Azure-winged magpies solve string-pulling tasks by partial understanding of the physical cognition

Lin WANG[#], Yunchao LUO[#], Xin WANG, Abudusaimaiti MAIERDIYALI, Hao CHANG, Zhongqiu LI[*]

School of Life Sciences, Nanjing University, Nanjing, 210023, China

*Address correspondence to Zhongqiu Li. E-mail: lizq@nju.edu.cn.

# Lin Wang and Yunchao Luo contributed equally

Handling editor: Ximena Nelson

Abstract

String-pulling is one of the most widely used paradigms in animal cognition research. We investigated how azure-winged magpies *Cyanopica cyanus* solve multiple-string problems that they have never encountered before. In Experiment 1, the strings were arranged in parallel, slanted, or crossed to investigate what rules azure-winged magpies use to solve multiple spatial relations of strings. Experiment 2 assessed whether the subjects understood the connection between the string and the bait while taking advantage of broken strings. In Experiment 3, the subjects were confronted with strings of different lengths attached to rewards in order to explore whether the string length, as a proxy for the pulling efficiency or reward distance, was crucial for the birds' choice of which string to pull. Generally, the birds were successful in tasks where the reward was close to the correct string's end, and they relied on a "proximity rule" in most cases. The results showed that azure-winged magpies had a partial understanding of the physical principles underlying the string-pulling but were stumped by complex spatial relations. They likely relied on simple strategies such as the proximity rule to solve the tasks. The effects of individual difference and experiential learning on string-pulling performance are also discussed.

Key words: string-pulling; azure-winged magpies; animal cognition; personality

Corvids are especially well known for performing amazing cognitive tasks, and they are extremely inquisitive to explore and manipulate objects with their beak and feet when foraging or playing. These birds are prolific tool users,

not only in the laboratory (Chappell and Kacelnik, 2002, 2004; Taylor et al. 2011; Jelbert et al. 2014), but also in the wild (Hunt 2014). At least 24 species of corvids are skilled in using tools (Lefebvre et al. 2002). Therefore, they are one of the key animals used for exploring animal cognition and intelligence. For example, rooks *Corvus frugilegus*, Eurasian jays *Garrulus glandarius*, and New Caledonian crows *Corvus moneduloides* are widely used to investigate the cognition of animals and showed remarkable performance in studies (Bird and Emery, 2009; Cheke et al. 2011; Jelbert et al. 2014). It seems that corvids possess remarkable cognitive ability in solving novel problems.

String-pulling is an extensively used approach in animal cognition research to evaluate the understanding of spatial and causal relationships . In these experiments, an out-of-reach food item is placed within an animal's field of vision, and the arrangement of strings can be varied in a number of ways to address an array of different psychological questions. The reward is attached to a string, the end of which can be accessed by the subjects. The tested animals can obtain the reward by pulling the correct string. The strings can be arranged either horizontally or vertically. The complexity of string-pulling tasks can be adjusted by manipulating the arrangement of strings, such as their number and formation. As Jacobs and Osvath reported, at least 50 variations of so-called string patterns have been created (Jacobs and Osvath 2015).

Since the first study a century ago (Kinnaman 1902), the string-pulling paradigm has been used to test 163 mammal and bird species in 208 studies (Jacobs and Osvath 2015). String-pulling has been used to test a variety of corvid species, and studies demonstrated that the performance of this species is comparable to that of monkeys and apes (Jacobs and Osvath 2015). Heinrich found that some tested ravens *Corvus corax* could solve tasks with crossed or slanted strings, thereby suggesting that they had an appreciation of the need for connectedness of the string to be a useful tool (Heinrich 1995). A study in 2012 indicated that hooded crows *Corvus corone cornix* were capable of solving slanted-string tasks but failed in crossed strings (Bagotskaya et al. 2012). Heinrich and Bugnyar (2005) found that ravens with prior experience of string-pulling were able to solve a counterintuitive problem, but naïve ravens were not. The naïve ravens' failure to solve a counterintuitive problem where the string had to be pulled down to move the food up led the authors to conclude that the ravens with prior experience did understand the cause – effect relationship between the string, food, and body. However, Taylor and colleagues (Taylor et al. 2010; Taylor et al. 2012) had different opinions. They found that both experienced and naïve New Caledonian crows had weak performance in horizontal and vertical setups when a visual restriction was introduced. Preventing perceptual feedback limits the corvid's understanding of the string problem. Taylor and colleagues also found that when the movement of the reward was prevented by the slack in the string, naïve wild birds stopped interacting with the apparatus. This finding suggested that the birds relied on the visual perception of approaching food as a reinforcement for their actions (Taylor et al. 2012). Therefore, Taylor and

colleagues concluded that planning and complex cognition were not the basis of the successful string-pulling but the "perceptual-motor feedback cycle" (Taylor et al. 2010; Taylor et al. 2012). In 2016, Hofmann et al. (2016) used string-pulling to test western scrub-jays *Aphelocoma californica* and found that the birds were not capable of solving the crossed strings, indicating that these birds did not understand the causal mechanisms underlying the string-pulling tasks.

Much has been written about the cognitive skills required to pull a string, such as, associative learning, trial and error, causal cognition, means-end understanding, imagination and insight (Wasserman et al. 2013). Two hypotheses are used to explain the cognitive processes of birds solving the string-pulling problem: the "perceptual-motor feedback cycle," which was mentioned above, and "insight." The "insight" hypothesis (Heinrich, 1995; Emery and Clayton, 2004; Heinrich and Bugnyar, 2005) suggests that birds mentally model their future actions; they imagine the effect on the position of the food when they repeatedly pull and step on the string, realize what actions will gain them the food, and then execute these actions. Though debates about the cognition behind string-pulling are ongoing, an increasing number of animal species have been tested.

The present study aims to investigate the strategies used by azure-winged magpies *Cyanopica cyanus* when they are confronted with various multiple-string problems. The azure-winged magpie is a social corvid, which can be found in Asia and Iberia. Previous behavioral studies of azure-winged magpies have focused on the reproductive system, such as cooperative breeding and helping behavior (Cruz et al. 2003; Avilés 2004; Horn et al. 2016; Da et al. 2018). However, studies on their cognitive abilities are rare. To assess the spatial and causal performance of azure-winged magpies, three string-pulling experiments were conducted. In Experiment 1, the spatial cognition ability was explored. The birds were tested using three patterned-string problems (Harlow and Settlage 1934; Dücker and Rensch 1977; Hofmann et al. 2016); they had to choose between two strings arranged in different patterns, of which only one was connected to the reward. In Experiment 2, two broken string problems were presented to the subjects to explore whether or not the birds understood string continuity. In Experiment 3, the subjects were required to distinguish between two strings of different lengths, both of which were connected to a reward. Thus, it was possible to observe cues that birds use to solve the task by manipulating the two strings with different lengths.

