by a small wall (0.5 cm × 5 cm). The table was divided into two identical parts by a central wall (1 cm × 5 cm). Each of the two parts had a hole cut in its center (16 cm × 30 cm). Two different panels could be inserted into the two table partitions and exchanged after each trial according to the test plan. One panel had a hole (23.5 cm × 15 cm) cut at a distance of 13.5 cm and 4 cm

from the end and sides of the table, respectively. The other panel had a blue rectangle identical in size and position to the hole of the other panel. The bottom part of the table was mounted on rails so that it could be pushed forward and backward from the subject. Two identical rakes positioned on each of the table partitions served as tools. The head of the rake was a piece of plywood (30 × 12 × 1 cm) that was attached to a wooden rod (\varnothing 2 cm, length 43.5 cm) that acted as handle.

In the Two-table condition, we used two identical small tables made of PVC material (32.5 cm × 20 cm), surrounded on three of its sides by a small wall (0.5 cm × 5 cm). A wooden dowel (40 cm length, 0.6 cm \varnothing) served as tool. We used grapes, monkey biscuits, or banana slices as rewards.

Procedure. The experimenter (E) placed the apparatus next to the mesh and sat down on a stool behind the apparatus. The ape was located on the other side of the mesh facing the apparatus and could observe the experimenter setting-up. There were two conditions:

One-table. The table was placed with its open side flush against the mesh (Figure 1a). The E deposited a reward in the center of each of panel at 37 cm from the mesh and placed a rake on each panel with their handles directed toward the subject. The rewards were placed simultaneously on both sides. The head of each rake was perpendicular to the panel and situated right behind the reward (from the subject's perspective). Then E slid the table toward the subject, so that the handles went through the mesh, thus allowing the subject access to them. We scored the first rake touched by the subject as her choice but subjects were allowed to manipulate both tools if they chose to do so. A trial ended after the subject either retrieved or lost the reward (by dropping it through the hole in one of the panels) or after one minute had elapsed without the subject manipulating the tools. After the subject had not manipulated the tools for one minute three times in a row, the session was terminated.

Two-table. The tables were placed side by side with their open sides facing the mesh (Figure 1b). One table was flush against the mesh while the other was positioned 15 cm back from the mesh. The experimenter deposited a reward in the center of each table, 15 cm from the mesh. For the table that was flush against the fence this meant that the reward was located near the back wall of the table while for the other table the reward was located at the front of the table near the edge. Then the experimenter introduced the stick through the mesh between the two tables and held it there until the subject picked it up. We scored the first grape to which the subject directed the stick as her choice and immediately removed the grape not targeted by the subject. A trial ended after the subject either retrieved or lost the reward (by dropping it into the gap between the recessed table and the mesh) or after one minute had elapsed without the subject manipulating the tool.

Each subject received two 10-trial sessions in each condition. The two sessions of one condition were conducted consecutively. We counterbalanced the order of the conditions across subjects and the position of the correct alternative across trials within a session so that it appeared the same number of times on the right and on the left side. Subjects received only one session per day.

Data scoring and analyses. All trials were videotaped. We scored the first rake touched by the subject (One-table condition) and the first grape targeted by the subject (Two-table condition) as her choice. JC scored 20% of a randomly determined choice of

Table 1
Subjects That Participated in the Study

Name	Sex	Age in years	Study participation	Previous experience on tool-use tasks
<i>Orangutans</i>				
Bimbo	male	24	1,2,4,5	a, b
Pini	female	17	1,2,4,5	a, b
Padana	female	7	1,2,4,5	—
Dokana	female	15	1,2,4,5	a, b
Dunja	female	32	1,2,5	a
Walter	male	16	4	a, b
Toba	female	11	4	a, b
<i>Chimpanzees</i>				
Unyoro	male	7	1,2,4	—
Jahaga	female	12	1,2,4,5	—
Trudi	female	12	1,2,4,5	—
Fifi	female	12	1,2,3,4,5	b
Alexandra	female	5	1,2,4,5	—
Alex	male	4	1,2,3,4,5	—
Annett	female	5	1,2,3,4,5	—
Robert	male	29	1,2	—
Frodo	male	11	1,2,3,4,5	—
Patrick	male	8	3,4,5	—
Natascha	female	25	2,4,5	—
Ulla	female	28	1,2,3,4	—
Pia	female	5	1,2,4,5	—
Swela	female	9	1,2,5	—
Corrie	female	28	2	—
Sandra	female	12	3,4,5	—
Riet	female	27	3,5	—
Brent	male	5	1,2,3	b
Dorien	female	24	2,4,5	—
Fraukje	female	29	3,4,5	—
<i>Gorillas</i>				
Gorgo	male	24	1,2,4	a
N'diki	female	27	1,2,4	a
Viringika	female	10	1,2,4,5	a, b
Bebe	female	~26	1,2,4	a
Ruby	female	7	2	—
<i>Bonobos</i>				
Ulindi	female	11	1,2,4,5	b
Yasa	female	7	1,2,4,5	—
Limbuko	male	9	1,2,4,5	—
Joey	male	22	1,2,4,5	b
Kuno	male	8	1,2,4,5	—

a, Mulcahy et al., 2005; b, Mulcahy & Call, 2006.

