Journal of Experimental Psychology: Animal Learning and Cognition

The Cognitive Underpinnings of Flexible Tool Use in Great Apes

Christoph J. Völter and Josep Call

Online First Publication, April 7, 2014. http://dx.doi.org/10.1037/xan0000025

Journal of Experimental Psychology: Animal Learning and Cognition

2329-8456/14/$12.00 DOI: 10.1037/xan0000025

The Cognitive Underpinnings of Flexible Tool Use in Great Apes

Christoph J. Völter
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany and University of Bielefeld

Josep Call
University of St Andrews and Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Nonhuman primates perform poorly in trap tasks, a benchmark test of causal knowledge in nonhuman animals. However, recent evidence suggests that when the confound of tool use is avoided, great apes' performance improves dramatically. In the present study, we examined the cognitive underpinnings of tool use that contribute to apes' poor performance in trap tasks. We presented chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), and orangutans (*Pongo abelii*) with different versions of a maze-like multilevel trap task. We manipulated whether the apes had to use their fingers or a stick to negotiate a reward through the maze. Furthermore, we varied whether the apes obtained visual information about the functionality of the traps (i.e., blockage of free passage) or only arbitrary color stimuli indicating the location of the traps. We found that (a) apes in the finger-maze task outperformed apes in the tool-use-maze task (and partially planned their moves multiple steps ahead), and (b) tool-using apes failed to learn to avoid the traps and performed similar to apes that did not obtain functional information about the traps. Follow-up experiments with apes that already learned to avoid the traps showed that tool use or the color cues per se did not pose a problem for experienced apes. These results suggest that simultaneously monitoring 2 spatial relations (the tool-reward and reward-surface relation) might overstrain apes' cognitive system. Thus, trap tasks involving tool use might constitute a dual task loading on the same cognitive resources; a candidate for these shared resources is the attentional system.

Keywords: primate cognition, tool use, dual task, causality, planning

Supplemental materials: http://dx.doi.org/10.1037/xan0000025.supp

The cognitive underpinnings of tool use have been the focus of a long-standing debate in the animal cognition literature (Bird & Emery, 2009; Hansell & Ruxton, 2008; Jalles-Filho, Teixeira Da Cunha, & Salm, 2001; Kacelnik, 2009; Matsuzawa, 2001; McCormack, Hoerl, & Butterfill, 2011; Seed & Byrne, 2010; Shumaker, Walkup, & Beck, 2011). In humans, tool use often involves causal understanding of object–object relations, planning of a sequence of actions toward an overarching goal, as well as sensorimotor abilities allowing for the coordination of perception and action during tool use (Roche, Blumenschine, & Shea, 2009; Seed & Byrne, 2010; Stout & Chaminade, 2007). The question arises to what extent these cognitive abilities are also present in other species, because tool use per se does not necessarily require such cognitive sophistication.

In the last two decades, researchers have employed trap tasks to shed light on causal knowledge and tool use of nonhuman primates (Girndt, Meier, & Call, 2008; Limongelli, Boysen, & Visalberghi, 1995; Martin-Ordas & Call, 2009; Martin-Ordas, Call, & Colmenares, 2008; Mulcahy & Call, 2006; Povinelli, 2000; Seed, Call, Emery, & Clayton, 2009; Visalberghi & Limongelli, 1994). In most of these tasks, subjects needed to extract a reward from an apparatus by means of a stick while avoiding a trap. Typically, subjects could move the reward either over a hole, which would cause the reward to fall into the trap (i.e., loss of reward), or it could be moved away from the trap, which would then enable the subject to extract the reward from the apparatus. In the classic trap-tube task pioneered by Visalberghi and Limongelli (1994), the food reward was located out of reach in a horizontally mounted, narrow Plexiglas tube. Crucially, in the middle of this tube, there was a hole in the bottom with a trap underneath. To access the food reward, the subjects had to insert a stick into the opening of the tube farthest from the reward to push it away from the trap.

Capuchin monkeys (*Cebus apella*) as well as chimpanzees (*Pan troglodytes*) performed poorly in this task: One of four capuchin monkeys (Visalberghi & Limongelli, 1994), two of five chimpanzees (Limongelli et al., 1995), and three of seven chimpanzees

Christoph J. Völter, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, and Department of Animal Behaviour, University of Bielefeld, Bielefeld, Germany; Josep Call, School of Psychology and Neuroscience, University of St Andrews, St Andrews, Fife, United Kingdom, and Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology.

Christoph J. Völter was supported by a scholarship of the German National Academic Foundation. We thank Raik Pieszek for constructing the experimental apparatuses, Roger Mundry for statistical advice with the GLMMs, Sylvio Tüpke for creating the illustrations, and the animal caretakers of the zoo at Leipzig. Special thanks go to Laura Martínez Plantón and Nereida Bueno Guerra for their support with data collection, and Mar Pereto for her help with reliability coding.

Correspondence concerning this article should be addressed to Christoph J. Völter, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. E-mail: christoph_voelter@eva.mpg.de

(Reaux & Povinelli, 2000) learned to solve the task but only after considerable experience (50 to 200 trials) with the task. A control experiment showed that the capuchin monkey who succeeded merely had learned to insert the tool farthest from the reward without taking into account the position or functionality of the trap. The successful chimpanzees, in contrast, did not merely learn such a fixed distance-based rule but adapted to changes in the trap location flexibly in line with an appreciation of the causal relations of the task. However, given the high number of trials that was necessary for the chimpanzees to master this task, an alternative associative account explaining chimpanzees' behavior might have been the acquisition of a procedural rule to avoid the trap. Additionally, some conceptual problems with these control conditions have been raised, as human adults (presumably understanding the causal relations of the task) also tended to avoid nonfunctional traps (inverted traps facing upward) and exhibited a bias for inserting the tool farthest away from the reward (Silva, Page, & Silva, 2005).

Later studies demonstrated the complexity of this task by uncovering the detrimental impact of several task constraints on apes' performance. Mulcahy and Call (2006), for example, presented apes with a modified trap-tube task that allowed the apes to rake the reward out of the tube instead of pushing the reward away from their own body. Three of eight apes (two orangutans [*Pongo abelii*], and one chimpanzee) solved this task in 24 to 60 trials using raking rather than pushing the reward to succeed. These subjects also passed the inverted-tube control (i.e., they did not avoid the inverted, nonfunctional trap), suggesting an understanding of the causal relations of the task rather than the acquisition of a procedural rule to avoid the trap.

Nevertheless, this kind of task seems to be very hard for apes and monkeys, as shown by the small number of successful individuals and the extent of experience necessary to acquire the correct solution. In an attempt to shed light on the complexity of spatial relational reasoning, Fragaszy and Cummins-Sebree (2005; see also Visalberghi & Fragaszy, 2006) developed a framework of the complexity of a spatial problem according to (a) the number of object–object relations to be considered, (b) ego- or allocentric spatial relations (e.g., moving an object in relation to the own body or to another external object), (c) the precision of the spatial relations (i.e., whether specific actions are needed), (d) the temporal duration for which the spatial relations need to be controlled, and (e) whether multiple spatial relations are relevant simultaneously. According to this framework, the trap-tube problem and similar tasks involving tool use ranks among the most difficult problems, as it requires, in addition to the egocentric relation between hand and tool, the consideration of two allocentric, spatial relationships (a and b) simultaneously (e). These allocentric relations include the relations between stick and reward and between the reward and the surface of the tube (including the traps). Moreover, to successfully retrieve the reward, specific actions are needed (c) over an extended period of time (d). Thus, attending to and processing three spatial relationships simultaneously might be a particularly difficult problem due to capacity limitations. In other words, the trap-tube problem might be challenging because it involves two simultaneous tasks, tool use and the actual trap task, which might both tax the same cognitive capacities.

Thus, removing the requirement of tool use should facilitate apes' performance in trap tasks, as it eliminates the allocentric relation between tool and reward. And this is exactly what Seed et al. (2009) found: When chimpanzees faced a trap-task that required no tool, Seed and colleagues found that all eight chimpanzees learned to solve the problem in less than 100 trials. That is to say, when removing the need of using a tool, more chimpanzees learned to solve the task and in fewer trials compared with previously administered trap problems (e.g., Limongelli et al., 1995; Mulcahy & Call, 2006; Povinelli, 2000). In a transfer task, Seed et al. (2009) manipulated whether apes had to use a tool or their fingers to operate the apparatus and whether subjects were experienced with regard to the previously administered no-tool trap task or not. They found that experienced subjects performed better than inexperienced subjects and that non-tool-using apes performed better than tool-using ones. However, only one out of eight inexperienced chimpanzees (who was in the no-tool condition) solved the transfer task. Due to this floor effect of inexperienced subjects, it is difficult to interpret the effect of tool use on these naïve subjects.

Nevertheless, the data by Seed et al. (2009) suggest that despite chimpanzees' tool proficiency, the use of tools still imposes a considerable cognitive load on chimpanzees—a load that may negatively impact on the flexible deployment of cognitive abilities. Seed et al. proposed three candidates that might contribute to the cognitive load of tool use in this context: the attentional system, cross-modal matching, and increased response variability. First, considering two object–object relations (the tool-reward and reward-surface relation) simultaneously might demand the splitting of attention or the shifting of attention back and forth repeatedly. Second, cross-modal matching refers to the difficulty of judging functional properties such as solidity and continuity (important for anticipating the effect of the traps) based on vision alone. Accordingly, tool use might exacerbate task difficulty, as it prevents the subject from acquiring direct tactile information about some aspects of the task. However, the apes in Seed et al.'s trap task got no haptic feedback about the traps, either in the tool or no-tool conditions. This suggests that cross-modal matching alone is insufficient to account for the difficulties imposed by tool use. Third, tool use increases the requirements for manual dexterity necessary to move the reward toward a certain goal. All these factors might contribute, to some extent, to apes' difficulties with the classic trap problems.

Using a different experimental approach, Kaneko and Tomonaga (2012) found that when chimpanzees needed to differentiate between a self-controlled cursor (via a trackball) and a distractor cursor (using prerecorded motion patterns) on a computer screen, they mainly paid attention to the goal directedness of the cursor and not to the motion kinematics. The authors suggested that chimpanzees' sensorimotor action monitoring failed to make explicit judgments about external objects (such as the cursor on the computer screen or tools). Based on this argument, one would predict that apes might also have problems monitoring the movements of a stick tool (and its effect on the food reward) in relation to the target. In line with such a notion, Povinelli, Reaux, and Frey (2010) provided some evidence suggesting that chimpanzees maintain separable representations for their hand and stick tools. Thus, the tool is not just represented as an extension of the body schema; instead, it might require explicit rather than implicit action monitoring—which seems to be hard for chimpanzees, as demonstrated by Kaneko and Tomonaga's findings.