Materials and Methods

Subjects

Six azure-winged magpies (named Marie, Daniel, Emily, Fatty, Joyce, and Tiny) served as subjects throughout the study. At the beginning of the study, the hand-raised birds were juvenile (aged 4 months old) and kept at Nanjing

University. They were single-housed in home cages (60 cm × 40 cm × 40 cm) and maintained on a regular diet. During the experiments, the subjects had access to water but not food. None of the subjects had participated in other studies, and they had no experience of string-pulling.

Experimental design

Red strings of different lengths were arranged in different spatial relations (Figure 1, 2). In Training 1, the birds were presented with a single mid-length string (14 cm) perpendicular to the edge of the acrylic box. In Training 2, the birds were given two parallel strings (14 cm), where only one was baited. In Experiment 1, two strings (14 cm) were arranged in parallel, slanted, or crossed to explore the spatial cognition ability of the birds. Experiment 2 assessed whether the subjects understood the connection between the string and the bait; therefore, long (14 cm) and short strings (6 cm) were arranged into broken strings. In Experiment 3, the distance of the reward attached to a 24 cm-long string was changed in relation to the reward attached to a 14 cm-long string.

All tasks involved a transparent acrylic box (20 cm × 10 cm × 2 cm), consisting of six transparent acrylic plates, where the top plate can be removed. There were four holes on the smallest side so that the string ends could protrude from the box approximately 2–3 cm. Different holes were chosen in different tasks. The strings were arranged horizontally in different spatial relations in the transparent box. This setup allowed the strings to be seen but not touched. The rewards were mealworms tied to red strings (0.8 mm in diameter). To prevent mealworms from wriggling, they were drowned in water. In previous studies, birds have generally been tested on vertical strings (Jacobs and Osvath, 2015). A horizontal string can be reeled in with a single pull, whereas a vertical string requires more complex coordination, such as reaching down, grasping and pulling, creating a loop, and repetition of these processes (Werdenich and Huber, 2006; Ellison et al. 2015). Compared with a vertical string, the subject has a better view of the arrangement in horizontal configuration and can more easily stay focused on the reward. To enhance the motivation of the tested birds, horizontal strings were used in this study.

Experimental procedures

The training and experiments were conducted between July and November 2017. In the experiments, the subjects were isolated so that they were in auditory but not in visual contact during testing. After 3 h of food deprivation, the bird was moved into a 40 cm × 40 cm cage to be tested individually. Marie, Emily, and Daniel were sensitive to the changed environment; thus, they were tested in their own living cage (same size as the testing cage).

In the testing cage, a simple plastic underlay sheet was placed and the acrylic box was presented on the underlay.

4

Thereafter, the birds were tested with multiple-string problems (Figure 2). The baited side was random and no one side was baited more than twice in succession. Each trial ended when a string (either rewarded or unrewarded) was removed from the box. The strings will be rearranged at the beginning of the next trial.

To ensure that the subjects remained testable during the course of the experiment, the experimental session lasted a maximum of 30 min. If the subject lost interest and was unwilling to approach the apparatus for more than 3 min, the experimental session of the subject was ended.

Before the training phase, the acrylic box was placed in the cage until the birds approached the apparatus reliably. Initially, the mealworms were placed at the edge of the acrylic box so that the birds could obtain the reward directly without pulling the strings. In Training 1, the birds were presented with a single mid-length string tied to a reward. In Training 2, two strings, the same ones used in Training 1, were presented to the birds concurrently. One of them was baited with a mealworm, and the other was empty. The two strings were arranged in parallel and perpendicular to the edge of the acrylic box. To avoid developing a side bias, the baited side was alternated randomly across trials, but no side was baited more than twice in a row. A training session lasted a maximum of 30 min. In addition, one tested bird was trained no more than 1 h a day. When a bird achieved five consecutive successful trials, we considered that the subject passed the training phase.

Three common patterned-string problems documented by Harlow and Settlage (Harlow and Settlage, 1934) were used in Experiment 1: parallel, slanted, and crossed strings. There were two levels for the slanted-string task. In Experiment 1b, the reward was close to the connected string end, whereas in Experiment 1c, it was close to the unconnected string end. To avoid developing a side bias, the subject was presented with strings slanted to the left or slanted to the right in Experiments 1B and 1c. Experiment 1d was a crossed-string problem where only one string was connected to the reward. Each task was repeated 50 times so that each individual was tested for a total of 200 trials. One of the two strings (14 cm) was tied with a reward, and once a string was removed from the box, the trial ended. The baited string was alternated randomly across trials, and no side was rewarded more than twice in a row. To prevent the formation of response habits, the four tasks in Experiment 1 and the three tasks in Experiment 3 were conducted in a randomly interleaved sequence. In addition, the intervals between the two tasks were more than 36 h. Studies showed that the subjects succeeded in patterned-string problems when they were given a large number of repetitions of the same condition but obtained contrary results in an intermixed design (King and Witt, 1966; Cha and King, 1969; Dücker and Rensch, 1977).

Experiment 2 consisted of two broken string problems to explore whether the subjects understood the connection between the string and the bait. In Experiment 2a, one mid-length (14 cm) and two shorter strings (6 cm) were arranged

as broken strings with an interval of 2 cm. In Experiment 2b, one mid-length string (14 cm) with a 90° angle and one shorter string (6 cm) were arranged as broken strings with an interval of 2 cm. The string with a 90° angle was tied to a reward, creating the illusion that the baited string was the shorter one. Each task was repeated 50 times so that each individual was tested for a total of 100 trials. When one of the two strings was removed, the trial ended. The strings were rearranged at the beginning of the next trial. The baited string was alternated randomly across trials, and no side was rewarded more than twice in a row.