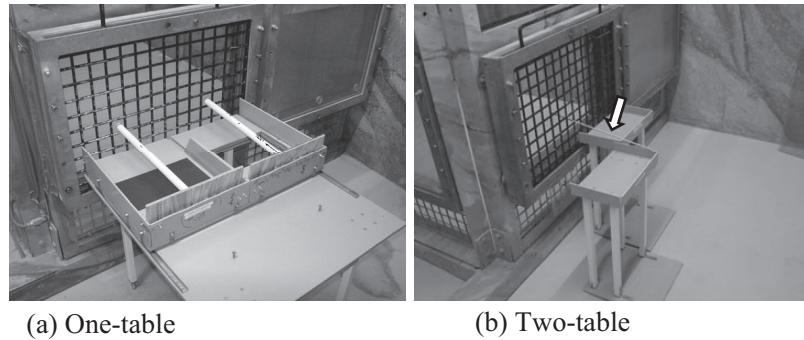


Figure 1. Experimental Setup for exp. 1 and exp. 2, conditions One-table (a) and Two-table (b). For the One-Table condition two prepositioned rakes served as tools (see Figure 5), whereas in Condition Two-Table the experimenter handed a wooden dowel of 40 cm length to the ape through the center of the mesh. The white arrow in Figure 1b indicates the position of the stick.

trials from videotapes to assess interobserver reliability. Interobserver reliability was excellent (Cohen's $\kappa = 0.84$). We analyzed the percentage of correct responses as a function of condition and session with a mixed ANOVA. Moreover, we tested whether subjects performed above chance levels using a one-sample t test (with 50% as the chance expected value). Finally, we used the binomial test to assess the subjects' performance on the first trial. All statistical tests were two-tailed.

Results

Figure 2 presents the mean percent correct trials as a function of condition and session. A $2 \times 2 \times 4$ ANOVA with condition and session as within-subject factors and species as between subject factor revealed that subjects performed significantly better in the Two-tables condition compared to the One-table condition, $F_{1,22} = 57.34$, $p < .001$, $\eta^2 = 0.72$. Additionally, subjects' performance improved across sessions, $F_{1,22} = 6.61$, $p = .017$, $\eta^2 = 0.23$. No other factors or interactions showed a significant effect. Subjects were above chance in both sessions of the Two-tables condition (1st session: $t_{26} = 8.17$, $p < .001$; 2nd session: $t_{26} = 8.13$, $p < .001$). In contrast, subjects did not perform above chance in either of the two sessions of the One-table condition (1st session: $t_{26} = 0.87$, $p = .39$; 2nd session: $t_{26} = 1.54$, $p = .14$). Overall, 74% and 7% of the subjects scored 75% correct or higher

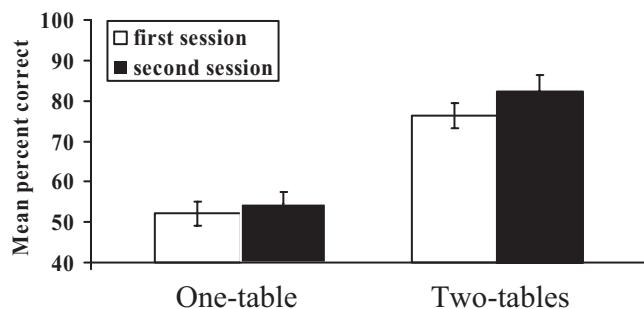


Figure 2. Experiment 1. Mean percent of trials (\pm SEM) in which subjects chose the correct alternative in Condition One-table and Condition Two-tables. The data are split in first and second half of the 20 trials.

in the Two-tables and One-table conditions, respectively. An analysis of the performance in the first trial corroborated these results. Twenty-one out of 26 subjects (81%) responded correctly in the first trial in the Two-tables condition (Binomial test: $p = .002$), but only 17 out of 26 subjects (65%) did so in the One-table condition (Binomial test: $p = .17$).