Predictive causal reasoning and/or the ability to plan actions ahead might also contribute to apes' difficulties in learning to avoid traps while using a tool. Tool use and trap tasks might involve predictive causal reasoning required to anticipate the effect of certain actions. In the current example, causal prediction means anticipating simultaneously the effect of moving the tool on the object, as well as the effect of moving the reward into the trap—something that might go beyond apes' capacity limits. Planning includes the predetermination of a sequence of subgoals toward an overarching goal before the current action is executed (Hayes-Roth & Hayes-Roth, 1979). Flexible tool use has repeatedly been suggested to involve this type of planning or subgoaling (e.g., Byrne, Morgan, & Sanz, 2013; Cox & Smitsman, 2006; Fragaszy, Johnson-Pynn, Hirsh, & Brakke, 2003). Accordingly, tool use increases the demands of trap tasks, as it introduces an additional step (i.e., a subgoal) in the means–end chain: Use the tool to move the reward (first step) and move the reward away from the trap (second step). Thus, limitations in apes' planning abilities may be responsible for the observed deficits in performance. These different factors are not mutually exclusive but are interrelated. The requirement of splitting one's attention, for example, might underlie simultaneous causal predictions and planning to some extent.

With regard to planning, we have recently shown that seven out of 12 great apes, as well as 4- and 5-year-old children, rapidly learned to avoid multiple traps in a maze-like task without the constraints identified in previous tasks (i.e., without the requirement of using a tool or of moving the reward away from one's own body; see Völter & Call, in press). Moreover, the decision-tree structure of that task allowed us to demonstrate that (younger) apes and children planned their actions up to two steps ahead. In the current study, we used the same task in combination with tool use to investigate the cognitive underpinnings of great apes' tool use in multistep settings. More specifically, we addressed three issues. First, we investigated whether the difficulties imposed by tool use were more pronounced when the trap configuration of the maze required planning ahead. If tool use requires the planning of actions beyond the current step, we hypothesized that apes' performance would be particularly affected in those trials in which they were required to plan their actions ahead. Moreover, apes that had to use a tool to displace the reward through the maze would perform worse than those apes who could use their fingers (the data of the latter group of subjects was obtained from Völter & Call, in press), particularly in trials requiring planning before making the first move.

Second, if tool use is limited by action monitoring of external objects, as suggested by recent evidence, we hypothesized that tool use would affect the general task performance of naïve subjects in the maze task because monitoring the tool might prevent subjects from paying attention to other task relevant aspects (e.g., traps). When allowed to move the reward with their fingers, action monitoring is thought to be implicit; thus, attending to the traps might be easier. However, the detrimental effects of tool use might subside with training as action monitoring becomes more implicit.

Third, manipulating the visual information about the function of the traps allowed us to assess the impact of causal knowledge on the task. If the apes were encoding functional information about the traps, we would expect that naïve subjects would perform better when they could see the traps blocking free passage. In contrast, if the apes solved the task by associating certain cues with success or failure, we expected them to perform on a similar level when presented with a version of the task that provided only noncausal (but equally salient and predictive) information about the trap location.

Finally, executive function (including planning) is known to decline with age in human and nonhuman primates (e.g., Bartus, Fleming, & Johnson, 1978; Hedden & Gabrieli, 2004; Lai, Moss, Killiany, Rosene, & Herndon, 1995; West, 1996). In particular, because age was a significant predictor of apes' performance in the no-tool version of the maze task (Völter & Call, in press), we examined whether younger apes would also perform better than older ones across the different experimental treatments of the present study.

To address each of these issues, we manipulated whether great apes (a) could operate the apparatus with their fingers or by means of a stick (Experiments 1 and 2), (b) obtained visual information on the function of the traps (Experiments 3 and 4), and (c) were naïve (Experiments 1 and 3) or experienced (Experiments 2 and 4) with the maze-like task. All comparisons between experimental manipulations reported throughout this article were planned before the first experiment was conducted (including the data reported in Völter & Call, in press). In addition, the assignment of subjects into different experimental groups was carried out before the first experiment was run.

Experiment 1

In this experiment, we compared two groups of nonhuman great apes. Apes in the tool group had to use a stick to move a reward through the maze-like trap task, whereas apes in the no-tool group (data obtained from Völter & Call, in press) could use their fingers to displace the reward. Both groups had not faced the task before, and the front side of the apparatus was transparent so that the functionality of the traps could be visually assessed.

Method

Subjects. Four bonobos (*Pan paniscus*), 14 chimpanzees (*Pan troglodytes*), and three orangutans (*Pongo abelii*) participated in this experiment. The subjects were housed at the Wolfgang Köhler Primate Research Center at the Leipzig Zoo (Leipzig, Germany). There were 13 females and eight males aged between 6 and 35 years ($M_{\text{age}} = 18.9$ years). Ten subjects were nursery-reared and 11 were mother-reared (see Table 1 for detailed information on each subject). Subjects lived in social groups of different sizes and had access to indoor (175 m^2 to 430 m^2) and outdoor areas (1,400 m^2 to 4,000 m^2). They were tested individually in special testing rooms (5.1 m^2 to 7.3 m^2). Subjects were not deprived of food, and water was available ad libitum during testing.

Subjects had participated in various cognitive tasks prior to the study. Eleven of the 14 chimpanzees of the current sample had participated in Seed et al. (2009), a study that is particularly relevant here. As in the current study, the subjects in Seed et al. also had to move a reward with their fingers either to the left or to the right while avoiding traps. In contrast to the present task, however, the Seed et al. task required no planning beyond the current decision. Additionally, Seed et al.'s apparatus (including its traps) differed in appearance from the ones used here.

Materials. The apparatus consisted of quadrangular box (height × length × depth = 47 cm × 67 cm × 5 cm) mounted to

Table 1
Species, Age, Sex, Rearing History, Condition (Refers to Experiment 1 and 3), and the Experiments in Which the Subjects Participated

Name	Species	Sex	Age	Rearing history	Condition	Experiment participation
Kuno	Bonobo	M	14	Nursery	No tool	1, 2, 3, 4
Ulindi	Bonobo	F	17	Mother	No tool	1, 3
Joey	Bonobo	M	28	Nursery	Tool	1, 3
Lexi	Bonobo	F	12	Nursery	Tool	1, 3
Jasongo	Bonobo	M	20	Mother	Opaque-cued	3
Yasa	Bonobo	F	14	Mother	Opaque-cued	3
Alex	Chimpanzee	M	10	Nursery	No tool	1, 2, 3, 4
Fifi	Chimpanzee	F	18	Mother	No tool	1, 2, 3, 4
Kofi	Chimpanzee	F	6	Mother	No tool	1, 2, 3, 4
Pia	Chimpanzee	F	12	Mother	No tool	1, 2, 3, 4
Riet	Chimpanzee	M	33	Nursery	No tool	1, 3
Robert	Chimpanzee	M	35	Nursery	No tool	1, 3
Sandra	Chimpanzee	F	18	Mother	No tool	1, 2, 3, 4
Annett	Chimpanzee	F	12	Nursery	Tool	1, 3
Corry	Chimpanzee	F	34	Nursery	Tool	1, 3
Lobo	Chimpanzee	M	7	Mother	Tool	1, 3
Swela	Chimpanzee	F	15	Mother	Tool	1, 3
Tai	Chimpanzee	F	9	Mother	Tool	1, 3
Trudi	Chimpanzee	F	18	Mother	Tool	1, 3
Ulla	Chimpanzee	F	34	Nursery	Tool	1, 3
Alexandra	Chimpanzee	F	12	Nursery	Opaque-cued	3
Dorien	Chimpanzee	F	30	Nursery	Opaque-cued	3
Fraukje	Chimpanzee	F	35	Nursery	Opaque-cued	3
Frodo	Chimpanzee	M	17	Mother	Opaque-cued	3
Jahaga	Chimpanzee	F	18	Mother	Opaque-cued	3
Kara	Chimpanzee	F	6	Mother	Opaque-cued	3
Lome	Chimpanzee	M	10	Mother	Opaque-cued	3
Kila	Orangutan	F	11	Mother	No tool	1, 2, 3, 4
Pini	Orangutan	F	23	Mother	No tool	1, 3
Bimbo	Orangutan	M	31	Nursery	Tool	1, 3
Dokana	Orangutan	F	22	Mother	Opaque-cued	3
Padana	Orangutan	F	13	Mother	Opaque-cued	3

the wall of the enclosure (see Figure 1). Inside the box there was a vertical maze consisting of three horizontal levels that were made of gray PVC material. Open gaps (width = 3 cm) located in these levels allowed to pass the food reward that was placed in the maze

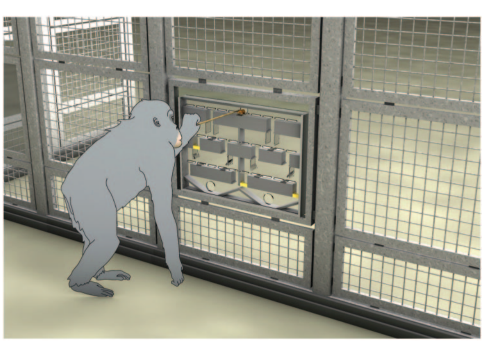

Figure 1. Illustration of an ape working on the maze apparatus in the tool condition. The trap configuration shows an example of level of planning (LoP) 2 and change in direction (CiD) 1, that is, the ape needed to consider traps located two levels ahead in their first decision, and they were required to change the direction of the reward once before they could retrieve it. The openings in the backside of the apparatus that allowed entering and removing the yellow trap elements are not depicted here for the sake of convenience. The color version of this figure appears in the online article only.

on to the next lower levels. In total, there were 10 gaps located over three of the maze's levels. The gaps were distributed symmetrically, that is, when our subjects moved the reward into a gap, it always fell in-between two gaps on the next level. On the first (uppermost) level, there were two gaps; on the second and third level, there were four gaps each. Furthermore, there was a vertical partition located in the middle of the apparatus between the second and third gap on the second and third level. Underneath each opening there were two transparent pieces of acrylic glass that channeled the food reward and a patch of rubber material glued to the apparatus to cushion the food reward when it fell from one level to the next. Both of these additional items served to prevent the food reward from accidentally skipping one level.