Experiment 3 consisted of three different string problems in which the birds were presented with two strings of different lengths. Each of the three tasks used the same strings (21 and 14 cm long). The distance between the reward attached to those strings and the subjects varied among the three tasks. The objective of this experimental design was to determine whether the string length, as a proxy for the pulling efficiency or reward distance, was crucial for the birds' choice of which string to pull. Each task was repeated 50 times so that each individual was tested in 150 trials. The sides for the long and mid-length strings varied randomly. In Experiment 3a, both strings lay straight which increased the distance of the reward by 17 cm on the long side and 10 cm on the short side. In Experiment 3b, the rewards were both placed at a distance of 10 cm from the edge of the box, and the excess of the long string was slack. In Experiment 3c, the reward attached to the long string was placed approximately 3 cm from the edge of the box, and the excess of the long string was slack. The short string was arranged in a straight line in all three tasks.

Data analysis

The time and results of the trials were recorded. For the analysis, only the initial choice was considered because only one pull was needed to obtain the reward in almost all cases. As all tasks were discrimination tasks with two different possibilities, a binomial test was applied to compare the birds' performance to chance. The number of correct responses out of the total number of trials was tested against a chance level of 50%. A chi-square test was used for analyzing the individual difference and experiential learning.

Results

Training

All tested birds except Emily passed Training 1 after three training sessions (90 min in total). Emily was unwilling to approach the acrylic box for the first few days, but she finally passed the first training phase after eight sessions of training (240 min in total). The numbers of trials in Training 2 that subjects received before reaching the criterion of

five successful trials in a row were different, and two of six birds had success rates that exceeded 50% (see Table S1). All birds successfully passed five consecutive trials in the second training phase.

In the training phase, the subjects did not develop side bias, and they remained testable. As a result all subjects were tested in the following experiments.

Experiment 1: common patterned-string problems

Experiment 1A was successfully solved by three out of the six birds with Marie and Tiny performing significantly above chance level ($t = 0.5$, $P < 0.001$). Experiment 1b, wherein the reward was closer to the end of the rewarded string, was successfully performed by three birds ($t = 0.5$, $P < 0.001$). In Experiment 1c, wherein the reward was opposite to the end of the unrewarded string, four of the six subjects chose the strings randomly. Daniel solved a more difficult version successfully ($t = 0.5$, $P = 0.007$), but Marie failed significantly above chance level ($t = 0.5$, $P = 0.003$). No subject was able to solve the cross string problem (Experiment 1d), and Tiny typically chose bare strings ($t = 0.5$, $P < 0.001$) (Table 1).

Experiment 2: broken string problems

Three out of the six birds solved the broken string problems (Experiment 2A) with Marie performing significantly high above chance level ($t = 0.5$, $P < 0.001$). Experiment 2B was successfully performed by four birds (Table 2).

Experiment 3: the choice of long and short strings

In Experiment 3A, three birds showed a significant preference for the shorter strings (14 cm). In Experiment 3b, two birds selected the shorter string ($t = 0.5$, $P < 0.001$). In Experiment 3c, four out of the six subjects had significant preferences. Marie and Fatty preferred the shorter strings, whereas Tiny and Joyce preferred the longer strings (21 cm) (Table 3).

Individual difference

In our study, the difference between individuals was significant. Thus, the success rate of 200 trials in Experiment 1 and 100 trials in Experiment 2 were calculated respectively and compared among the subjects via the chi-square tests. The performance of the birds in Experiment 1 was not above chance level ($\chi^2 = 5.000$, $df = 5$, $P = 0.416$), and the correct

numbers of the birds in Experiment 1 were: 116 (Marie), 117 (Daniel), 114 (Emily), 99 (Fatty), 105 (Tiny) and 111 (Joyce). The performance of the birds in Experiment 2 varied significantly above chance level ($\chi^2 = 50.089$, $df = 5$, $P < 0.001$), indicating an obvious individual difference (see Figure 3 and Table S2). The choices of the birds in Experiment 3c varied significantly above chance level ($\chi^2 = 42.514$, $df = 5$, $P < 0.001$); two of six subjects preferred short distance (the long string) and two of six subjects preferred short strings (see Figure 4). Besides, the performance of the birds in training phase also indicated an obvious individual difference.

Experiential learning

To demonstrate whether there is learning behavior, 50 trials were divided into two parts to determine if the birds have better performance in the second half. Of the total 54 samples (9 tasks of each subject), 28 improved, 6 remained constant, and 20 declined in the second half (see Table 4). The chi-square test showed no effects of experiential learning on solving the string-pulling tasks ($\chi^2 = 8.629$, $df = 10$, $P = 0.568$). However, there was still an obvious individual difference. For example, as shown in Figure 5, Daniel improved significantly in the second half of trials compared with the first half (Experiment 1d: $\chi^2 = 9.680$, $df = 1$, $P = 0.01$; Experiment 2b: $\chi^2 = 9.191$, $df = 1$, $P = 0.002$; Experiment 3b: $\chi^2 = 6.640$, $df = 1$, $P = 0.01$). The first 10 and last 10 trials of each task were analyzed, and the results are shown in the Supplemental Materials (Tables S3 and S4).

Discussions

The results of Experiment 1 showed that the subjects were capable of solving several easy patterned-string tasks but failed in other difficult tasks. Compared with Experiment 1a, five birds improved their accuracy in Experiment 1b, with the exception of Tiny. One possibility was that increased experience with string-pulling could have resulted in improved performance. What was surprising was that Marie did well in Experiments 1A and 1B but failed in Experiment 1C. This finding might be considered as evidence that the subjects chose according to the proximity rule and they could not understand the causal connection between the string and the reward. In the same way, Tiny showed a significant preference for the bare string. Surprisingly, Daniel and Emily did not perform significantly above chance level in easy task but were successful in more complicated tasks. Other studies also found that some birds succeeded in difficult tasks but failed in easy tasks. For example, Hofmann et al. (2016) tested a western scrub-jay, which failed in parallel but succeeded in slant strings. Bagotskaya et al. (2012) tested a hooded crow, which failed in a single string task but successfully coped with multiple-string problems. One reason for this finding could be due to the increased exposure of

8

the subjects to the string-pulling experiment. The outcome of Experiment 1 revealed that the birds prefer to choose according to the proximity rule. In addition, the birds were not able to understand the logical structure and the underlying physical principles of string-pulling. The results were similar to those observed in western scrub-jays (Hofmann et al. 2016).