Discussion

Subjects' performance dramatically differed between conditions. In the One-table condition, apes failed to select the rake on the side without the trap, thus replicating (with a larger sample size) the results obtained by Povinelli (2000). In contrast, apes targeted the reward located on the table without a trap in the Two-table condition. This result was evident in the first trial. The contrast between the One- and Two-table conditions is even starker when one considers that many subjects that had succeeded in the Two-table condition failed in the One-table condition. The reverse never occurred.

Thus, this experiment established that subjects were able to spontaneously avoid a trap table. However, the factors responsible for this improved performance compared to the original task are unclear. The form of tool presentation, the position of the trap (forward in the Two-table and back in the One-table condition), the number of tables, the type of tool, the differential reinforcement, or even the apparent distance of the grape to the fence may have been responsible for the observed differences. In the next four experiments, we systematically varied some of these factors and controlled others to identify the factors that made the original task so difficult.

Experiment 2: One Table: Two Rakes Versus One Stick

In this and subsequent experiments, we used the table from the One-Table condition (Figure 1a) and always administered two conditions. In the current experiment, one condition presented two prepositioned rakes, while the other condition presented a single stick in the center.

Method

Subjects. Thirty-one apes were tested in this experiment. There were 10 males and 21 females with an age range of 4 years

to 32 years. Twenty-seven subjects had participated in the previous experiment and three naïve chimpanzees and one gorilla were new to this experiment (see Table 1). Five chimpanzees (Brent, Corrie, Dorien, Robert, and Natascha) and two gorillas (Gorgo, Ruby) failed to complete at least one of the sessions and were subsequently dropped from the analyses.

Apparatus. We used the same single table as in the one-table condition of Experiment 1, except that we widened the hole on its lateral sides to prevent subjects from moving the reward around the hole and against the lateral walls of the table—something that one subject managed to do on one occasion in a pretest of the experiment. To match the new size of the hole, we also changed the size of the painted rectangle on the other panel accordingly. We used the same tools as in Experiment 1 except that we increased the length of the stick to 50 cm to accommodate for the larger size of the table.

Procedure. The experimenter placed the table flush against the mesh, the ape sat on the other side of the mesh facing the table and the experimenter baited the table as in Experiment 1. There were two conditions: In the Two-rakes condition the subjects received two prepositioned rakes, whereas in the One-stick condition subjects were given a stick through the center of the mesh. Each subject received two 10-trial sessions in each condition. We counterbalanced the order of the conditions across subjects and the position of the correct alternative across trials within a session so that it appeared the same number of times to the right and to the left. Subjects received only one session per day. All trials were videotaped. We used the same scoring, reliability assessment, and analyses as in Experiment 1. Interobserver reliability was excellent (Cohen's $\kappa = 0.96$).

Results

Figure 3 presents the mean percent of correct trials as a function of condition and session. A $2 \times 2 \times 4$ ANOVA with condition and session as within-subject factors and species as between subject factor revealed that subjects performed significantly better in the One-stick condition compared to the Two-rakes condition, $F_{1,20} = 7.08$, $p < .015$, $\eta^2 = 0.26$. Additionally, subjects' performance improved across sessions, $F_{1,20} = 5.63$, $p = .028$, $\eta^2 = 0.22$. No other factors or interactions showed a significant effect. Subjects were above chance in both sessions of both conditions, $t_{23} = 3.77$, $p < .002$, in all cases. Overall, 73% and 23% of the subjects scored

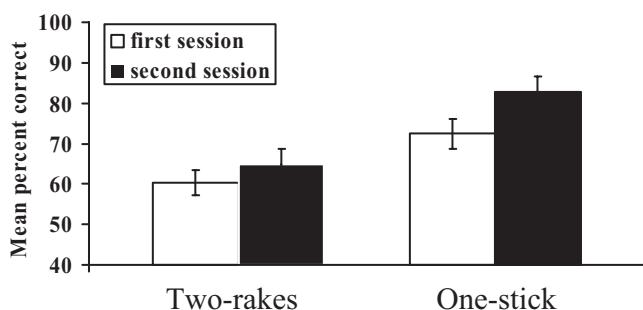


Figure 3. Experiment 2. Mean percent of trials (\pm SEM) in which subjects chose the correct alternative in Condition Two-rakes and Condition One-stick. From this experiment on, we only used one table instead of two tables. The data are split in first and second half of the 20 trials.

75% correct or higher in the One-stick and Two-rakes conditions, respectively. Twenty out of 24 subjects (83%) responded correctly on the first trial in the One-stick condition, Binomial test: $p = .002$, but only 13 out of 24 subjects (54%) did so in the Two-rakes condition, Binomial test: $p = .84$.