The front side of the apparatus (which was accessible to the apes) was made of transparent acrylic glass and contained three horizontal rows of 10 slits each (height × width = 1.8 cm × 5 cm; distance between the slits = 1 cm) that allowed the apes in the no-tool condition to stick in their fingers, and to move the food reward to the left and to the right on all three levels (see Figure 2a). In the tool condition there was an additional transparent panel mounted to the front side of the apparatus. This additional panel had 1-cm-wide slits on the same level as the openings for the fingers and two circular holes on the same level as the holes in the apparatus (see Figure 2b). This add-on prevented the apes from sticking in their fingers into the apparatus, as these slits were too

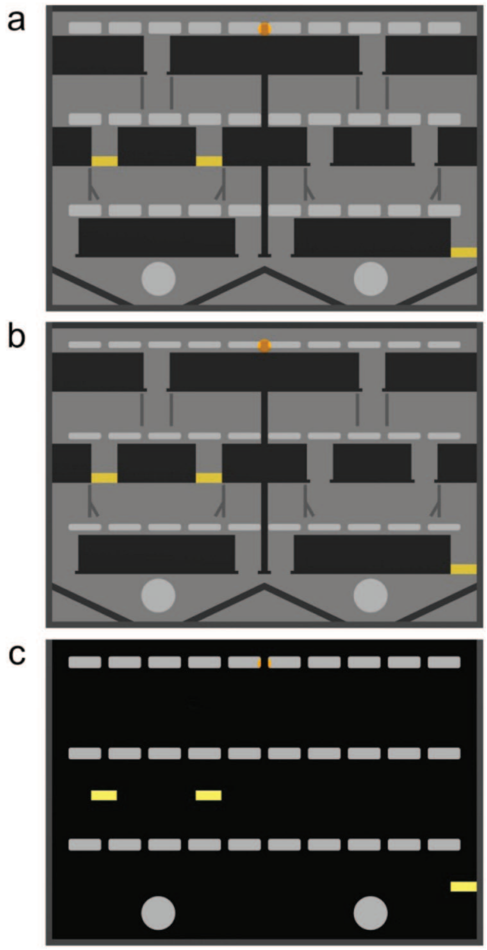

Figure 2. Examples of the three experimental conditions used across the current study: (a) no-tool condition, (b) tool condition, and (c) opaque-cued condition. In the no-tool (a) and opaque-cued condition (c), the apes could negotiate the reward (located at the uppermost level) through the maze directly with their fingers via slits in the front side of the apparatus. In the tool condition, these slits were too narrow for the apes' fingers. Therefore, the apes needed to use a stick tool to move the reward. The yellow squares show the traps that block free passage through the gaps. In the no-tool and tool condition, the apes could visually inspect the function of the traps, whereas in the opaque-cued version, the apes merely saw color cues. The depicted trap configuration is an example of level of planning (LoP) 1 and change in direction (CiD) 1, that is, the apes needed to consider traps located one level ahead in their first decision, and they were required to change the direction of the reward once before they could retrieve it. The color version of this figure appears in the online article only.

narrow. In the tool condition, the apes were provided with 25-cm-long wooden sticks (diameter = 0.6 cm). The apes in both conditions could extract the food reward from the apparatus via two large, circular holes on the lower side of the maze (diameter = 5.4 cm). Two ramps on the left and right of each hole ensured that the reward would roll behind one of the holes when the subjects displaced the reward from the third level to the bottom of the apparatus.

The experimenter could access the backside of the apparatus, made of transparent acrylic glass. It contained a circular opening in the middle of the uppermost level that we used for baiting purposes. Furthermore, there were 10 openings (height × length = 3.0 cm × 4.5 cm) corresponding to the location of each of the maze's gaps. The experimenter inserted yellow traps through these openings to block the gaps, and thus to prevent the passage of the reward. The subjects could visually inspect the yellow traps, made of painted acrylic glass (height × length × depth = 1.5 cm × 4 cm × 3.5 cm) when the experimenter inserted them in the apparatus. We used monkey chow pellets (height × length × depth = 2.0 cm × 3.0 cm × 2.0 cm) as rewards. The pieces of monkey chow were solid and thick enough to prevent subjects from just pulling them through the slits in the front side of the apparatus.

Procedure and design. At the beginning of each trial, the experimenter placed three traps into the apparatus in full view of the subjects so that they could visually assess the traps and their function (blockage of free passage). Then the experimenter introduced the reward into the apparatus via the baiting hole in the backside of the apparatus on the uppermost level. In the tool condition, the experimenter gave the tool to the ape by introducing it through the mesh underneath the apparatus. The trial ended when the subject obtained the reward or pushed it into a trap. When the subject obtained the reward, we moved on to the next configuration. When the subject failed, the experimenter removed the trapped reward and dropped it into the food bucket. After 3 to 5 s, the experimenter started the next trial by inserting another piece of food into the apparatus.

Half of the subjects received the no-tool condition, and the other half received the tool condition. Assignment to the groups was random, with the restriction that both groups were counterbalanced as much as possible for species, age, and sex (no-tool: M_{age} = 17.9, n = 11, 64% females; tool: M_{age} = 20.0, n = 10, 70% females). There were two rounds of 24 trial-unique trap configurations per subject (48 configurations in total). We employed a maximum of 16 trials per configuration. When the subjects failed to obtain the reward after 16 trials with a given configuration, we skipped it and administered the next configuration. Each session consisted of a maximum of 16 trials or three configurations (depending on which criterion was reached first).

There were two independent variables regarding maze complexity: level of planning (LoP) and changes in direction (CiD). With regard to the LoP, the subjects had to consider, at the beginning of a trial, only the first (uppermost) level, the first and second (intermediate) level, or all three levels to obtain the reward. For each LoP, there were eight different configurations: In LoP 0, one trap was located on the uppermost level; the other two traps were either both on the second level or on the third level, either at Positions 1 and 4 or at Positions 2 and 3. To solve such configurations, the apes only had to take into account the traps at the current level, that is, the level where the reward was currently located. In LoP 1, two traps were blocking both openings on one side of the second level; the third trap was located on the other side, either also on the second or third level (for an example, see Figure 2). Hence, the apes could not solve this task by only taking into account the uppermost level in the beginning of a trial (as there were no traps in the uppermost level). Instead, they had to look one level ahead, that is, when making the first decision on the uppermost level, the traps in the second level had to be considered. For LoP 2, two traps were placed on one side of the third level (i.e., this side was completely blocked; for an example, see Figure 1), and the third trap was on the other side, either also on the third or the second level. Therefore, the apes had to consider the traps on the third level when the reward was still at the uppermost level when making the decision where to move the food reward. Hence, LoP was a measure of spatial distance of the task-relevant items (i.e., the traps) from the starting point. The factor LoP allowed us to manipulate how many subgoals the subjects had to consider in making an informed first decision.

The second independent variable was the number of CiDs that the subject had to perform to gain the reward after making the first decision. In half of the 24 configurations, there was no change in direction necessary, that is, the apes either had to push the reward completely to the left or to the right to get the reward. In the other half of configurations, they had to change the direction of the reward once, that is, after moving to the right or left side on the uppermost level, the apes had to change the direction of the reward in the second or third level to the opposite side. Thus, CiD can be considered as measure of path complexity by manipulating the degree of motor control required to solve the maze. The number of CiDs was completely balanced across the different LoP configurations: For each LoP, four configurations involved no change in direction (CiD 0) and four involved one change in direction (CiD 1). The "correct" side of the maze was balanced across the 24 configurations. Finally, the order of configurations was pseudorandomized, with the restriction that no more than two configurations of the same LoP were presented in a row.

Scoring and analysis. We videotaped all trials. We scored the following three dependent measures: (a) percentage of configurations in which subjects obtained the reward in the first trial (T1 success, chance level = 25% correct), (b) percentage of T1 trials in which subjects moved the reward toward the correct side in the uppermost level for the final solution (T1 first decision, chance level = 50% correct), and (c) the mean number of trials per configuration the subjects needed to obtain the reward (chance level = 4 trials; for the results of this variable, see supplemental results of the online supplemental materials). A second coder scored 20% of the trials to assess interobserver reliability, which, according to Fleiss (1981), is excellent (Cohen's κ: clear condition, T1 success $\kappa = 1.0$, $n = 113$, $p < .001$; T1 first decision $\kappa = 1.0$, $n = 113$, $p < .001$; tool condition, T1 success $\kappa = 0.98$, $n = 105$, $p < .001$; T1 first decision $\kappa = 0.98$, $n = 105$, $p < .001$).

Above-chance performance in T1 success was indicative of successful sequential decision making without necessarily taking into account upcoming levels. If the first decision was made completely randomly, subjects could still have obtained the reward in 50% of T1 trials (which was significantly above the chance level of 25% correct). Above-chance performance in T1 first decision (in particular in LoP 1 and LoP 2 configurations) was indicative of planning, that is, upcoming levels were considered when the first decision was made.

We used Pearson's correlations to test the relation of age with the dependent variables. All p values reported here are exact and two-tailed. The assumption of normality was met for the current data (Kolmogorov–Smirnov test: $p > .05$). At the individual level, we used binomial tests for the binary variables T1 success and T1 first decision.

To test whether the dependent variables T1 success and T1 first decision were influenced by the factors condition (tool, no-tool), LoP, CiD, repetition of configurations, and the age of the subjects we used a generalized linear mixed model (GLMM; Baayen, 2008) that included these five predictors as fixed effects, and subject as well as configuration identity as random effects. The models were fitted in R (R Development Core Team, 2011) using the function lmer of the R-package lme4 (Bates & Maechler, 2010). The significance of the full model compared with the null model (comprising only random effects; Forstmeier & Schielzeth, 2011) was established using a likelihood ratio test (Dobson, 2002). Therefore, we used the R function anova with argument test set to "Chisq." All models reported here were found to be significant ($p < .01$).

The response variables "success in T1" and "first decision in T1" were binary (success/failure); therefore, we specified binomial errors and the "logit" link function. We z-transformed all predictors to a mean of zero and a standard deviation of one to obtain comparable estimates. The intercepts of the models represented the sample mean assumed by the models. In the logistic models, the fitted mean is revealed by the inverse logit-transformation of the intercept (i.e., exp[intercept]/(1 + exp[intercept]). The corresponding p value of the intercept indicates whether the intercept deviates significantly from the equal distribution, which was the null hypothesis for the variable "first decision in T1" (proportion of correct decisions = 0.5). However, for the dependent variable "success in T1," the chance value was 0.25. Therefore, we subtracted the logit-transformed chance value from the estimate of the intercept and calculated the corresponding z and p value based on this adjusted estimate. However, because the dependent variables were not based on a simple linear function of the given predictor variables in the models, there was a minimal deviation of the sample mean assumed by the model from the actual sample mean. We corrected for this small deviation by adjusting the scaled variables by adding a constant value chosen such that the absolute difference between the actual sample mean and the fitted mean was minimized (the corresponding function was written by Roger Mundry and is available upon request). Doing so did not affect any terms of the model except for the intercept. Thereby, the intercept in these models became a reliable test of subjects' performance against chance while controlling for the covariates and random effects.

Results

Success in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: $\chi^2 = 44.86$, $df = 7$, $p < .001$). The GLMM indicated significant effects of repetition of configurations, age, and a significant interaction between tool use and CiD (the main effects of tool use and CiD are therefore not considered) on success in T1 (see Table 2): Subjects improved in the second round compared with the first one, and younger subjects performed better than older ones. Correlation analyses revealed that the age effect was more pronounced in the no-tool condition (Pearson r: no-tool $r = -0.72$, $n = 11$, $p = .012$; tool $r = -0.58$, $n = 10$, $p = .081$). Furthermore, non-tool-using subjects performed better than tool users when no change in direction was required ($z = -2.50$, $p = .013$), but not when a change in direction was necessary ($z = -0.58$, $p = .559$; see Figure 3). In contrast, there was no significant main effect of LoP or an interaction of LoP with tool use.