The result of Experiment 2 was unexpected and suggested that some birds could solve broken string tasks based on their ability to follow the connection between the string and the bait and assess the value of the string to a certain extent. However, no bird was able to solve the crossed-string task in Experiment 1. Moreover, the result of the crossed-string task suggested that the subject could not understand the causal connection between the string and the reward, which seemed inconsistent with the result of Experiment 1C. We infer that the subjects can realize the necessity of string continuity but are confused with the complex spatial relations in Experiment 1c. It is surprising that the azure-winged magpies perform very well in broken-string tasks. According to published studies, great apes (*Gorilla*) can solve a similar task with food resting on a piece of cloth (Herrmann et al. 2008). Pigeons *Columba livia* cannot solve the broken-string task (Schmidt and Cook 2006). Some spectacled parrotlets *Forpus conspicillatus* can solve the broken-string task spontaneously, whereas lorikeets *Vini kuhlii*, macaws (*Psittacidae*), and cockatoos (*Cacatuidae*) cannot (Krasheninnikova and Wanker 2010).

The result of Experiment 3A showed that some subjects preferred the proximal reward and others chose randomly, indicating that some birds chose on a "proximity rule." In Experiment 3B, a minority preferred the short strings. When the distance of both strings was approximately 10 cm, the birds had difficulty in distinguishing the two strings. What was interesting was that the results of Experiment 3C were contradictory: two birds preferred longer strings and two the shorter strings. A possible explanation is that the birds that prefer the longer strings are strongly motivated to obtain the reward closer to them irrespective of the effort necessary to do so. The distance of food has an effect in mammals and birds and can be more influential than the functionality of the tool (Fujita et al. 2003).

Individual differences are common in studies that explore the cognitive ability of animals. The European magpie *Pica pica* is the only bird that has passed the mirror test and possesses self-awareness (Prior et al. 2008). However, only two of five European magpies passed the test completely. This result provided evidence for the existence of individual differences. Similarly, research in 2014 investigated the causal understanding of water displacement of New Caledonian crows, and only two of six subjects completed all the experiments (Jelbert et al. 2014). Jacobs and Osvath (2015) discussed several factors that might influence or confound the performance of birds in string-pulling test, such as age and personality. Age is not a key factor in our study because all birds are juveniles, with less than 2 weeks' age gap. However, personality may play a part in solving the tasks. To illustrate the influence factors of individual difference,

further studies are needed.

Increased experience with string-pulling could have resulted in improved performance. However, individual difference still exists. For example, Daniel and Tiny improved significantly in six of nine tests, but Marie declined significantly in five of nine tests. The results indicated that the string-pulling performance was complicated and the explanation, with experiential learning, might be simplistic. Many studies illustrate that improvement in string-pulling over a number of sessions shows the positive effects of learning and experience, with a corresponding decrease in exploratory behavior, solution time, dropping errors, and incorrect choices (Jacobs and Osvath, 2015). However, previous experience does not always have a clear directional effect; cats show poor performance in perpendicular conditions with increased experience (Whitt et al. 2009). The experience of rhesus macaques did not have any strong effect on subsequent perpendicular performance but adversely affected the performance under the crossed-string conditions (Mason and Harlow, 1961). In a recent study of western scrub-jays, some individuals improved, whereas others declined, similar to our findings (Hofmann et al. 2016). To advance the understanding of the role of learning and experience in solving string-pulling problem, researchers should shift their focus to the kind of experience the animals use in performing the task.

The objective of this study is to explore the physical cognition of azure-winged magpies by presenting them with a selection of horizontal string-pulling tasks in three experiments. The results of Experiment 1 did not provide abundant evidence that the azure-winged magpies understood the causal relationship underlying the string-pulling problem and the functionality of a string. Instead, they appeared to rely on a "proximity rule," that is, pulling the string end nearest to the reward. Meanwhile, performance of the birds in Experiment 1A, 1B and 1C could be explained by associative learning. Subjects carried over knowledge from training phase to pull on a string closed to the worm, however, subjects failed as soon as strings were crossed over and the associative rule no longer applied. The results of Experiment 2 demonstrated that some birds could realize the connection between the string and the bait. Furthermore, Experiment 3 illustrated that birds relied on different strategies to select strings. The hypothesis of Taylor and colleagues claim that the performance of many species of corvids can be explained by a "perceptual motor feedback loop." That is to say, food moving toward the subject is regarded as an internal psychological reinforcer that motivates the subject to pull the strings (Taylor et al. 2010; Taylor et al. 2012). In this study, subjects learn over multiple trials which behaviors bring food closer, and performing that action preferentially over food items that are closer to them, than those further away. The proximity of the food is therefore a vital part of the feedback loop mechanism. In this study, the longest string is 21 cm long. This means that the subjects only need to pull once or twice for the string to come out from the box. Meanwhile, the number of switches (i.e., subjects choose one string and then switch to another)

is rare. That is to say, the first strings the subjects touched are same as the first strings they pull out of the apparatus. Thus, trial-and-error mechanism is not discussed in this study.

Overall, the string-pulling performance of azure-winged magpies is similar to that of other corvids and parrots (Jacobs and Osvath 2015), even though they behave better in broken-string tests. Few animals have a complete comprehension of the functionality of the strings except ravens and some primates (Mayer et al. 2014). Other species, including squirrel monkeys *Saimiri sciureus* (Cha and King 1969), common marmosets *Callithrix jacchus* (Gagne et al. 2012), rhesus monkeys *Macaca mulatta* (Mason et al. 1956), parrots (*Psittaciformes*) (Schuck-Paim et al. 2009), and other corvids (Taylor et al. 2010; Bagotskaya et al. 2012; Hofmann et al. 2016) as well as our azure-winged magpies, mostly could not completely understand the causal relationship underlying the string-pulling problem.