Discussion

We obtained very similar results to those in Experiment 1. Subjects performed significantly better when they could insert the stick rather than choose between two prepositioned rakes. Again, subjects performed above chance in the first trial of the One-stick condition. The use of the same table in both conditions allowed us to rule out the possibility that the observed differences between conditions in Experiment 1 were due to features inherent to the table such as the type and position of the gap, or the possibility that subjects perceived the reward closer when it was right behind a wall. More importantly, the results showed that subjects can solve the trap table task on the first trial when what they have to do is to insert a tool and rake in one of two alternatives.

Interestingly, subjects also performed above chance in the Two-rake condition, which represents an improvement on the results of Experiment 1, even though this improvement was not apparent on the first trial. Numerically, however, the apes' overall performance (65.1%) was not so different from the values observed in Experiment 1 and those reported by Povinelli (2000). Three quarters of the subjects scored 75% correct or higher in the One-stick condition compared to only one quarter in the Two-stick condition. One possible explanation for the continued differences observed between the stick and rakes conditions is that they are due to the differential reinforcement that subjects received in the former but not the latter. In the next experiment, we contrasted the performance of two groups of chimpanzees in the Two-rake conditions under different reinforcement regimes and with a sample size more comparable to that of the original Povinelli (2000) study.

Experiment 3: The Effect of Reinforcement

In this experiment we compared the performance of two groups of chimpanzees under different reinforcement regimes. One group was nondifferentially reinforced so that they were allowed to use any rake they wanted until they obtained the reward. This meant that subjects were rewarded regardless of which rake they chose first. The other group was differentially reinforced by allowing subjects to only pull the first rake that they touched, which meant that subjects only got the reward if they pulled first from the rake placed on the panel without a trap.

Method

Subjects. We tested four totally naïve chimpanzees (see Table 1) and compared their performance to the performance of six chimpanzees in the One-table condition of Experiment 1. As those chimpanzees received Condition One-table first, they were naïve to the setup, too. We did not retest those chimpanzees in the current experiment but used their data from Experiment 1. There were 4 males and 6 females with an age range of 4 years to 29 years (see Table 1). All chimpanzees belonged to the social groups

housed at the Wolfgang Köhler Primate Research Center in the Leipzig Zoo.

Apparatus. We Used the Same Single Table and Two Prepositioned Rakes of Experiment 1.

Procedure

The basic procedure was identical to the One-table condition of Experiment 1, except that the four naïve subjects were differentially reinforced for their choices as we had done in the Two-table condition of Experiment 1. Subjects were only allowed to make one choice by removing the reward that they had not targeted in their first attempt. The six experienced subjects followed the exact nondifferential reinforcement procedure as in the One-table condition of Experiment 1. Each subject received two 10-trial sessions. We counterbalanced the position of the correct alternative across trials within a session so that it appeared the same number of times on the right and on the left. Subjects received only one session per day. All trials were videotaped and we used the same scoring, reliability assessment, and analyses as in Experiment 1. Interobserver reliability was excellent (Cohen's $\kappa = 1.0$).

Results

A 2×2 ANOVA with session as within-subject factor and reinforcement regime as between subject factor on the percent of correct trials revealed no significant effects for reinforcement regime, $F_{1,8} = 0.16, p = .72$, eta² = 0.02, session, $F_{1,8} = 1.03, p = .34$, eta² = 0.11, or reinforcement regime x session, $F_{1,8} = 3.60, p = .095$, eta² = 0.31. Subjects were not above chance in the first session, $t_9 = 0.43, p = .68$ (mean percent correct = 48.0, sem = 4.7) or second session, $t_9 = 0.65, p = .53$ (mean percent correct = 53.4, sem = 5.3). Overall, none of the subjects reached the 75% correct score.

Discussion

We found no evidence that the reinforcement regime had an effect on performance after 20 trials. Subjects did not perform above chance levels in either the first or the second session. Thus, differential reinforcement did not improve performance, which failed to reach above chance levels, in the two rake condition after 20 trials. One outstanding question is whether the type of tool, independent of other factors, affected the performance of subjects. Note that subjects were invariably more successful with the stick than the rakes. In the next experiment, we compared performance with one and two tools, in Experiment 4a with one or two sticks, and in Experiment 4b with one or two rakes. We also used differential reinforcement in all conditions to see whether subjects would improve their performance in those conditions with two tools present.

Experiment 4a: One Stick Versus Two Sticks

This experiment contrasted the subjects' performance on two prepositioned sticks with a single stick offered between the two rewards under a differential reinforcement regime.