Table 2
Experiment 1: Output of GLMMs for the Different Dependent Variables

	T1 success			T1 first decision		
Model terms	Est	95% CI	*p*	Est	95% CI	*p*
Tool use	−0.46	[−0.96, 0.04]	0.07	−0.40	[−0.76, −0.05]	0.03
LoP	−0.04	[−0.23, 0.14]	0.64	−0.06	[−0.24, 0.12]	0.50
CiD	−0.34	[−0.52, −0.15]	<0.01	−0.14	[−0.32, 0.04]	0.12
Age	−0.51	[−0.76, −0.25]	<0.01	−0.30	[−0.47, −0.12]	<0.01
Repetition	0.29	[0.15, 0.43]	<0.01	0.17	[0.04, 0.29]	0.01
Tool Use × LoP	−0.05	[−0.33, 0.23]	0.72	−0.02	[−0.28, 0.23]	0.86
Tool Use × CiD	0.31	[0.04, 0.59]	0.03	0.23	[−0.03, 0.48]	0.08

Note. GLMM = generalized linear mixed model; T1 = trial 1; CI = confidence interval; LoP = level of planning; CiD = changes in direction.

Overall, tool users did not perform above chance (29.2 ± 4.1% correct; $z = 1.30$, $p = .194$), whereas non-tool-users did (41.5 ± 5.9% correct; $z = 4.07$, $p < .001$). In line with the results of the GLMM, only subjects in the no-tool condition scored above chance when no change in direction was required (CiD 0) and in the second round (but not in the first one). Across LoP, non-tool-using subjects solved more trials in T1 than expected by chance. At the individual level, four chimpanzees and one bonobo performed above chance overall in the no-tool condition (all *p*s < .001). In the tool-use condition, there were two chimpanzees that performed above chance ($p < .01$).

First decision in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: $\chi^2 = 23.56$, $df = 7$, $p = .001$). The GLMM indicated a significant effect of tool use, age, and repetition (see Table 2): Non-tool-using subjects performed better than tool users, younger subjects performed better than older ones, and subjects performed better in the second round. Again, the age effect on performance was mainly driven by the non-tool-using subjects (Pearson correlation: no-tool $r = -0.85$, $n = 11$, $p < .001$; tool $r = -0.22$, $n = 10$, $p = .548$; see also Figure 4). In contrast, there was no significant main effect of LoP and CiD, or an interaction of LoP and CiD with tool use (see Figure 5).

Overall, non-tool-using subjects performed better than expected by chance in their first decision in T1 (59.3 ± 4.6% correct; $z = 3.55$, $p < .001$), whereas tool users performed at chance (48.3 ± 3.6% correct; $z = 0.68$, $p = .500$). Considering the individual data, we found that, overall, four non-tool-using subjects performed significantly above chance (three chimpanzees and one bonobo, all *p*s < .05). One tool-using chimpanzee tended to perform above chance (Annett, $p = .059$; see Video 1 of the online supplemental materials).

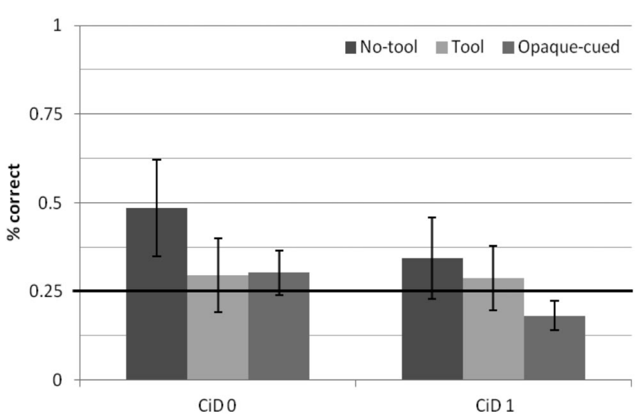

Figure 3. Experiments 1 and 3: Trial 1 success (mean ± 95% CI) as function of changes in direction (CiD) and condition.

Discussion

All apes in the tool condition spontaneously used the stick to negotiate the reward through the maze. However, using the tool prevented most of the apes from learning to avoid the traps, that is, they moved the reward by means of the stick randomly to the left or to the right without considering the yellow traps that blocked free passage to the next level (except for two chimpanzees who learned to avoid the traps in the tool condition). This sharply contrasts with the apes who were allowed to use their fingers to operate the apparatus. These apes did in fact learn to avoid the traps in the first trial of a given configuration within the two rounds of 24-trap configurations, and some of the apes—in particular, the younger ones—even planned their moves up to two steps ahead (see Völter & Call, in press, for an extended discussion on the latter finding). Even in those configurations without a change in direction on the second and third levels (a feature that makes configurations easier to solve), subjects using the tools, unlike those not using them, performed randomly. However, both groups failed in those configurations requiring a change in direction.

In line with previous research (Seed et al., 2009), these results suggest that tool use places a significant cognitive load on the apes, preventing them from learning to avoid the traps. In contrast, tool use did not impact the number of trials per configuration (see the online supplemental results), suggesting no influence of tool use on post-error adjustments. A possible explanation for the lack of perseveration in the tool-use condition is that in the no-tool condition, the apes often switched their hands to operate on the apparatus when changing the direction of the reward (as required in CiD 1 configurations). When using a tool, such a change of the operating hand was not observed. Changing the direction of the reward therefore seems to be more costly in the no-tool condition than in the tool condition.

A potential candidate for the cognitive load imposed by tool use is action monitoring of an external object. Accordingly, apes had

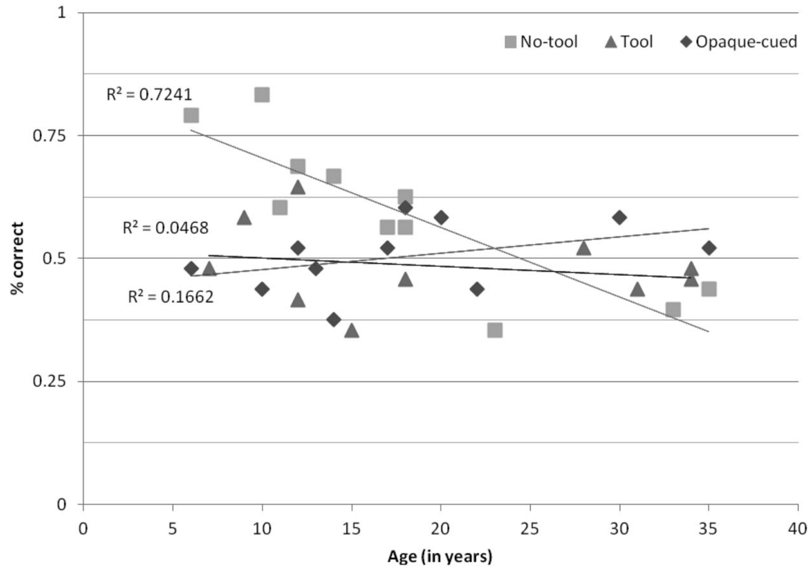

Figure 4. Experiments 1 and 3: Trial 1 first decision as function of age and condition.

to explicitly monitor their actions to move the reward by means of the stick, but not when they could simply use their fingers to move the reward, in which case action monitoring was presumably implicit and thus less demanding. Previous research suggests monitoring the actions of an external effector toward a target is difficult for chimpanzees (Kaneko & Tomonaga, 2012), and that stick tool use involves a separate representation of the tool and not just an extension of the body schema (Povinelli et al., 2010). The current results can be reconciled with these previous findings. Maintaining such a representation might be costly and prevent the apes from simultaneously attending to other task-relevant cues, such as the yellow traps. In contrast, we found no evidence that tool use was related to the ability to plan multiple steps ahead (as indicated by the lack of a significant interaction between tool use and LoP). However, as the tool-using subjects failed to learn to avoid the traps, which was a prerequisite for planning in the current task, this conclusion awaits further confirmation.

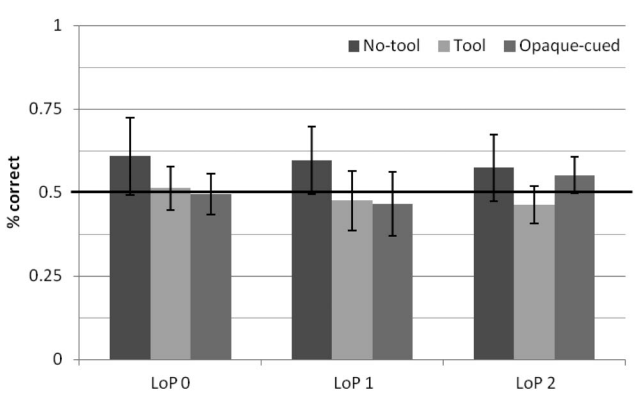

Figure 5. Experiments 1 and 3: Trial 1 first decision (mean ± 95% CI) as function of level of planning (LoP) and condition.

Experiment 2

In Experiment 1, we established that tool use imposed a significant cognitive load on great apes. In particular, we found that the need to use a tool prevented the apes from solving the maze task above chance levels in the first trial of each configuration. The question arises whether tool use would also have a detrimental effect on apes who had experience with the no-tool version of the task and thus already had learned to avoid the traps, or whether these difficulties introduced by tool use were specific to the acquisition phase of the task.

Method

Subjects. The seven subjects who passed the no-tool condition in Experiment 1 participated in this experiment. They included one bonobo, five chimpanzees, and one orangutan (see Table 1). There were four females and three males between the ages of 6 and 18 years (M_{age} = 12.7 years). Two subjects were nursery-reared and five were mother-reared.

Between Experiment 1 and the current experiment, the apes were presented with different follow-ups of the same experimental setup, including a version of the task with four traps inserted, a visually restricted version of the apparatus (opaque-cued condition; Experiment 4), and a memory rehearsal condition (in which the apes were to memorize the location of the traps before operating on the apparatus). Thus, by the time the current experiment took place, these seven individuals were experts with regard to the no-tool version of the task.

Materials. The apparatus was identical to that used in Experiment 1.

Procedure and design. The procedure was identical to Experiment 1, with the exception that our subjects were presented with both conditions—the tool and no-tool conditions—and we administered one round of 24 trial-unique configurations per subject and condition. The order of condition was balanced across subjects. There was a maximum of 24 trials or eight configurations (depending on which criterion was reached first) per session. The order of presentations was pseudorandomized, with the restriction that no more than two configurations of the same LoP were presented in a row.

Scoring and analysis. The scoring and analysis were the same as in Experiment 1.

Results

Success in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: $\chi^2 = 28.63$, $df = 7$, $p < .001$). The GLMM indicated significant effects of age, LoP, and CiD, but not tool use (see Table 3): Younger subjects performed better than older ones, subjects performed better in LoP 0 ($z = -2.71$, $p = .007$) and LoP 1 ($z = -3.07$, $p = .002$) than in LoP 2; no difference was found between LoP 0 and LoP 1 ($z = 0.63$, $p = .529$), and subjects performed better when no change in direction was required (see Figure 6). Correlation analyses revealed that the age effect was more pronounced in the novel condition involving tool use (Pearson correlation: no-tool $r = -0.54$, $n = 7$, $p = .214$; tool $r = -0.78$, $n = 7$, $p = .039$). There was no significant interaction of LoP and CiD with tool use.