Our results suggest that azure-winged magpies may not fully understand the causal mechanisms underlying string-pulling tasks, and they generally rely on proximity rule to solve the tasks. The performance of the birds in string-pulling tasks suggests that they have a partial understanding of the physical cognition. New Caledonian crows are well known for their remarkable physical cognition, but they still struggle with some string-pulling tasks. Hence, further investigations are necessary to explore the cognition of azure-winged magpies in different paradigms. For instance, testing the azure-winged magpies using other benchmark tests of physical cognition would be interesting, such as the two-trap trap-tube test (Seed et al. 2006; Taylor et al. 2009), Aesop's fable water task (Bird and Emery, 2009; Cheke et al. 2011; Cheke et al. 2012; Jelbert et al. 2014), object permanence task (Zucca et al. 2005; Zucca et al. 2007; Hoffmann et al. 2011; Ujfalussy et al. 2013), or mirror test (Reiss and Marino, 2001; Plotnik et al. 2006; Medina et al. 2011; Soler et al. 2014; Ma et al. 2015), so that we may obtain a sufficient understanding of their physical cognitive abilities.

Ethics statement

All applicable international, national, and institutional guidelines for the care and use of animals were followed. Under Chinese law, no specific approval was required for this noninvasive study. In addition, this study does not involve human participants.

References

Avilés JM, 2004. Egg rejection by iberian azure-winged magpies *cyanopica cyanus* in the absence of brood parasitism. J. Avian Biol. 35(4):295–299.

Bagotskaya MS, Smirnova AA, Zorina ZA, 2012. Corvidae can understand logical structure in baited string-pulling tasks. Neurosci. Behav. Physiol. 42(1):36–42.

Bird CD, Emery NJ, 2009. Rooks use stones to raise the water level to reach a floating worm. Curr. Biol. 19(16):1410–1414.

Cha JH, King JE, 1969. The learning of patterned strings problems by squirrel monkeys. Anim. Behav. 17(69):64–67.

Chappell J, Kacelnik A, 2002. Tool selectivity in a non-primate, the new caledonian crow *Corvus moneduloides*. Anim. Cogn. 5(2):71–78.

Chappell J, Kacelnik A, 2004. Selection of tool diameter by new caledonian crows *corvus moneduloides*. Anim. Cogn. 7(2):121–127.

Cheke LG, Bird CD, Clayton NS, 2011. Tool-use and instrumental learning in the eurasian jay *Garrulus glandarius*. Anim. Cogn. 14(3):441–455.

Cheke LG, Loissel E, Clayton NS, 2012. How do children solve aesop's fable? PLoS ONE 7(7):e40574.

Cruz CDL, Solís E, Valencia J, Chastel O, Sorci G, 2003. Testosterone and helping behavior in the azure-winged magpie *Cyanopica cyanus*: natural covariation and an experimental test. Behav. Ecol. Sociobiol. 55(1):103–111.

Dücker G, Rensch B, 1977. The solution of patterned string problems by birds. Behaviour 62(1/2):164–173.

Da XW, Xian LL, Luo JJ, Gao LF, Du B, 2018. Azure winged magpies *cyanopica cyanus* trade off reproductive success and parental care by establishing a size hierarchy among nestlings. Ibis.

Ellison AM, Watson J, Demers E, 2015. Testing problem solving in turkey vultures *Cathartes aura* using the string-pulling test. Anim. Cogn. 18(1):111–118.

Emery NJ, Clayton NS, 2004. The mentality of crows: Convergent evolution of intelligence in corvids and apes. Science 306(5703):1903–1907.

Fujita K, Kuroshima H, Asai S, 2003. How do tufted capuchin monkeys *Cebus apella* understand causality involved in tool use? J. Exp. Psychol. Anim. B. 29(3):233–242.

Gagne M, Levesque K, Nutile L, Locurto C, 2012. Performance on patterned string problems by common marmosets *Callithrix jacchus*. Anim. Cogn. 15(5):1021–1030.

Harlow HF, Settlage PH, 1934. Comparative behavior of primates. Vii. Capacity of monkeys to solve patterned string tests. J. Comp. Psychol. 18(3):423–435.

Heinrich B, 1995. An experimental investigation of insight in common ravens *Corvus corax*. Auk 112(4):994–1003.

Heinrich B, Bugnyar T, 2005. Testing problem solving in ravens: string-pulling to reach food. Ethology 111(10):962–976.

Herrmann E, Wobber V, Call J, 2008. Great apes' (*Pan troglodytes, Pan paniscus, Gorilla gorilla, Pongo pygmaeus*) understanding of tool functional properties after limited experience. J. Comp. Psychol. 122(2):220–230.

Hoffmann A, Rüttler V, Nieder A, 2011. Ontogeny of object permanence and object tracking in the carrion crow *corvus corone*. Anim. Behav. 82(2):359–367.

Hofmann MM, Cheke LG, Clayton NS, 2016. Western scrub-jays *Aphelocoma californica* solve multiple-string problems by the spatial relation of string and reward. Anim. Cogn. 19(6):1103–1114.

Horn L, Scheer C, Bugnyar T, Jjm M, 2016. Proactive prosociality in a cooperatively breeding corvid, the azure-winged magpie *Cyanopica cyana*. Biol. Letters 12(10):20160649.

Hunt GR, 2014. New caledonian crows' *Corvus moneduloides* pandanus tool designs: Diversification or independent invention? Wilson J. Ornithol. 126(1):133–139.

Jacobs IF, Osvath M, 2015. The string-pulling paradigm in comparative psychology. J. Comp. Psychol. 129(2):89–120.

Jelbert SA, Taylor AH, Cheke LG, Clayton NS, Gray RD, 2014. Using the aesop's fable paradigm to investigate causal understanding of water displacement by new caledonian crows. PLoS One 9(3):e92895.

King JE, Witt ED, 1966. The learning of patterned strings problems by rock squirrels. Psychonomic Sci. 4(9):319–320.

Kinnaman AJ, 1902. Mental life of two macacus rhesus monkeys in captivity. Am. J. Psychol. 13(2):173–218.

Krasheninnikova A, Wanker R, 2010. String-pulling in spectacled parrotlets *Forpus conspicillatus*. Behavior 147(5):725–739.

Lefebvre L, Nicolakakis N, Boire D, 2002. Tools and brains in birds. Behavior 139(7):939–973.

Ma X, Jin Y, Luo B, Zhang G, Wei R et al.,2015. Giant pandas failed to show mirror self-recognition. Anim. Cogn. 18(3):713–721.

Mason WA, Blazek NC, Harlow HF, 1956. Learning capacities of the infant rhesus monkey. J. Comp. Physiol. Psychol. 49(5):449–453.

