Published in final edited form as:

Anim Cogn. 2012 September ; 15(5): 1021–1030. doi:10.1007/s10071-012-0511-0.

Performance on patterned string problems by common marmosets (*Callithrix jacchus*)

Matthew Gagne, Kathryn Levesque, Lauren Nutile, and Charles Locurto

College of the Holy Cross, Department of Psychology, 1 College Street, Worcester, MA 01610

Abstract

This experiment examined the performance of common marmosets (*Callithrix jacchus*) on a series of patterned string problems to assess the marmosets' understanding of means-ends relationships. One marmoset, Jet, was exposed to a series of problems that were ordered in terms of perceived difficulty during two testings that were separated by one year. In the second testing Jet received problems that had been used during the first testing along with three new problems. Each of the new problems was designed to be an exemplar of the type of problem that Jet had experienced difficulty with in the first testing. A second marmoset, Peaches, was tested on the same set of problems given to Jet in the second testing. Results indicated that the marmosets' performance on these problems fell into three categories. In one category, some problems were solved without evidence of trial and error learning. In a second category, there were problems in which the marmosets responded at chance levels initially but evidenced improvement as a function of extended testing. In a third category, some problems appeared to be virtually unsolvable even with extended testing. Taken together, these results indicate that the marmosets were able to learn the means-ends connection between pulling a string and obtaining food. This learning was best characterized as a trial and error process for some problem forms, while for others there appeared to be rapid learning that did not require extensive practice. The instances of rapid learning may be the result of the application of a simple spatial proximity rule in which the marmosets chose the string that was closest to an imaginary line drawn between the marmoset and the reinforcer.

Keywords

string pulling; insight; trial and error; marmosets

Comparative psychologists have long been interested in the issue of whether nonhumans display evidence of learning that is not directly the result of trial and error experience. Historically, this issue was notably illustrated by Köhler's observations of chimpanzees on the island of Tenerife (Köhler, 1925/1976). As is well known, Köhler reported that at least some of his chimpanzees were able to solve problems that involved obtaining food that was out of their immediate reach by constructing and using tools. Their performance appeared to follow a pattern in which the chimpanzees would first experience a failure on the problem when trying to obtain the food by direct reach. They would then pause and they appeared to search for potential tools among the objects in their enclosure such as sticks to extend their search or boxes to stand on. They would then attempt the problem using these tools. The chimpanzees' successes led Köhler to describe their performance as involving insight, defined by him as a sudden and flexible re-organization of the chimpanzees' behavior (see Thorpe, 1964, for a similar definition). It was Köhler's contention that in circumstances like

those arranged for his chimpanzees in which the subject could perceive all relevant aspects of the problem needed for solution, the subject may solve the problem cognitively, that is, without the necessity of overt trial and error practice.

Not surprisingly, Köhler's work, particularly his interpretation, drew considerable attention. The principal criticism was that Köhler could not possibly have known the entire experiential history of his subjects. Was it not possible, therefore, that considerable prior exposure to the components of the tools used by his chimpanzees underlay their apparent sudden success? Similarly, there was skepticism expressed by some workers as to whether the behavior of his chimpanzees was indeed flexible and showed evidence of insight. Films of Köhler's chimpanzees engaged in these problems suggested to some workers that the chimpanzees were far from flexible and insightful in their problem-solving behavior. As described by Gould and Gould (1994), for example, "The crate stacking behavior, after nearly three years of practice, remains laughably inept: the animals still try to balance on a corner, or even position the crates so that the open end faces up, which makes it difficult either to stack additional crates or find stable footing on the existing pile" (p. 79). Other workers noted that in attempts to replicate Köhler's work, it was vital that the chimpanzees had considerable opportunities to manipulate the objects that were to be later used as tools (Birch, 1945, Schiller, 1957). Additionally, work by Epstein et al. (1984) indicated that with sufficient trial and error training, pigeons could perform successfully on problems that were based on Köhler's work with chimpanzees. A reasonable summary of Köhler's work with chimpanzees would credit him for bringing to the fore a challenge to a behavioristic interpretation of problem solving, but would recognize that his interpretation had not been fully supported.

A number of other types of problems have been used to study the same issue of trial and error learning versus an understanding of physical causality. Among them, patterned string problems have been applied to a broad sample of species, including Köhler's chimpanzees. In these problems the subject is confronted with one or more strings. One string has a reinforcer at its far end, and if pulled, will bring the reinforcer within reach of the subject. The remaining string or strings are distracters and do not if pulled lead to reinforcement. While the strings can be arranged in a number of different patterns, two basic configurations have been most frequently used. One is a parallel configuration in which two or more strings are arranged in parallel to each other, and one string is baited with the reinforcer at its tip. The parallel strings may be perpendicular to the subject or presented at an acute angle to the subject. Examples of string configurations can be seen in Figure 1 which illustrates the string configurations used in the present work. In all of these configurations the subject is located at the open end of the strings while the reinforcer and a distracter (filled circle and heart-shaped object, respectively) are located at the far end of the strings. Problems B, C and J represent parallel string configurations. Problems B and C are presented perpendicular to the subject whereas in problem J the parallel strings are presented at an acute angle to the subject in all four forms of that problem. In the second often-used configuration two or more strings are crossed in some manner, or would cross if extended, with one string baited. In Figure 1, problems H, I, K, L, M, N, and O, are examples of this second type of string configuration.