Overall, subjects performed above chance in the tool condition (77.4 ± 4.3% correct; $z = 4.74$, $p < .001$) as well as in the no-tool condition (81.5 ± 3.1% correct; $z = 6.61$, $p < .001$). This was also true across LoP and CiD, respectively. At the individual level, all subjects performed above chance overall in both the tool and no-tool conditions (all $ps < .001$).

First decision in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: $\chi^2 = 24.41$, $df = 7$, $p < .001$). The GLMM indicated significant effects of age, LoP, and CiD, but no effect of tool use (see Table 3): Younger subjects performed better than older ones, subjects performed better in LoP 0 ($z = -2.71$, $p = .007$) and LoP 1 ($z = -3.06$, $p = .002$) than in LoP 2 (see Figure 7; no difference was found between LoP 0 and LoP 1, $z = 0.64$, $p = .522$), and subjects performed better when no change in direction was required. Correlation analyses revealed that the age effect was mainly driven by the novel condition involving tool use (Pearson correlation: no-tool $r = -0.54$, $n = 7$, $p = .214$; tool $r = -0.81$, $n = 7$, $p = .026$). There was no significant interaction of LoP and CiD with tool use.

Overall, subjects performed above chance in the tool condition (80.9 ± 4.5% correct; $z = 4.08$, $p < .001$) as well as in the no-tool condition (81.5 ± 3.1% correct; $z = 3.80$, $p < .001$). This was also true across the different levels of planning and CiD. At the individual level, all subjects except one chimpanzee performed above chance in the no-tool condition ($p < .05$). In the tool condition, all except two chimpanzees performed above chance ($p < .05$).

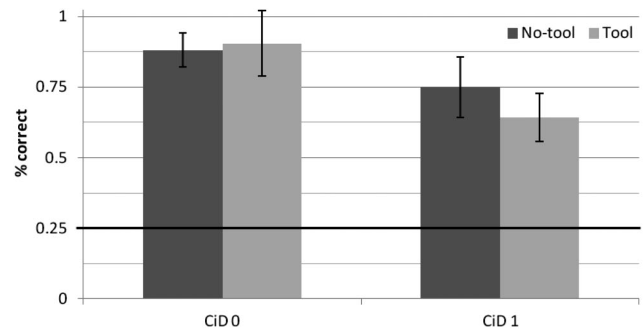

Figure 6. Experiment 2: Trial 1 success (mean ± 95% CI) as function of changes in direction (CiD) and tool use.

Discussion

The experienced subjects who had learned to avoid the traps using their fingers exhibited no performance decrement when we introduced the need to use a tool. Moreover, there was no effect of tool use on apes' ability to plan their actions ahead (LoP) or on motor control (CiD). This indicates that the detrimental effect of tool use on apes' performance in the current maze task was specific to the acquisition phase, that is, when apes were learning to avoid the traps. These data also suggest that the increased motor demands of tool use alone are insufficient to explain the deficits of the naïve apes in the tool condition of Experiment 1. Moreover, we expected that if tool use involved planning actions ahead, it should particularly disrupt apes' performance when the task required planning ahead. However, as the experienced apes were able to plan their moves also in the tool condition, we found no empirical support for the hypothesis that tool use disrupted apes' ability for planning ahead.

Thus, negotiating the reward through the maze with a tool while avoiding the traps seems to be difficult for apes only when the task is novel. Recent findings indicate that chimpanzees tend to fixate on the target of an external effector (Kaneko & Tomonaga, 2012).

Table 3
Experiment 2: Output of GLMMs for the Different Dependent Variables

	T1 success			T1 first decision		
Model terms	Est	95% CI	*p*	Est	95% CI	*p*
Tool use	−0.14	[−0.82, 0.54]	0.69	0.04	[−0.66, 0.74]	0.91
LoP	−0.64	[−1.10, −0.18]	0.01	−0.64	[−1.11, −0.17]	0.01
CiD	−0.49	[−0.94, −0.04]	0.03	−0.49	[−0.96, −0.03]	0.04
Age	−0.45	[−0.75, −0.15]	<0.01	−0.51	[−0.83, −0.20]	<0.01
Order of condition	−0.08	[−0.38, 0.22]	0.60	−0.07	[−0.39, 0.24]	0.66
Tool use × LoP	0.10	[−0.51, 0.70]	0.76	0.11	[−0.51, 0.72]	0.73
Tool use × CiD	−0.42	[−1.05, 0.20]	0.19	−0.24	[−0.87, 0.39]	0.46

Note. GLMM = generalized linear mixed model; T1 = trial 1; CI = confidence interval; LoP = level of planning; CiD = changes in direction.

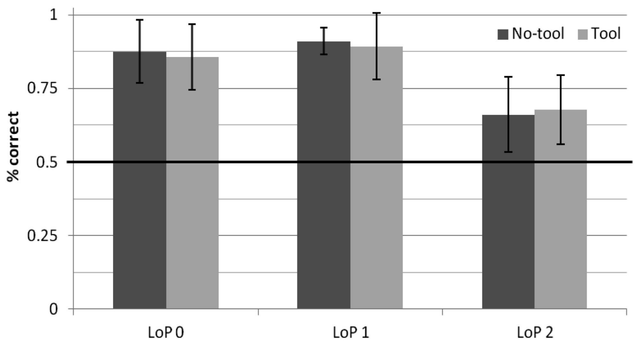

Figure 7. Experiment 2: Trial 1 first decision (mean ± 95% CI) as function of level of planning (LoP) and tool use.

Accordingly, the introduction of a tool might lead to a shift in apes' attention toward the target of the external effector, that is, the food reward at the endpoint of the stick tool. However, if tool use makes the apes fixate on the reward located at the endpoint of the stick tool, this might prevent them from paying attention to the traps. For experienced apes, this process of avoiding the traps might have become so implicit that it required less attention.

Interestingly, the negative relation between age and planning performance observed for naïve apes using their fingers, but not for tools (see Experiment 1), was also present in the novel tool-use condition. This means that the age effect was especially pronounced when we confronted the apes with a novel situation, thus supporting the idea of higher cognitive flexibility of younger apes compared with older ones.

Experiment 3

Next we addressed the question of what naïve apes had learned about the traps in Experiment 1, that is, whether they avoided the traps based on causal or functional knowledge, or on associative knowledge. Therefore, we introduced a control condition in the current setup in which we maintained the color information but removed functional information about the traps, that is, blockage of free passage. We expected that if the apes avoided the traps in Experiment 1 based on purely associative knowledge, the performance of the no-tool condition of Experiment 1 and the opaque-cued control condition should be similar. Furthermore, we wanted to know how subjects in the visually restricted version would perform compared with the subjects in the tool condition.

Method

Subjects. Two bonobos (*Pan paniscus*), seven chimpanzees (*Pan troglodytes*), and two orangutans (*Pongo abelii*) were tested in Experiment 3 with the novel opaque-cued condition. All subjects were naïve with regard to the maze task. We compared these data with the data for the tool and no-tool conditions (Experiment 1). Assignment to the conditions (no-tool, tool, and opaque-cued) was random, with the restriction that the groups were counterbalanced as much as possible for species, age, and sex (no-tool: M_{age} 17.9, $n = 11$, 64% females; tool: M_{age} 20.0, $n = 10$, 70% females; opaque-cued: M_{age} 17.9, $n = 11$, 73% females).

Materials. The apparatus for the opaque-cued condition was identical in construction to that used in the no-tool condition, with the modification that its front side was painted black, except at the locations where the traps could be introduced. At these locations, a 1.2 cm × 4 cm portion was left unpainted so that black or yellow screens (height × length = 3.5 × 4 cm) inserted from the backside of the apparatus could be seen from the front. Yellow screens indicated the location of the traps; black screens indicated free passage (see Figure 2c). The traps in the opaque-cued condition were transparent pieces of acrylic glass (4.2 × 1.9 cm) that could not be seen by the apes when inserted in the apparatus behind the preinserted screens.

Procedure and design. The procedure was similar to Experiment 1, with the exception that in the opaque-cued condition, the experimenter first inserted the screens that were occluding the view to the gaps before the first trial of each configuration. There were seven black (i.e., no trap present) screens and three yellow (i.e., trap present) screens. The actual traps in the opaque-cued condition were introduced after the screens were in place and thus outside the subjects' view.

Scoring and analysis. We scored and analyzed the data in the same way as in previous experiments. Additionally, we compared the data for the tool and no-tool condition of Experiment 1 with the new data obtained with the opaque-cued condition. A second coder scored 20% of the opaque-cued trials to assess interobserver reliability, which, according to Fleiss (1981), is excellent (Cohen's κ: T1 success $\kappa = 0.97$, $n = 118$, $p < .001$; T1 first decision $\kappa = 0.93$, $n = 118$, $p < .001$).

Results

Success in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: $\chi^2 = 51.17$, $df = 10$, $p < .001$). The GLMM indicated significant effects of repetition of configurations, age, CiD, the opaque-cued versus no-tool condition, and a significant Opaque-Cued Versus Tool × CiD interaction on success in T1 (see Table 4): subjects improved in the second round compared with the first one, younger subjects performed better than older ones, and subjects in the no-tool condition (front side of the apparatus transparent) outperformed subjects in the opaque-cued condition. Correlation analyses revealed that the age effect was not driven by the subjects in the opaque-cued condition (Pearson correlation: $r = -0.16$, $n = 11$, $p = .630$) but by the subjects in the no-tool condition (see Experiment 1). Furthermore, tool-using subjects performed better than subjects in the opaque-cued condition when a change in direction was required ($z = 2.01$, $p = .045$), but not when no change in direction was necessary ($z = 1.27$, $p = .204$; see Figure 3). In contrast, there was no significant main effect of LoP or a Condition × LoP interaction.

Overall, subjects in the opaque-cued condition did not perform above chance (24.2 ± 1.6% correct; $z = -0.3$, $p = .774$), in line with the tool-using subjects and in contrast to non-tool-using subjects (see Experiment 1). In line with the results of the GLMM, subjects in the opaque-cued condition scored below chance when a change in direction was required (CiD 1: $z = -2.27$, $p = .023$), but at chance for CiD 0 configurations ($z = 1.46$, $p = .145$). At the individual level, there was no subject who scored above chance in the opaque-cued condition (all ps > .1).