Mason WA, Harlow HF, 1961. The effects of age and previous training on patterned-strings performance of rhesus monkeys. J. Comp. Physiol. Psychol. 54(6):704–709.

Mayer C, Call J, Albiach-Serrano A, Visalberghi E, Sabbatini G et al., 2014. Abstract knowledge in the broken-string problem: evidence from nonhuman primates and pre-schoolers. PLoS ONE 9(10):e108597.

Medina FS, Taylor AH, Hunt GR, Gray RD, 2011. New caledonian crows' responses to mirrors. Anim. Behav.

82(5):981–993.

Plotnik JM, de Waal FB, Reiss D, 2006. Self-recognition in an asian elephant. P. Natl. Acad. Sci. USA 103(45):17053–17057.

Prior H, Schwarz A, Güntürkün O, 2008. Mirror-induced behavior in the magpie *Pica pica*: evidence of self-recognition. PLoS Biol. 6(8):e202.

Reiss D, Marino L, 2001. Mirror self-recognition in the bottlenose dolphin: a case of cognitive convergence. P. Natl. Acad. Sci. USA 98(10):5937–5942.

Schmidt GF, Cook RG, 2006. Mind the gap: means-end discrimination by pigeons. Anim. Behav. 71(3):599–608.

Schuck-Paim C, Borsari A, Ottoni EB, 2009. Means to an end: Neotropical parrots manage to pull strings to meet their goals. Anim. Cogn. 12(2):287–301.

Seed AM, Tebbich S, Emery NJ, Clayton NS, 2006. Investigating physical cognition in rooks, *corvus frugilegus*. Curr. Biol. 16(7):697–701.

Soler M, Perez-Contreras T, Peralta-Sanchez JM, 2014. Mirror-mark tests performed on jackdaws reveal potential methodological problems in the use of stickers in avian mark-test studies. PLoS One 9(1):e86193.

Taylor AH, Elliffe DM, Hunt GR, Emery NJ, Clayton NS et al., 2011. New caledonian crows learn the functional properties of novel tool types. PLoS ONE 6(12):e26887.

Taylor AH, Hunt GR, Medina FS, Gray RD, 2009. Do new caledonian crows solve physical problems through causal reasoning? Proc. Biol. Sci. 276(1655):247–254.

Taylor AH, Knaebe B, Gray RD, 2012. An end to insight? New caledonian crows can spontaneously solve problems without planning their actions. Proc. Biol. Sci. 279(1749):4977–4981.

Taylor AH, Medina FS, Holzhaider JC, Hearne LJ, Hunt GR et al., 2010. An investigation into the cognition behind spontaneous string pulling in new caledonian crows. PLoS ONE 5(2):e9345.

Ujfalussy DJ, Miklosi A, Bugnyar T, 2013. Ontogeny of object permanence in a non-storing corvid species, the jackdaw *Corvus monedula*. Anim. Cogn. 16(3):405–416.

Wasserman E, Nagasaka Y, Castro L, Brzykcy S, 2013. Pigeons learn virtual patterned-string problems in a computerized touch screen environment. Anim. Cogn. 16(5):737–753.

Werdenich D, Huber L, 2006. A case of quick problem solving in birds: String pulling in keas *nestor notabilis*. Anim. Behav. 71(4):855–863.

Whitt E, Douglas M, Osthaus B, Hocking I, 2009. Domestic cats *Felis catus* do not show causal understanding in a

string-pulling task. Anim. Cogn. 12(5):739.

Zucca P, Antonelli F, Vallortigara G, 2005. Detour behaviour in three species of birds: Quails (*coturnix sp.*), herring gulls *Larus cachinnans* and canaries *Serinus canaria*. Anim. Cogn. 8(2):122–128.