Method

Subjects. Thirty apes were tested in this experiment. All had participated in the previous experiments (see Table 1). There were 10 males and 20 females, with an age range of 4 years to 32 years.

Apparatus. We Used the Same Single Table and Sticks as in Experiment 2.

Procedure. The procedure was identical to that of Experiment 2 except that we replaced the rakes for sticks. Consequently, we administered two conditions: One-stick and Two-sticks. Each subject received one 10-trial session per condition. We counterbalanced the order of the conditions and the position of the correct alternative as in Experiment 1. Subjects received only one session per day. We videotaped all trials and used the same scoring, reliability assessment, and analyses as in Experiment 1. Interobserver reliability was excellent (Cohen's $\kappa = 0.94$).

Results

Figure 4 presents the mean percent of correct trials as a function of condition. A 2×4 ANOVA with condition as within-subject factors and species as between subject factor revealed that subjects performed significantly better with one stick than with two sticks, $F_{1,27} = 16.18, p < .001$, eta² = 0.38. No other factors or interactions showed a significant effect. Subjects were above chance both in the One-stick, $t_{30} = 9.47, p < .001$, and Two-stick conditions, $t_{30} = 3.91, p < .001$. Overall, 74% and 35% of the subjects scored 75% correct or higher in the One-stick and Two-sticks conditions, respectively. Twenty-nine out of 31 subjects (94%) responded correctly in the first trial in the One-stick condition, Binomial test: $p < .001$, but only 12 out of 31 subjects (39%) did so in the Two-stick condition, Binomial test: $p = .28$.

There was no improvement across trials in the One-stick (Cochran Q = 4.98, df = 9, $p = .84$) or Two stick conditions (Cochran Q = 6.11, df = 9, $p = .73$). Comparing the 1st and the 10th trial produced analogous results (Sign test: One stick: $p = .125$; Two stick: $p = 1.0$).

Discussion

We obtained the same results as in Experiments 1 and 2. Namely, subjects performed significantly better when they were offered a single tool between the two alternatives than when they found two prepositioned tools, one in each alternative. This means

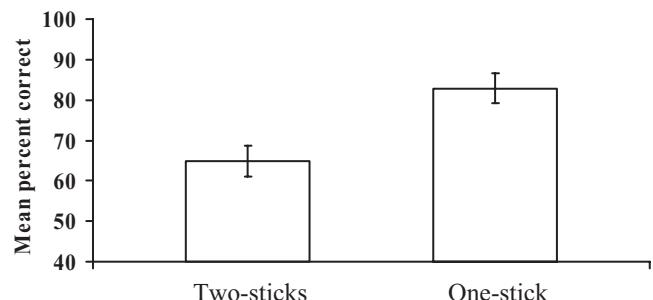


Figure 4. Experiment 4a. Mean percent of trials (\pm SEM) in which subjects chose the correct alternative in Condition Two-sticks and Condition One-stick. The subjects received 10 trials per condition.

that their low performance in the two-rake condition is not solely attributable to the type of tool, as we also observed the same effect with sticks. In the next experiment, we tested this hypothesis further by using the same conditions as in the current experiment except that we replaced the sticks for rakes. We predicted that subjects would perform better in the one rake compared to the two rake condition.

Experiment 4b: One Rake Versus Two Rakes

Method

Subjects. Twenty-six apes were tested in this experiment. All had participated in the previous experiments (see Table 1). There were seven males and 19 females, with an age range of 4 years to 32 years.

Apparatus. We Used the Same Single Table and Rakes as in Experiment 2.

Procedure. The procedure was identical to that of Experiment 4a except that we replaced the sticks with rakes (see Figure 5). Each subject received one 10-trial session per condition. We counterbalanced the order of the conditions and the position of the correct alternative as in Experiment 1. We videotaped all trials and used the same scoring, reliability assessment, and analyses as in Experiment 1. Interobserver reliability was excellent (Cohen's $\kappa = 0.98$).

Results

Figure 6 presents the mean percent of correct trials as a function of condition. A 2×4 ANOVA with condition as within-subject factor and species as between subject factor revealed that subjects tended to perform better with one rake than with two, $F_{1,23} = 3.52$, $p = .074$, $\eta^2 = 0.13$. No other factors or interactions showed a significant effect. Removing species from the model confirmed the effect of condition on performance, $F_{1,26} = 5.20$, $p = .031$, $\eta^2 = 0.17$. Subjects were above chance both in the One-rake, $t_{26} = 6.62$, $p < .001$, and Two-rakes conditions, $t_{26} = 3.51$, $p = .002$. Overall, 59% and 30% of the subjects scored 75% correct or higher in the One-rake and Two-rakes conditions, respectively. Nineteen out of 27 subjects (70%) responded correctly in the first trial in the One-rake condition, Binomial test: $p = .034$, but only 14 out of 27 subjects (52%) did so in the Two-rakes condition, Binomial test: $p = .85$.