Table 4
Experiment 3: Output of GLMMs for the Different Dependent Variables

	T1 success			T1 first decision		
Model terms	Est	95% CI	*p*	Est	95% CI	*p*
Opaque-cued vs. No tool	0.83	[0.38, 1.28]	<0.01	0.38	[0.04, 0.72]	0.03
Opaque-cued vs. Tool	0.35	[−0.11, 0.81]	0.14	−0.05	[−0.39, 0.30]	0.80
LoP	0.16	[−0.06, 0.37]	0.15	0.09	[−0.08, 0.27]	0.28
CiD	−0.35	[−0.57, −0.13]	<0.01	−0.08	[−0.25, 0.09]	0.38
Age	−0.37	[−0.55, −0.18]	<0.01	−0.16	[−0.31, −0.02]	0.02
Repetition	0.20	[0.09, 0.31]	<0.01	0.10	[−0.001, 0.20]	0.05
Opaque-cued vs. No tool × LoP	−0.20	[−0.48, 0.07]	0.15	−0.16	[−0.40, 0.09]	0.22
Opaque-cued vs. Tool × LoP	−0.25	[−0.54, 0.04]	0.09	−0.18	[−0.43, 0.07]	0.16
Opaque-cued vs. No tool × CiD	0.02	[−0.26, 0.30]	0.89	−0.06	[−0.31, 0.18]	0.62
Opaque-cued vs. Tool × CiD	0.33	[0.04, 0.62]	0.03	0.16	[−0.09, 0.41]	0.20

Note. GLMM = generalized linear mixed model; T1 = trial 1; CI = confidence interval; LoP = level of planning; CiD = changes in direction.

First decision in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: $\chi^2 = 21.14$, $df = 10$, $p = .020$). The GLMM indicated a significant effect of age, the opaque-cued versus no-tool condition, and a marginally significant effect of repetition of configurations (see Table 4): younger subjects performed better than older ones, subjects in the no-tool condition (in contrast to subjects in the tool condition) outperformed those in the opaque-cued condition, and subjects tended to perform better in the second round. Again, the age effect on performance was not driven by subjects in the opaque-cued condition (Pearson correlation: r = .41, $n = 11$, $p = .213$; see also Figure 4) but by the subjects in the no-tool condition (see Experiment 1). In contrast, there were no significant main effects of LoP, CiD, or interactions between condition and LoP (see Figure 5) or CiD.

Overall, subjects in the opaque-cued condition did not perform above chance (50.4 ± 2.1% correct; $z = 0.17$, $p = .862$), in line with the tool-using subjects and in contrast to non-tool-using subjects (see Experiment 1). At the individual level, there was no subject who scored above chance overall in the opaque-cued condition (all *p*s > .1).

Discussion

All apes in the opaque-cued condition (i.e., that did not get information about the causal relations of the task) operated the apparatus right away across all configurations presented to them. However, the apes did not benefit from the color cues that indicated the location from the traps. Interestingly, the first trial performance shows a similar performance of tool-using apes and apes in the opaque-cued condition. Apes in both of these conditions failed to avoid the traps. In contrast, the apes in the no-tool condition rapidly learned to avoid the traps and outperformed the subjects in the other two conditions. Furthermore, we found an interaction between CiD and tool use compared with the opaque-cued condition. As in Experiment 1, the apes that operated the apparatus only with their fingers (no-tool and opaque-cued conditions) performed worse when a change in direction was required (CiD 1) compared with no change in direction (CiD 0). The difference between the opaque-cued and no-tool conditions was that when causal information was accessible, the apes performed generally better than when only arbitrary color cues were visible. In contrast, for tool-using apes, no effect of CiD was found. With regard to the number of trials per configuration (see the online supplemental materials), we found that the apes in the opaque-cued condition needed more trials than expected by chance and performed significantly worse than those in the clear version of the task (tool and no-tool conditions).

Data on the first trial performance suggests that the information about the causal function of the traps helped the apes to solve this problem. Together, these findings support the notion that great apes encode information about the function of the traps in the current task. Encoding such functional properties of the traps is also in line with previous results by Seed et al. (2009), who found that one of eight chimpanzees spontaneously transferred to two novel conditions, which shared functional but not perceptual cues with the original trap task. Interestingly, this high-performing individual in the study by Seed and colleagues was the female chimpanzee Annett, who, in the current study, was one of only two naïve apes (with regard to this maze task) who learned to avoid the traps in the tool condition (T1 success), and was the only tool-using subject that tended to perform above chance in her first decision in T1. It is important to note that the trap tasks used by Seed and colleagues and the maze task used in the current study differed markedly in their appearance, including the color, shape, and size of the traps. Yet the previously administered tasks and the one used here share structural aspects (i.e., both require to avoid trap elements that block free passage). Thus, it is possible that this chimpanzee transferred the functional knowledge that she had acquired in the previous study to our current task. Therefore, she might have needed less attentional resources to avoid traps, which would explain why she was the only "naïve" ape who was able to plan her moves while using a tool.

Again, the most likely explanation for the interaction between the tool and opaque-cued conditions with regard to CiD is that the tool users were not required to switch their hands to change the direction of the reward, which made it less costly for them to change the direction of the reward. The results of the number of trials per configuration revealed an important difference between the visually restricted condition and tool-use condition. The tool users, in contrast to the apes in the opaque-cued condition, obtained feedback of what happened with the reward when it moved into a trap. This additional piece of information might have helped the tool users to adjust their behavior on a trial-to-trial basis more efficiently than those in the opaque-cued condition.

Experiment 4

In this final experiment, we examined whether apes that had learned to solve the task in the no-tool condition of Experiment 1 could transfer their knowledge to the visually restricted version of the task. We wanted to know whether the apes who had learned to avoid the traps would generalize their knowledge to the arbitrary color cues of the opaque-cued condition. Furthermore, we included a further condition without any obvious visual cues about the trap location to control for the usage of other inadvertently provided cues emanating either from the apparatus, the procedure, or the experimenter.

Method

Subjects. The subjects in this experiment were the same experienced subjects who had learned to avoid the traps in the no-tool condition of Experiment 1 ($n = 7$; see Table 1).

Materials. We used the same apparatuses as in Experiment 3, that is, a clear version of the apparatus for the no-tool condition and an opaque version (with the front side painted black) for the opaque-cued and opaque conditions. Regarding the opaque condition, the only difference to the opaque-cued condition was that we dropped the yellow screens and only used black screens. Thus, in the opaque condition, unlike in the opaque-cued condition, there were no color cues indicating the location of the traps.

Procedure and design. In this experiment, we presented our experienced subjects with all three conditions: the no-tool, opaque-cued, and opaque conditions. We administered one round of 24 trial-unique configurations per subject and condition. The no-tool and opaque-cued conditions were identical to Experiment 3. In the opaque condition, all screens that were occluding subjects' view to the 10 openings in the maze apparatus were black. The actual traps in the opaque-cued and opaque conditions were inserted after the screens were in place and thus outside the subjects' view. The order of conditions was balanced across subjects. There was a maximum of 24 trials or eight configurations (depending on which criterion was reached first) per session.

Scoring and analysis. The scoring and analysis were the same as in Experiment 1.

Results

Success in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: $\chi^2 = 71.66$, $df = 10$, $p < .001$). The GLMM indicated a significant effect of condition (see Table 5): Subjects performed better in the no-tool condition compared with the opaque-cued and opaque conditions. Moreover, subjects performed better in the opaque-cued condition compared with the opaque condition. In addition, we found a significant order effect of the conditions and a differential effect of LoP between the no-tool and opaque conditions. Subjects' performance decreased as a function of the order of presentation, potentially due to motivational issues. With regard to LoP, only in the no-tool condition, we found a significant effect on performance ($z = 2.24$, $p = .025$). For LoP 0 configurations, subjects performed significantly better in the no-tool condition compared with the opaque ($z = 5.48$, $p < .001$) and opaque-cued ($z = 3.60$, $p < .001$) conditions, and subjects performed better in the opaque-cued than in the opaque condition ($z = 2.39$ $p = .017$). Similarly, for LoP 1 configurations, subjects performed significantly better in the no-tool condition compared with the opaque ($z = 3.86$, $p < .001$) and opaque-cued ($z = 3.47$, $p < .001$) conditions. In contrast, subjects did not perform significantly better in the opaque-cued than in the opaque condition ($z = 0.47$, $p = .640$). For LoP 2 configurations, subjects performed significantly better in the no-tool condition compared with the opaque condition ($z = 2.75$, $p = .006$). In contrast, we found no significant difference between the no-tool and opaque-cued conditions ($z = 1.80$, $p = .072$), or between the opaque-cued and opaque conditions ($z = 1.04$, $p = .298$). There was also no significant main effect of age, CiD, or an interaction between CiD and condition (see Figure 8).

Overall, subjects performed better than expected by chance in the no-tool (70.2 ± 5.7% correct; $z = 8.83$, $p < .001$) and opaque-cued (42.3 ± 2.7% correct; $z = 4.87$, $p < .001$) conditions,

Table 5
Experiment 4: Output of GLMMs for the Different Dependent Variables

Model terms	T1 success			T1 first decision		
	Est	95% CI	*p*	Est	95% CI	*p*
No tool vs. Opaque	−1.74	[−2.23, −1.26]	<0.01	−0.89	[−1.37, −0.42]	<0.01
No tool vs. Opaque-cued	−1.23	[−1.69, −0.76]	<0.01	−0.65	[−1.13, −0.17]	0.01
Opaque-cued vs. Opaque	−0.52	[−0.98, −0.06]	0.03	−0.24	[−0.68, 0.20]	0.28
LoP	−0.40	[−0.79, −0.02]	0.04	−0.36	[−0.75, 0.03]	0.07
CiD	−0.11	[−0.50, 0.27]	0.56	−0.19	[−0.58, 0.19]	0.32
Age	−0.08	[−0.26, 0.11]	0.42	0.00	[−0.18, 0.19]	0.97
Order of condition	−0.19	[−0.39, 0.002]	0.047	−0.20	[−0.38, −0.01]	0.04
No tool vs. Opaque × LoP	0.62	[0.13, 1.11]	0.01	0.48	[0.01, 0.96]	0.05
No tool vs. Opaque-cued × LoP	0.37	[−0.10, 0.84]	0.12	0.30	[−0.18, 0.78]	0.22
Opaque-cued vs. Opaque × LoP	0.25	[−0.21, 0.71]	0.28	0.18	[−0.25, 0.62]	0.41
No tool vs. Opaque × CiD	0.21	[−0.27, 0.69]	0.39	0.19	[−0.28, 0.66]	0.42
No tool vs. Opaque-cued × CiD	0.09	[−0.38, 0.55]	0.71	0.27	[−0.20, 0.74]	0.26
Opaque-cued vs. Opaque × CiD	0.12	[−0.34, 0.58]	0.60	−0.08	[−0.52, 0.36]	0.73

Note. GLMM = generalized linear mixed model; T1 = trial 1; CI = confidence interval; LoP = level of planning; CiD = changes in direction.