Zucca P, Milos N, Vallortigara G, 2007. Piagetian object permanence and its development in eurasian jays *Garrulus glandarius*. Anim. Cogn. 10(2):243–258.
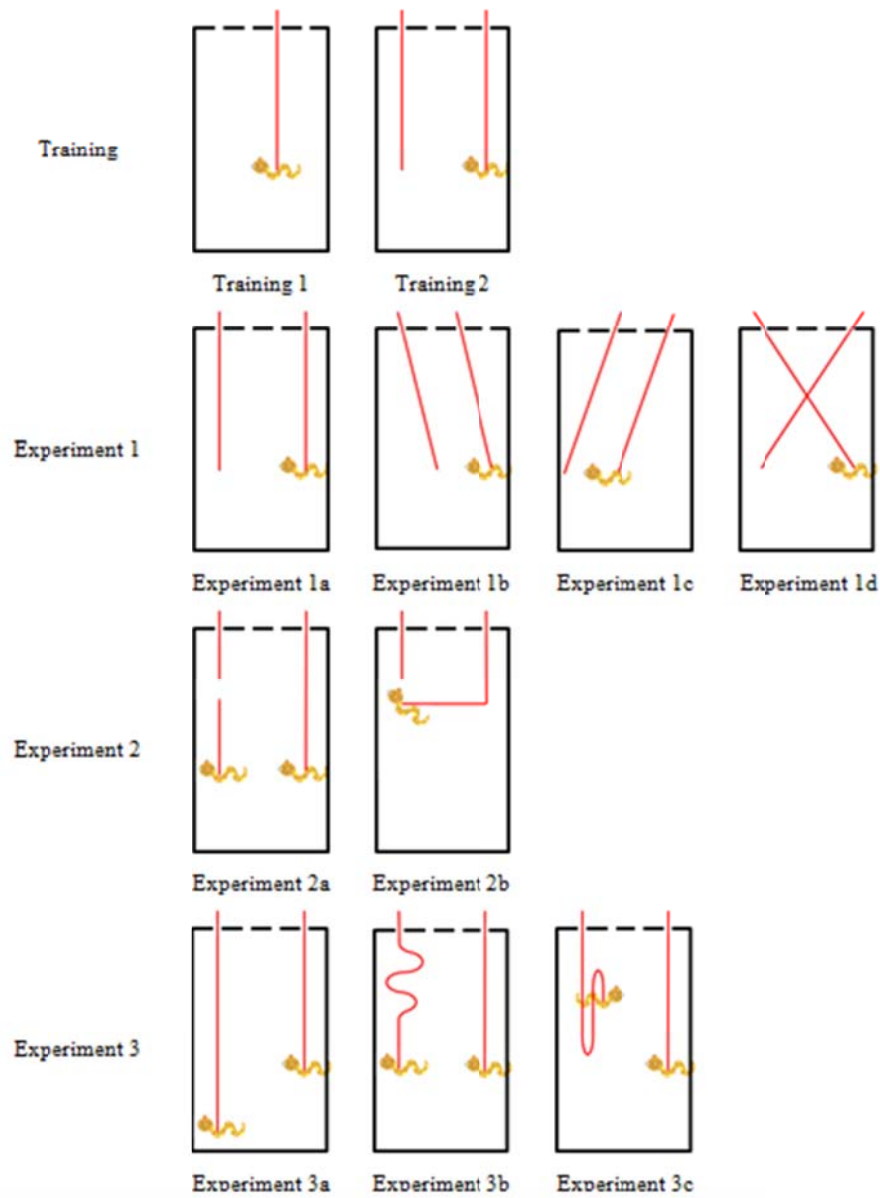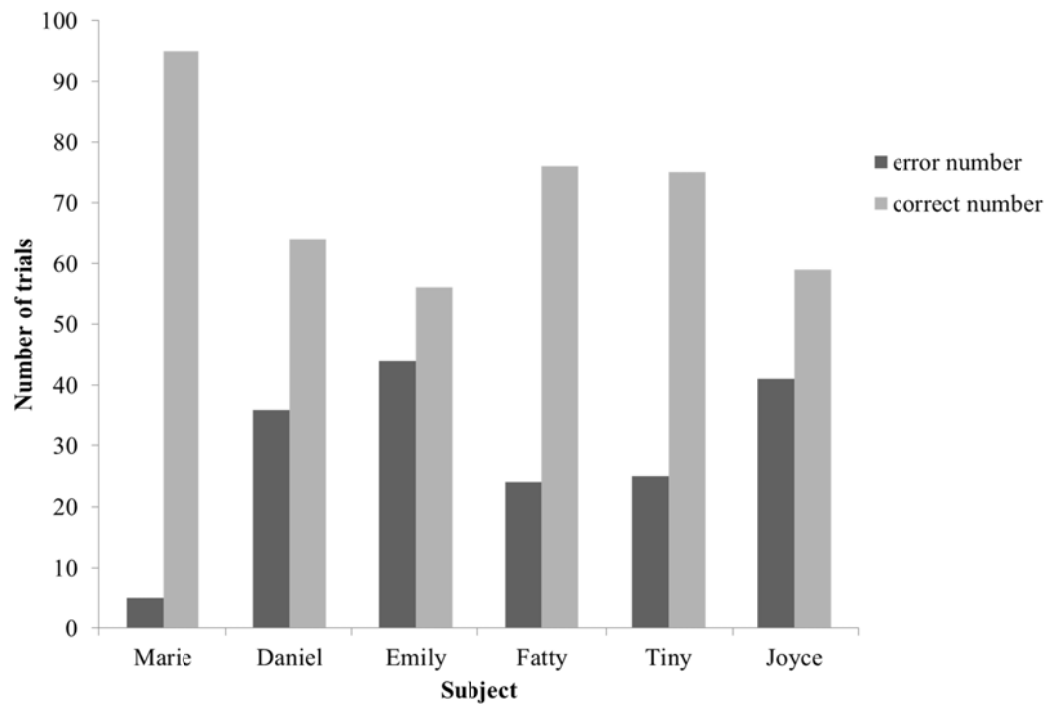
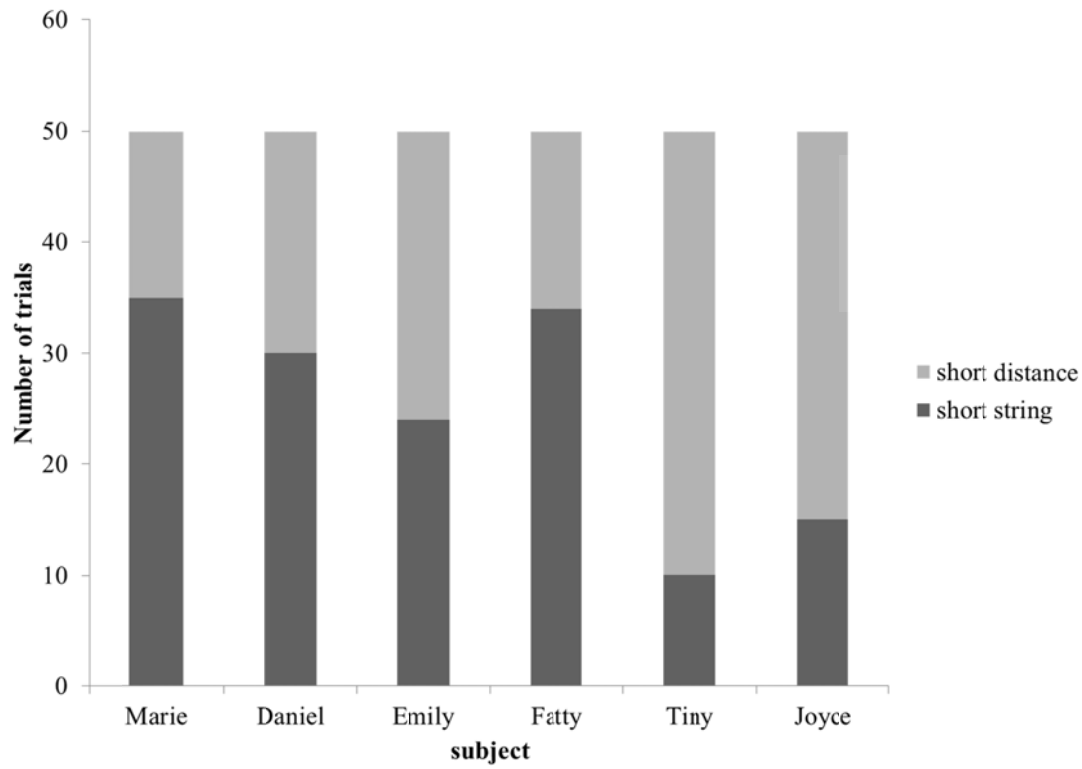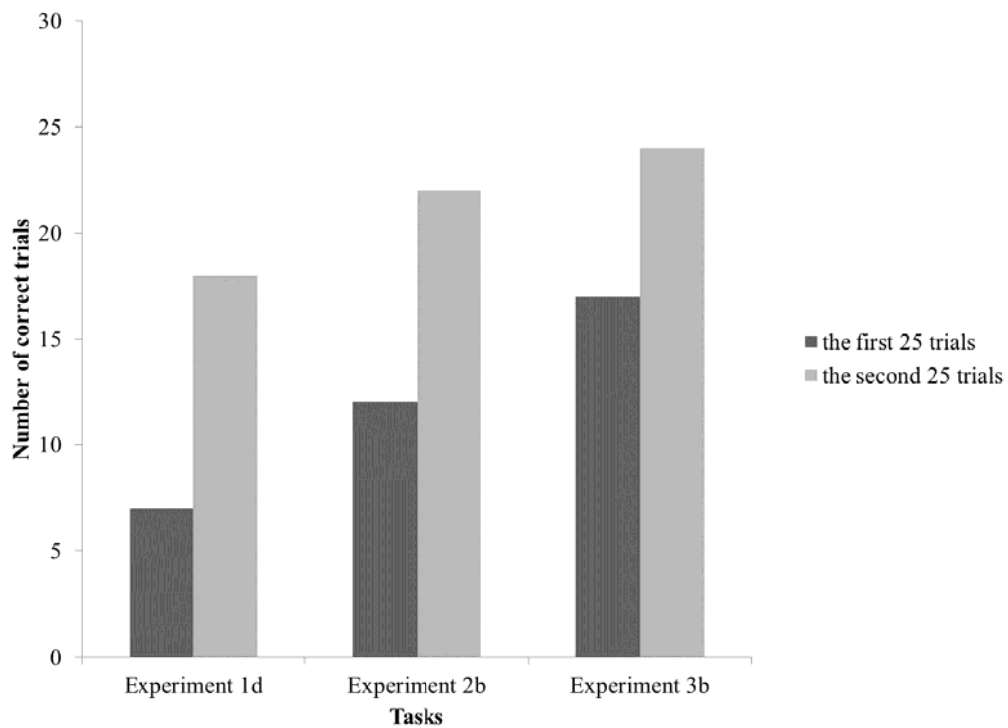**Figure 1.** String arrangement of Training, Experiment 1, Experiment 2, and Experiment 3
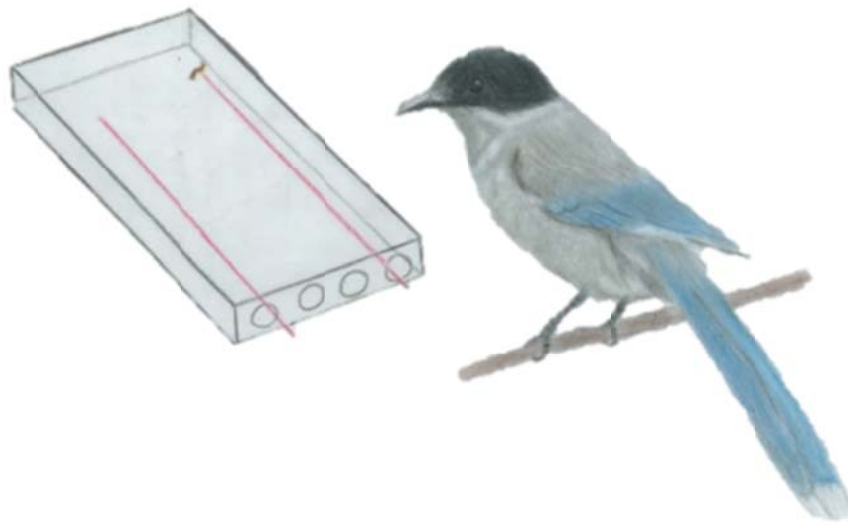
Figure 2 Pattern diagram of string-pulling in this study