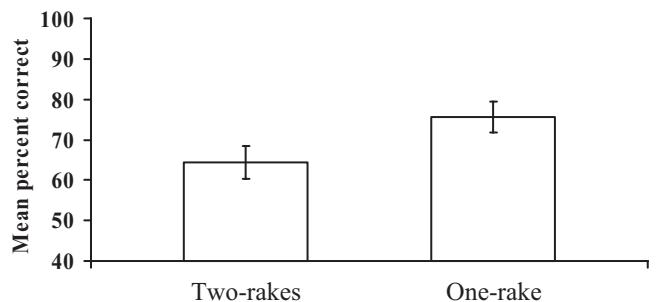


Figure 6. Experiment 4b. Mean percent of trials (\pm SEM) in which subjects chose the correct alternative in Condition Two-rakes and Condition One-rake. The subjects received 10 trials in each of the conditions.

There was no significant improvement across trials in the One-rake ($Cochran Q = 13.80$, $df = 9$, $p = .13$) or Two-rakes conditions ($Cochran Q = 13.03$, $df = 9$, $p = .16$). However, comparing the first and the 10th trial revealed that subjects improved their performance in the One-rake condition (Sign test: $p = .016$) but not in the Two-rake condition (Sign test: $p = 1.0$).

We also compared the results of the last two experiments. A 2×2 ANOVA with tool type (stick vs. rake) and tool quantity (one vs. two) as within-subject factors revealed that subjects performed significantly better with one tool compared to two tools, $F_{1,23} = 16.97$, $p < .001$, and with the stick compared to rake, $F_{1,23} = 9.52$, $p = .005$. However, there was also a significant tool type \times tool quantity effect, $F_{1,23} = 4.92$, $p = .037$. Post hoc tests indicated that subjects performed better in the stick than the rake in the one tool condition, $t_{23} = 3.77$, $p = .001$, but not in the two tools condition, $t_{23} = 0.23$, $p = .82$.

Discussion

Once again subjects performed significantly better when they were offered a tool in the center location than when they found two prepositioned tools, thus confirming that this effect was independent of the type of tool. However, the joint analysis of Experiments 4 and 5 indicated that subjects performed better with the stick than the rake in the one tool condition, which suggests some contribution of the type of tool as well. This effect, however, was not large enough to prevent subjects from being above chance in the first

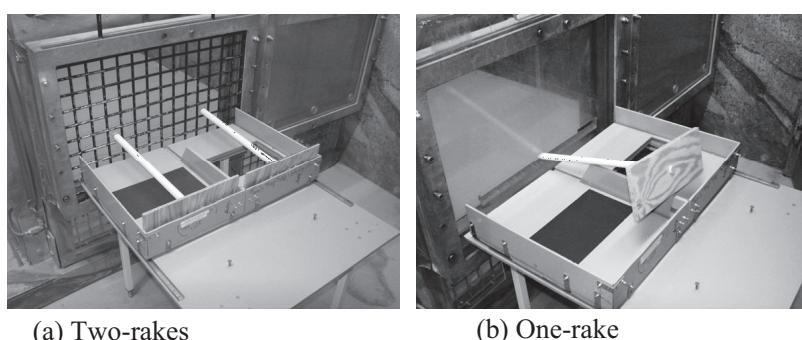


Figure 5. Experimental setup for exp. 4b, Conditions Two-rakes (a) and One-rake (b). The cut-out in the Plexiglas could be used to maneuver the rake to the sides. In each condition subjects received 10 trials

trial of the one rake condition. It is very likely that the difference between the stick and the rake arose from the more complex actions needed to manipulate the rake in the One-tool condition (there were no differences between the rake and the stick in the Two-tool condition), which compared to the stick was heavier and composed of two pieces. Indeed, some of the observed mistakes may have been attributable to a lack of motor control rather than to selecting the grape on the trap-less panel.

General Discussion

These experiments demonstrated that apes were sensitive to the position of a trap on a platform when attempting to drag a reward toward them. Apes' high performance in the first trial ruled out the possibility that subjects learned to respond in this way during the test. These positive findings contrast with previous results (Povinelli, 2000), which we also replicated in the current study. It appears that a critical factor in explaining this discrepancy was the position of the tools on the platform. In particular, presenting a pair of tools prepositioned on the platform led subjects to make numerous errors whereas giving a single tool in between the two options resulted in a substantially higher performance. Other factors investigated, including the type of table, the type of trap, or the kind of tool had little effect on performance. Finally, we found no evidence of species differences.