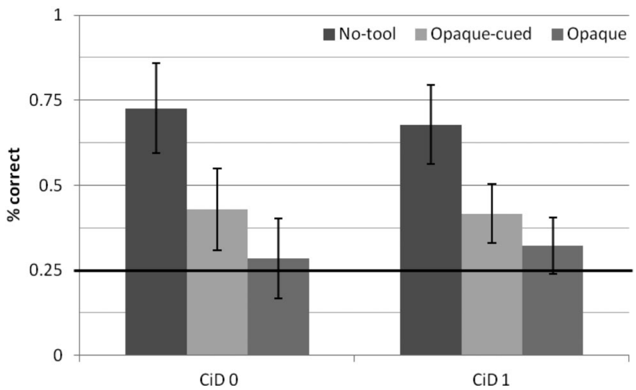

Figure 8. Experiment 4: Trial 1 success (mean ± 95% CI) as function of changes in direction (CiD) and condition.

but not in the opaque condition (30.4 ± 4.2% correct; z = 1.55, p = .120). At the individual level, all seven subjects performed significantly better than expected by chance in the no-tool condition, two out of seven in the opaque-cued condition, and none of the subjects in the opaque condition.

First decision in T1. Overall, the full model was significant compared with the null model (likelihood ratio test: χ^2 = 25.81, df = 10, p = .004). The GLMM indicates a significant effect of condition (see Table 5): Subjects performed better in the no-tool condition compared with the opaque-cued and opaque conditions. In contrast, we found no difference between the opaque-cued and opaque conditions. Additionally, we found again a significant order effect of the conditions and a differential effect of LoP between the no-tool and opaque conditions. Subjects' performance decreased as a function of the order of presentation, potentially due to motivational issues. With regard to LoP, we found a significant effect on performance only in the no-tool condition (z = 2.24, p = .025; see Figure 9). For LoP 0 configurations, subjects performed significantly better in the no-tool condition compared with the opaque condition (z = 3.03, p = .002). In contrast, we found no difference between the no-tool and the opaque-cued conditions (z = 1.64, p = .101), or between the opaque-cued and opaque conditions (z = 1.51, p = .131). In LoP 1 configurations, subjects performed significantly better in the no-tool condition compared with the opaque (z = 2.62, p = .009) and opaque-cued (z = 2.62, p = .009) conditions. In contrast, there was no significant difference between the opaque-cued and opaque conditions (z = 0.03, p = .980). In LoP 2 configurations, subjects did not perform significantly different across conditions (no-tool vs. opaque: z = 0.52, p = .605; no-tool vs. opaque-cued: z = 0.16, p = .872; opaque-cued vs. opaque: z = 0.37, p = .715). There was also no significant main effect of age, CiD, or an interaction between CiD and condition.

Overall, subjects performed better than expected by chance in the no-tool (73.8 ± 4.5% correct; z = 5.09, p < .001) and opaque-cued (60.1 ± 2.4% correct; z = 2.60, p < .001) conditions, but not in the opaque condition (53.6 ± 3.8% correct; z = 0.88, p = .380). At the individual level, four out of seven subjects performed significantly better than expected by chance in the no-tool condition. In contrast, none of the subjects scored significantly above chance in the opaque-cued or opaque conditions.

Discussion

The experienced subjects who had learned to avoid the traps using their fingers were able to transfer to the opaque-cued condition, as indicated by their above-chance performance in the first trial of each configuration and number of trials per configuration (see the online supplemental materials). Thus, they were able to use the yellow color cues to guide their decisions at the current level. However, they were not able to plan their actions ahead based on these arbitrary color cues, as the apes did not perform better than in the opaque condition with regard to the first decision in T1. In the opaque condition, the apes performed as expected at chance in the first trial, showing that inadvertent cues were not at work. Most subjects also performed at chance with regard to the number of trials per configuration, except for two individuals who were able to adjust their behavior flexibly after errors occurred, even in the absence of visual cues. We can rule out that these two subjects were using inadvertent cues, as they were performing at chance in the first trial of each configuration. These two individuals, one juvenile chimpanzee and one orangutan (Kofi and Kila), were also among the best-performing individuals in the clear and opaque-cued condition, suggesting that they were particularly attentive when errors occurred. Previous experience with the opaque-cued condition was not necessary for their post-error adjustments in the absence of visual cues, at least for the juvenile chimpanzee, as he was confronted with the opaque condition before the opaque-cued condition.

Together with the results of Experiment 3, these findings indicate that although color stimuli alone were insufficient for the apes to learn the relevant task contingencies, after some experience with the clear version of the apparatus, the apes could generalize the acquired knowledge about traps to the color cues. The initially presented clear version of the apparatus (Experiment 1) allowed the apes to encode functional information about the yellow traps. Most likely, this functional information was critical for their current transfer of learning. Interestingly, however, these apes failed to use the color information to plan their actions ahead. This suggests that the information that the apes had initially used to plan their actions in the clear version of the apparatus went beyond mere color cues. What is missing in the opaque-cued condition is the causal information about the blockage of the openings. Thus,

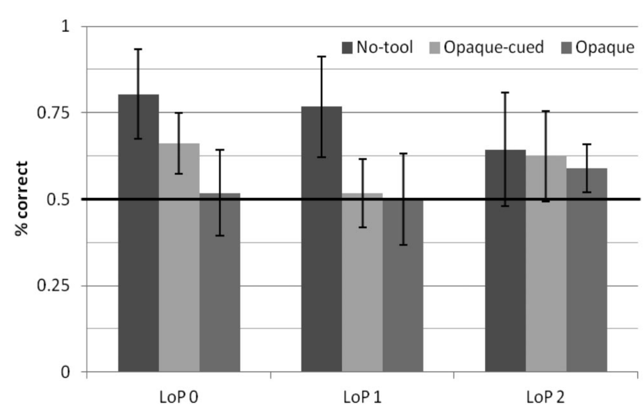

Figure 9. Experiment 4: Trial 1 first decision (mean ± 95% CI) as function of level of planning (LoP) and condition.

this functional information seems to be critical for planning in the present context.

General Discussion

Our study revealed three main findings. First of all, tool use imposes a cognitive load on nonhuman great apes confronted with trap tasks, thus extending previous results (Seed et al., 2009). We found that this cognitive load was particularly pronounced when the apes were confronted with a novel task. In particular, the apes exhibited severe problems to learn about the reward–trap contingency while using the stick as tool to move the reward. This raises the possibility that tool use and the trap task tax a shared cognitive resource.

Second, although experienced apes, unlike naïve ones, adapted well to the requirement of tool use and arbitrary cues, they failed to plan in advance when only arbitrary cues were present. This suggests that apes generalize but need functional information to plan their actions ahead.

Third, when we removed the information on the function of the traps (opaque-cued condition), naïve apes failed to solve the task in the first trial of each configuration above chance levels. This suggests that the successful apes in the no-tool condition used information about functional properties of the traps when making their decisions.

Based on the current findings, the most likely candidate for this shared cognitive process underlying trap tasks and tool use is managing two object–object relations simultaneously (Fragaszy & Cummins-Sebree, 2005). Accordingly, due to capacity limitations, the inclusion of the primary task (tool use) results in a drop in performance in the secondary (trap) task. Therefore, classic trap tasks involving tool use, such as the trap-tube task, can be regarded as dual tasks with two tasks loading on the same cognitive capacity, which might, in turn, explain apes surprisingly poor performance.

We found no evidence that tool use simply exacerbated the task based on the enhanced motor demands of tool use per se. Experienced subjects did not show a drop in performance whatsoever when confronted with the need to use a tool in the context of the trap task for the first time. With regard to the different LoPs, we found just as little evidence for a capacity limitation in the context of tool use. When required to use a tool, naïve apes even failed the condition that did not involve planning (LoP 0), and experienced apes performed similarly with or without the need to use a tool across different levels of planning.

The current findings are also in line with a recent study on great apes' planning abilities using a novel apparatus that required proto-tool use (defined as object-substrate manipulation; Parker & Gibson, 1977). Tecwyn, Thorpe, and Chappell (2013) developed the so-called paddle box paradigm, in which orangutans and bonobos were to rotate multiple paddles (i.e., the substrate) in an appropriate sequence to obtain a food reward (i.e., the object) that was located on top of one of these paddles. The apparatus consisted of eight paddles distributed over three levels. The task was to move the reward to a goal location, from which the reward could be retrieved. The only way to achieve this goal was to rotate the paddles in such a way that the reward would roll toward the target location. The authors found that although the apes were able to sequentially rotate paddles to move the reward, step by step, toward the target location, they failed to predetermine relevant actions and to adjust the orientation of the upcoming paddles before the reward was moved. Crucially, these results confirm the findings obtained in the tool condition of the current study: When the reward could only be moved indirectly by means of a (proto-) tool, the apes failed to predetermine their actions beyond the current step.

Based on neuropsychological evidence, Goldenberg and Spatt (2009) made the distinction between manipulation knowledge and mechanical reasoning in the context of tool use. Manipulation knowledge involves knowledge about how a tool is used ("behavioral routines"), whereas mechanical reasoning deals with how a tool should be applied, given a certain problem-solving situation. In humans, different neural substrates have been identified that are either associated with manipulation knowledge or mechanical reasoning. Using this dichotomy, our results suggest that the limitation of apes tool use in the context of the present trap task is not based on lack or overload of cognitive resources related to the manipulation knowledge, but on the resources underlying apes mechanical reasoning capacities. Next, we discuss the three candidates for the cognitive load of tool use on apes' mechanical reasoning capacities proposed by Seed et al. (2009; i.e., splitting attention, cross-modal matching, and response variability) in the light of the current findings. Two of them received some support, whereas the third did not.

First of all, naïve and experienced apes needed to split or shift their attention between tool, reward, and the three traps in order to make an informed first decision. The failure of the apes in the visually restricted version of the task, despite the cues and response contingencies associated with these cues, suggests that naïve apes encoded functional information about the task. Experienced apes were not required to encode such functional information any longer and could also make their decisions on mere color cues. What made the tool-use task harder for the naïve apes was, thus, learning to avoid the traps based on the functional information about the traps. This initial causal learning to avoid the traps might have been hindered by the need to use a tool. Recent evidence by Kaneko and Tomonaga (2012) indicates that chimpanzees have difficulties in monitoring the movements of external effectors and tend to focus on the target of this effector. In the present context, the requirement to use a tool might have forced the apes to pay more attention to the movement of the reward (via the stick tool) at the cost of the traps. Accordingly, this lack of attentional resources resulted in their failure to avoid the traps in the first trial of each unique configuration.

Second, cross-modal matching might have affected apes as well and contributed, to some extent, to the observed difference between apes in the tool and no-tool conditions. Recall that although tactile information about the traps could not be acquired in either of the conditions, the apes in the no-tool condition got tactile feedback about the continuity and solidity of the surface on which the reward was moved. Thus, displacing the reward with their fingers along the surface might have helped the apes to shift their visual attention away from the reward to plan their next steps (to avoid the traps). In other words, the tactile information in the no-tool condition might have been involved in implicit action monitoring, whereas tool use (by hindering tactile feedback on the location of the reward) might have required explicit action monitoring. Finally, if enhanced response variability was the main limitation of tool use, we would have expected that also experienced subjects that needed to use the tool for the first time were affected by the increased motor demand, at least to some extent. However, this was not the case, suggesting that increased motor demands of tool use per se were not sufficient for explaining the drop in performance in the current trap task.