Figure 3 Individual difference of performance in Experiment 2 out of 100 trials

Figure 4 Individual difference of performance in Experiment 3C out of 50 trials

Figure 5 String-pulling performance of Daniel in Experiments 1D, 2B, and 3B

Table 1 Number of correct choices in Experiment 1 out of 50 trials

	Marie	Daniel	Emily	Fatty	Tiny	Joyce
Experiment 1A	39***	27	31	26	41***	37**
Experiment 1B	40***	30	39***	28	32	39***
Experiment 1C	14**	35**	21	24	22	18
Experiment 1D	23	25	23	21	10***	17

Significant results (according to a two-tailed binomial test with chance level at 50%) are marked with * ($P < 0.05$),

** ($P < 0.01$), or *** ($P < 0.001$)

Table 2 Number of correct choices in Experiment 2 out of 50 trials

	Marie	Daniel	Emily	Fatty	Tiny	Joyce
Experiment 2A	48***	30	27	31	36**	35**
Experiment 2B	47***	34*	29	45***	39***	24

Significant results (according to a two-tailed binomial test with chance level at 50%) are marked with * (P < 0.05),

** (P < 0.01), or *** ($P < 0.001$)

Table 3 Number of choices of the short string in Experiment 3 out of 50 trials

	Marie	Daniel	Emily	Fatty	Tiny	Joyce
Experiment 3A	41***	39***	25	20	27	37**
Experiment 3B	41***	41***	30	29	30	28
Experiment 3C	35**	30	24	34*	10***	15**

Significant results (according to a two-tailed binomial test with chance level at 50%) are marked with * ($P < 0.05$), ** ($P < 0.01$), or *** ($P < 0.001$)

Table 4 Contrast of the first 25 trials and the second 25 trials

	1A	1B	1C	1D	2A	2B	3A	3B	3C
Marie	↓	-	↓	↑	-	↓	↑	↓	↓
Daniel	↑	↑	↓	↑	↓	↑	↑	↑	↓
Emily	↑	↓	↑	↓	↓	↑	↓	↑	↑
Fatty	↑	↑	-	↑	↑	↑	-	↓	↓
Tiny	↑	↑	↓	↑	↑	↑	↓	↑	-
Joyce	↓	↓	-	↑	↑	↓	↓	↑	↑

"↓" means that the subject performed better in the first half of the test than in the second half. "↑" has the opposite meaning. In addition, "-" means the correct rate is invariable. As the table indicates, the correct rate of tests is varied. The colored arrows mean that the date is significant.