Previous negative results have been interpreted as reflecting a lack of knowledge about the critical features of the task (Fujita et al., 2003; Povinelli, 2000; Visalberghi & Limongelli, 1994). The current results should make us cautious about this conclusion. Individuals may have possessed sufficient knowledge about the nature of the task, but such knowledge may have been masked by their poor attentional or motor control. Thus, subjects may have grabbed the first tool (or inserted the tool in the first opening) that captured their attention regardless of its appropriateness, and subsequently experienced great difficulty releasing that tool regardless of its perceived effectiveness. It is even conceivable that apes did not see grabbing the rake as a choice between the two rewards, but simply as a way of procuring a tool that they were going to use. So what the experimenter saw as a choice between two rewards placed on two panels, for the ape may have been a choice between two identical tools. This is not to say that subjects were unable to learn to choose the correct tool. They did, but their choices may have been based on associating certain perceptual configurations with reinforcement, not necessarily on knowledge about the critical features of the task.

Another explanation for the apes' failure is that recruiting the necessary cognitive resources to inhibit certain motor responses and producing others may have drained the resources needed to judge which of two options was correct. Cognitive psychologists have often observed a marked performance decrease in a cognitively demanding task after subjects are required to simultaneously perform a second simple and often unrelated task. Unlike the previous alternative, this explanation would not focus on inhibition of certain responses but on the management of a common pool of attentional and motor skills. Regardless of which one of these two alternatives is correct, subjects' performance improved dramatically after changing the way the tools were presented to them.

One paradoxical outcome of this research is that providing prepositioned tools in an attempt to simplify the problem, so that

subjects did not have motor control problems because they simply had to pull from one of the alternatives, or that they did not have to decide where to insert the tool (thus reducing the number of steps required to complete the task), actually made the problem harder. One possible explanation for this result lies on the hierarchical structure inherent in problem solving. In effect, problem solving consists of making decisions at certain points regarding things like which tool to use, where to insert it, and what action to perform. Providing subjects with some ready-made decisions may be detrimental to their performance if those decisions do not coincide with the decisions that subjects themselves would have also made, especially if they have difficulty discarding the offered alternatives. For instance, primates tested in the trap task were required to push the reward out of the tube, but Mulcahy and Call (2006) found that most apes when given both options, preferred to rake (not push) the reward out of the tube. Interestingly, three of the apes tested by Mulcahy and Call (2006) solved the trap-tube problem faster than any other nonhuman animal tested before. Moreover, these three animals also ignored the position of the trap when the trap was inverted—again something that no other apes had done before. Similarly, offering apes prepositioned tools, as in the current study, interferes with their ability to select the correct alternative because of their bias to grab the first tool that they see independently of the position of the trap. When such a strong predisposition was eliminated, subjects performed significantly better.

Thus, it is conceivable that apes cope better with more open-ended problems in which they can decide between the different options that they have available. Attempting to simplify the task by removing some of the options may substantially reduce their probability of success. This finding may have important consequences for those studies that assess causal knowledge about tool features by presenting pairs of rewards near two different tools on a platform. For instance, one reward is placed next to a hoe while the other is placed next to a straight stick. Subjects are required to select the tool that can bring the reward closer, in this case the hoe. Invariably, monkeys and apes initially fail to discriminate between the two tools (Fujita et al., 2003; Herrmann et al., submitted; Hauser, Kralik, & Botto-Mahan, 1999; Hauser, Pearson, & Seelig, 2002; Povinelli, 2000, but see Cunningham et al., 2006), but with sufficient training monkeys and apes can learn to choose the correct alternative. Some subjects can even generalize to new problems on the platform (e.g., Fujita et al., 2003). It is unclear, however, what individuals had learned and how well this knowledge would transfer to more open-ended situations. For instance, are those individuals that select the hoe also capable of selecting and using the hoe to capture a string from which a reward is suspended? This is still unknown and future research should establish the relation between forced-choice tasks and open ended problems. The current study shows that failure on the two-choice platform task does not predict failure on the more open-ended task. It would be particularly important to test capuchin monkeys in the more open-ended situation used in the current study that would require them to direct the tool toward one of the two alternatives. This research across tasks is particularly important because generalizations to novel situations that share functional properties is one of the main indicators of flexible cognition.