There are a number of problems in Figure 1 that could be expected to be solved quickly based on prior work. We used these problems as introductory problems to habituate the subject to the testing protocol and arena. Some of these problems are single string problems, namely problems A, D, E, F, and G. Others are two-string parallel configurations in which the strings lie perpendicular to the subject's position, namely, problems B and C. It can be observed that in some but not all of the parallel string configurations the correct string lies closest to an imaginary line drawn from the reinforcer to the plane defined by the tips of the

strings. Examples of this idea are found in problems B, C, H, J-b and J-d, K-b and K-c. In these problems the subject might choose the correct string by following the simple rule of choosing the string that was closest to that imaginary line. We will hereafter refer to these types of problems as linear problems. In some parallel problems forms, namely, J-a and J-c, and in the crossed or near-crossed string configurations, the correct string lies further from that imaginary line than does the distracter string. As an example, in the two-string crossed or “X” configuration, problem L, the end of the string closest to that imaginary line is the distracter. Problem forms K-a and K-d are additional examples of this type of problem. We will hereafter refer to these problems as nonlinear problems. This distinction between configurations that can be solved by pulling the string closest to the imaginary line, linear problems, versus those that cannot be solved in this manner, nonlinear problems, is relevant to the results of prior experiments and the present experiment.

The standard strategy for studying these problems is to expose subjects to a limited number of trials on a given string configuration to determine whether that configuration can be mastered without extensive trial and error practice, thereby suggesting that some form of understanding may be operative. A study using dogs by Osthaus et al. (2005) illustrates this strategy. In this study dogs were exposed to a number of different string patterns. These patterns may be grouped into three types of configurations. One type consisted of single string problems, similar to our problems A, D, E, F, and G. The second problem type consisted of two parallel string configurations, corresponding to our problems B, C, and the four forms of problem J. The third string type consisted of crossed strings, much like our problem L. The procedural strategy was to give subjects a limited number of trials, 10 – 20 trials on a given problem, with the entire test protocol administered in a single session. Results indicated that dogs were highly successful with problems of the first pattern type, that is, with single string problems. They were also successful with parallel string arrangements that resembled our linear problems. The subjects were least successful with problems that were nonlinear in nature. Indeed, when two strings were crossed, as in our problem L, 8 of 12 dogs performed at below chance levels, suggesting that they were responding as though the problem was linear in nature and they chose the string closest to the imaginary line between the reinforcer and themselves. The remaining four dogs were at chance on this problem. It appears from this work that dogs gave some evidence of understanding of physical causality in the simpler linear problems, but performance did not reveal evidence of anything beyond associative learning in the nonlinear problems. A later study obtained similar results with cats (Whitt et al., 2009).

As earlier noted, a wide variety of species have been studied on these problems, ranging from Old and New World primates to avian species. Tomasello & Call (1997) provide a summary of primate work (Figure 2.7, pp. 50–51.). Hauser, Kralik, & Botto-Mahan (1999) studied cotton-top tamarins, a New World species as are the common marmosets used in the present work. Interestingly, Beck’s (1967) work with gibbons suggests that this diminutive or “lesser” ape, performed as well or better than did Kohler’s chimpanzees on a set of these problems. The avian species studied include goldfinches and siskins (Seibt & Wickler, 2006), kea (Huber & Gajdon, 2006), ravens (Heinrich & Bugnyar, 2005), grey parrots (Pepperberg, 2004), and greenfinches and canaries (Vince, 1961). In avian studies, food is attached to a string that is hung from a perch, and the subject is required to pull the string up by using a combination of pull-and-hold movements, with the collected string held by the bird’s feet. This format allows for parallel string configurations as well as crossed-string problems.

Taken together there appears to be rather limited systematic evidence in any of this work, across a wide range of species, that patterned strings are solved in a manner that clearly reflects an understanding of physical causality rather than the results of trial and error

associative learning. In an effort to explore the limits of this conclusion the present work contained several features not commonly found in prior work. We exposed marmosets to a larger set of both linear and nonlinear problems than was the case in previous research. We also continued testing on problems that were not solved within the limits of our acquisition criterion to observe whether extended testing on a problem might reveal improvement that was not evident in a smaller set of trials. Additionally, we included a type of prompt and fade strategy on the most difficult problem, problem L, to determine whether successful performance might be encouraged if the problem were made less difficult, much in the manner of prompt and fade strategies that are used in behavioral therapy (e. g., Alberto & Troutman, 2006). We also replicated the testing for one marmoset, Jet, one year after initial testing to determine whether there were savings over time that may not have been predictable from initial performance.