The negative correlation of apes' performance in the current task with age requires further discussion. With regard to the naïve subjects, this correlation was mainly due to apes in the no-tool condition, to a lesser extent in the tool-using apes, and completely absent for the apes in the opaque-cued condition. Hence, age only affected the performance when information about the functional properties of the traps was available suggesting that this correlation reflects age-related changes in cognitive flexibility rather than changes in associative learning speed. In line with such a notion, we found the age effect again in experienced subjects but now in the tool-use condition, which was novel for them and, thus, cognitively more challenging compared with the previously mastered no-tool condition.

Our findings might also have important implications for other research paradigms, including computerized tasks that require subjects to respond by operating a joystick or trackball as opposed to directly touching the computer screen. Indeed, it has been suggested that the input device matters in a computer-based numerical ordering task, that is, using a joystick might make the actual computer task harder for chimpanzees compared with responding directly with their fingers on a touch-screen (Beran, Pate, Washburn, & Rumbaugh, 2004; Kawai & Matsuzawa, 2000; for divergent results, see Scarf, Danly, Morgan, Colombo, & Terrace, 2011). Finally, we would like to point out that because some of the conclusions of this study are based on cross-experiment comparisons, caution is required (even though these were planned comparisons). Unidentified variables might have contributed, to some extent, to the differences in apes' performance across experiments.

In summary, the current results suggest that great apes' tool use is cognitively demanding and hinders, in some situations, successful problem solving. More specifically, the present study addressed three questions. First, the difficulties imposed by tool use were not specific to planning. Second, we found evidence that tool use was limited by action monitoring of external effectors. Accordingly, tool use disrupted naïve subjects' performance in the trap tasks, as monitoring the tool prevented them from paying attention to the traps. Third, functional information about the traps was essential for naïve subjects to solve this task. Even for experienced subjects, the need to use arbitrary cues hindered their planning abilities. Together, the study pinpoints apes' cognitive limitations to consider multiple object–object relations simultaneously. When required to analyze two object–object relations at the same time, apes cognitive capabilities, such as the attentional system, seem to be somewhat overextended, as indicated by the drop of performance in the tool-use condition. Finally, the current findings support the notion that trap tasks are well suited for examining the causal understanding and planning abilities of nonhuman animals if certain task constraints such as tool use are avoided.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511801686

Bartus, R. T., Fleming, D., & Johnson, H. (1978). Aging in the rhesus monkey: Debilitating effects on short-term memory. *Journal of Gerontology, 33,* 858–871. doi:10.1093/geronj/33.6.858

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–35. 2010. Retrieved from http://CRAN.R-project.org/package=lme4

Beran, M. J., Pate, J. L., Washburn, D. A., & Rumbaugh, D. M. (2004). Sequential responding and planning in chimpanzees (*Pan troglodytes*) and rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes, 30,* 203–212. doi:10.1037/0097-7403.30.3.203

Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 106,* 10370–10375. doi:10.1073/pnas.0901008106

Byrne, R. W., Morgan, D. B., & Sanz, C. M. (2013). Chimpanzees plan their tool use. In C. M. Sanz, J. Call, & C. Boesch (Eds.), *Tool use in animals: Cognition and ecology* (pp. 48–64). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511894800.004

Cox, R. F., & Smitsman, W. (2006). Action planning in young children's tool use. *Developmental Science, 9,* 628–641. doi:10.1111/j.1467-7687.2006.00541.x

Dobson, A. J. (2002). *An introduction to generalized linear models*. Boca Raton, FL: Chapman & Hall/CRC press.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York, NY: Wiley.

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology, 65,* 47–55. doi:10.1007/s00265-010-1038-5

Fragaszy, D. M., & Cummins-Sebree, S. E. (2005). Relational spatial reasoning by a nonhuman: The example of capuchin monkeys. *Behavioral and Cognitive Neuroscience Reviews, 4,* 282–306. doi:10.1177/1534582306286573

Fragaszy, D. M., Johnson-Pynn, J., Hirsh, E., & Brakke, K. (2003). Strategic navigation of two-dimensional alley mazes: Comparing capuchin monkeys and chimpanzees. *Animal Cognition, 6,* 149–160. doi:10.1007/s10071-002-0137-8

Girndt, A., Meier, T., & Call, J. (2008). Task constraints mask great apes' ability to solve the trap-table task. *Journal of Experimental Psychology: Animal Behavior Processes, 34,* 54–62. doi:10.1037/0097-7403.34.1.54

Goldenberg, G., & Spatt, J. (2009). The neural basis of tool use. *Brain: A Journal of Neurology, 132,* 1645–1655. doi:10.1093/brain/awp080

Hansell, M., & Ruxton, G. D. (2008). Setting tool use within the context of animal construction behaviour. *Trends in Ecology & Evolution, 23,* 73–78. doi:10.1016/j.tree.2007.10.006

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3,* 275–310. doi:10.1207/s15516709cog0304_1

Hedden, T., & Gabrieli, J. D. E. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Nature Reviews Neuroscience, 5,* 87–96. doi:10.1038/nrn1323

Jalles-Filho, E., Teixeira Da Cunha, R. G., & Salm, R. A. (2001). Transport of tools and mental representation: Is capuchin monkey tool behaviour a useful model of Plio-Pleistocene hominid technology? *Journal of Human Evolution, 40,* 365–377. doi:10.1006/jhev.2000.0461

Kacelnik, A. (2009). Tools for thought or thoughts for tools? *PNAS Proceedings of the National Academy of Sciences of the United States of America, 106,* 10071–10072. doi:10.1073/pnas.0904735106

Kaneko, T., & Tomonaga, M. (2012). Relative contributions of goal representation and kinematic information to self-monitoring by chimpanzees and humans. *Cognition, 125,* 168–178. doi:10.1016/j.cognition.2012.07.006

Kawai, N., & Matsuzawa, T. (2000). Numerical memory span in a chimpanzee. *Nature, 403,* 39–40. doi:10.1038/47405

Lai, Z. C., Moss, M. B., Killiany, R. J., Rosene, D. L., & Herndon, J. G. (1995). Executive system dysfunction in the aged monkey: Spatial and object reversal learning. *Neurobiology of Aging, 16,* 947–954. doi:10.1016/0197-4580(95)02014-4

Limongelli, L., Boysen, S., & Visalberghi, E. (1995). Comprehension of cause-effect relations in a tool-using task by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 109,* 18–26. doi:10.1037/0735-7036.109.1.18

Martin-Ordas, G., & Call, J. (2009). Assessing generalization within and between trap tasks in the great apes. *International Journal of Comparative Psychology, 22,* 43–60.

Martin-Ordas, G., Call, J., & Colmenares, F. (2008). Tubes, tables and traps: Great apes solve two functionally equivalent trap tasks but show no evidence of transfer across tasks. *Animal Cognition, 11,* 423–430. doi:10.1007/s10071-007-0132-1

Matsuzawa, T. (2001). Primate foundations of human intelligence: A view of tool use in nonhuman primates and fossil hominids. In T. Matsuzawa (Ed.), *Primate origins of human cognition and behavior* (pp. 3–25). Osaka, Japan: Springer. doi:10.1007/978-4-431-09423-4

McCormack, T., Hoerl, C., & Butterfill, S. (Eds.). (2011). *Tool use and causal cognition*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199571154.001.0001

Mulcahy, N. J., & Call, J. (2006). How great apes perform on a modified trap-tube task. *Animal Cognition, 9,* 193–199. doi:10.1007/s10071-006-0019-6

Parker, S. T., & Gibson, K. R. (1977). Object manipulation, tool use and sensorimotor intelligence as feeding adaptations in Cebus monkeys and great apes. *Journal of Human Evolution, 6,* 623–641. doi:10.1016/S0047-2484(77)80135-8

Povinelli, D. J. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works*. New York, NY: Oxford University Press.

Povinelli, D. J., Reaux, J. E., & Frey, S. H. (2010). Chimpanzees' context-dependent tool use provides evidence for separable representations of hand and tool even during active use within peripersonal space. *Neuropsychologia, 48,* 243–247. doi:10.1016/j.neuropsychologia.2009.09.010

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Project for Statistical Computing. Retrieved from http://www.R-project.org

Reaux, J. E., & Povinelli, D. J. (2000). The trap-tube problem. In D. J. Povinelli (Ed.), *Folk physics for apes: A chimpanzee's theory of how the world works* (pp. 108–131). New York, NY: Oxford University Press.

Roche, H., Blumenschine, R. J., & Shea, J. J. (2009). Origins and adaptations of early Homo: What archeology tells us. In F. E. Grine, J. G. Fleagle, & R. E. Leakey (Eds.), *The First Humans: Origin and Early Evolution of the Genus Homo* (pp. 135–147). New York, NY: Springer. doi:10.1007/978-1-4020-9980-9_12

Scarf, D., Danly, E., Morgan, G., Colombo, M., & Terrace, H. S. (2011). Sequential planning in rhesus monkeys (*Macaca mulatta*). *Animal Cognition, 14,* 317–324. doi:10.1007/s10071-010-0365-2

Seed, A., & Byrne, R. (2010). Animal tool-use. *Current Biology, 20,* R1032–R1039. doi:10.1016/j.cub.2010.09.042

Seed, A. M., Call, J., Emery, N. J., & Clayton, N. S. (2009). Chimpanzees solve the trap problem when the confound of tool-use is removed. *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 23–34. doi:10.1037/a0012925

Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal tool behavior: The use and manufacture of tools by animals*. Baltimore, MD: Johns Hopkins University Press.

Silva, F. J., Page, D. M., & Silva, K. M. (2005). Methodological-conceptual problems in the study of chimpanzees' folk physics: How studies with adult humans can help. *Learning & Behavior, 33,* 47–58. doi:10.3758/BF03196049

Stout, D., & Chaminade, T. (2007). The evolutionary neuroscience of tool making. *Neuropsychologia, 45,* 1091–1100. doi:10.1016/j.neuropsychologia.2006.09.014

Tecwyn, E. C., Thorpe, S. K., & Chappell, J. (2013). A novel test of planning ability: Great apes can plan step-by-step but not in advance of action. *Behavioural Processes, 100,* 174–184. doi:10.1016/j.beproc.2013.09.016

Visalberghi, E., & Fragaszy, D. M. (2006). What is challenging about tool use? The capuchin's perspective. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 529–552). New York, NY: Oxford University Press.

Visalberghi, E., & Limongelli, L. (1994). Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology, 108,* 15–22. doi:10.1037/0735-7036.108.1.15

Völter, C. J., & Call, J. (in press). Younger apes and human children plan their moves in a maze task. *Cognition, 130,* 186–203. doi:10.1016/j.cognition.2013.10.007

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120,* 272–292. doi:10.1037/0033-2909.120.2.272

Received November 29, 2013
Revision received February 26, 2014
Accepted February 27, 2014 ■