In conclusion, we found that apes were sensitive to the position of a trap when raking a reward located on a platform. However,

such good performance depended on whether subjects could choose where to insert the tool. Selecting between two prepositioned tools was much harder for them. Their systematic high performance in the first trial is inconsistent with an explanation based on learning procedural rules to solve the problem during the test. Although these data are consistent with the idea that apes possess causal knowledge governing the interaction between the tool, the reward, and the obstacle, more research is needed to confirm this. In particular, future studies should be devoted to rule out certain explanations (e.g., apes have an innate predisposition to avoid holes or prefer blue holes, see Povinelli, 2000) and to explore to what extent apes can transfer their high performance across functionally equivalent tasks. Obviously, keeping in mind that good performance in those transfer tasks may critically depend on the presence (or absence) of certain task features whose investigation was the primary goal of the current study. Although one could argue that if subjects are not able to inhibit or rectify certain responses they do not have a full understanding of the problem—a position championed by Piaget among others—this study highlights how easily good performance may be masked by the constraints of the particular testing situation.

References

- Chappell, J., & Kacelnik, A. (2002). Tool selectivity in a non-primate, the new caledonian crow (*Corvus monedula*). *Animal Cognition*, 5(2), 71–78.
- Cunningham, C. L., Anderson, J. R., & Mootnick, A. R. (2006). Object manipulation to obtain a food reward in hoolock gibbons, *Bunopithecus hoolock*. *Animal Behaviour*, 71, 621–629.
- Fujita, K., Kuroshima, H., & Asai, S. (2003). How do tufted Capuchin monkeys (*Cebus apella*) understand causality involved in tool use? *Journal of Experimental Psychology: Animal Behavior Processes*, 29(3), 233–242.
- Guillaume, P., Meyerson, I. (1930). "Recherches sur l'usage de l'instrument chez les singes. I. *Le probleme du detour.*" *Journal de Psychologie Normale et Pathologique*, 27, 177–236.
- Hauser, M., Kralik, J., & Botto-Mahan, C. (1999). Problem solving and functional design features: Experiments on cotton-top tamarins, *Saguinus oedipus oedipus*. *Animal Behaviour*, 57, 565–582.
- Hauser, M., Pearson, H., Seelig, D. (2002). "Ontogeny of tool use in cottontop tamarins, *Saguinus oedipus*: innate recognition of functionally relevant features". *Animal Behaviour*, 64, 299–311.
- Herrmann, E., Wobber, V. E., Call, J. (submitted). "Apes' understanding of tool functional properties after limited experience". *Journal of Comparative Psychology*.
- Köhler, W. (1925). *The mentality of apes*. London: Routledge and Kegan Paul.
- Limongelli, L., Boysen, S. T., & Visalberghi, E. (1995). Comprehension of cause-effect relations in a tool-using task by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 109(1), 18–26.
- Mulcahy, N., Call, J., Dunbar R. (2005). Gorillas (*Gorilla gorilla*) and orangutans (*Pongo pygmaeus*) encode relevant problem features in a tool-using task. *Journal of Comparative Psychology*, 119(1), 23–32.
- Mulcahy, N. J., Call, J. (2006). How great apes perform on a modified trap-tube task. *Animal Cognition*, 9(3), 193–199.
- Natale, F. (1989). Causality II: The stick problem. In F. Antinucci (Ed.), *Cognitive structure and development in nonhuman primates* (pp. 121–133). Hillsdale, NJ: Erlbaum.
- Pepperberg, I. M. (2004). "Insightful" string-pulling in grey parrots (*Psittacus erithacus*) is affected by vocal competence. *Animal Cognition*, 7(4), 263–266.
- Povinelli, D. (2000). Folk physics for apes. Oxford: University Press
- Santos, L. R., Pearson, H. M., Spaepen, G. M., Tsao, F., Hauser, M. D. (2006). Probing the limits of tool competence: Experiments with two non-tool-using species (*Cercopithecus aethiops* and *Saguinus oedipus*). *Animal Cognition*, 9(2), 94–109.
- Seed, A. M., Tebbich, S., Emery N. J., Clayton, N. S. (2006). Investigating physical cognition in rooks, *Corvus frugilegus*. *Current Biology*, 16, 697–701.
- Tebbich, S., Bshary, R. (2004). Cognitive abilities related to tool use in the woodpecker finch, *Cactospiza pallida*. *Animal Behaviour*, 67, 689–697.
- Visalberghi, E., & Limongelli, L. (1994). Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 108(1), 15–22.

Received June 5, 2007

Revision received October 10, 2007

Accepted October 11, 2007 ■