Method

Subjects

One subject was Jet, a seven year-old female Common marmoset (*Callithrix jacchus*). Jet had been born and was housed at the New England Primate Research Center, Southborough Massachusetts. She was housed with another female marmoset, Peaches, age seven, who was used along with Jet at the time of the one-year retesting of Jet. For testing purposes the marmosets' cage was fitted with a divider that placed each marmoset in a separate half of the cage and obscured one marmoset's view of the testing of the other marmoset. The marmosets were divided in this manner two to three hours prior to a session. Their daily feeding regimen consisted of a morning feeding and a second feeding in the afternoon. On testing days they received their morning feeding for approximately two hours. Food was removed three hours prior to testing, which took place in the early afternoon. The marmosets received their afternoon feeding immediately following a session. Testing was conducted on average four days each week. Both marmosets had been exposed to a number of cognitive tasks prior to this experiment, including discrimination problems, tool use problems (using a tool form adapted from Hauser, Pearson, & Seelig, 2002, Experiment 1), and reverse contingency problems. Neither marmoset had been exposed to string pulling prior to this experiment. Subjects were maintained in accordance with the guidelines of the Committee on Animals of the Harvard Medical School.

Apparatus

The marmosets remained in their home cages during testing. The string problems were presented to them on a cardboard tray that was approximately 36 cm in length and 24 cm in width. The tray was secured horizontally to the front of the home cage at the lower edge of the cage door via hooks. For each problem the subject reached through the cage door to make a choice. Wooden dowels approximately .5 cm in diameter extended approximately 2 cm above the board and were arranged in a matrix around the board to allow the strings to be configured for a particular problem. We used two boards, each of which had a slightly different arrangement of wooden dowels. The different arrangement of dowels facilitated the configuration of the strings for a particular problem. The middle of the board was aligned with the middle of the cage door for each problem, with the result that the subject faced the middle of the board when choosing a string.

The strings were approximately 48 cm in length and were composed of metal link chains. One of the two strings had a small metal bottle cap at one end in which was placed the reinforcer, a mini-marshmallow (Swiss Miss). This string was always the correct response option. For problems that involved two strings, the second string had a wooden heart at one end that was approximately the same size as the bottle cap. This string was always the

incorrect response option. The other end of each string hung off the side of the board nearest the cage and was within reach of the subject.

Procedure

The arrangement of the strings for each problem is given in Figure 1. The circular object at the end of one string represents the bottle cap into which the reinforcer was placed. The subject faced each problem from the side at which the strings terminated. Problems A through L were used in the first testing of Jet and were presented in alphabetical order. A subset of these problems, along with problems M, N, and O were used for Jet's second testing and for Peaches' testing. For each problem, different forms of the problem were used such that the string or strings were arranged in different positional forms. For example, in problem A where only the correct string was used, that string was arranged in one of three possible positions an equal number of times across the trials within a session. Similarly, in problem B, where two strings were used for the first time, each string occurred in each possible position an equal number of times across the trials within a session.

The problems of principal interest were the nonlinear problem forms J-a and J-c, K-a and K-d, and problem L. As noted earlier, in these problems the correct string was further from an imaginary line dropped between the reinforcer and the subject than was the incorrect string. For Jet's second testing and Peaches' testing, problems M, N, and O constituted additional exemplars of this problem type.

A string was considered selected if the subject grasped a string and moved the object at the end of the string, either the bottle cap or the wooden heart. If the subject grasped the incorrect string first and moved that string, and then chose the correct string, we counted these self-corrections as errors. Self-corrections accounted for approximately 1% of Jet's errors across the first and second testing. They constituted approximately 5% of Peaches' choices.

For all problems, a trial terminated when a string was selected and moved, or when 120 sec elapsed, timed from the presentation of the tray to the subject with the problem in place. When one trial ended a cardboard sheet was placed in front of the subject's cage to block the subject's view during the set-up for the next problem. The intertrial interval (ITI) was approximately 30 sec. The criterion for passing a problem was set at a proportion correct for an entire session in which $P \geq .05$. The sign test was used for all tests involving proportion correct.

Jet's first testing—For Jet's first testing, problems A and B were considered adaptation problems. Both problems were presented in the first session of training, problem A for 3 trials and problem B for 10 trials. We scheduled these problems for only one session unless Jet did not reach the criterion of $P \geq .05$ across the two problems. Problems C and H which each consisted of two string forms, were presented in sessions that consisted of 10 trials, with each string arrangement presented in randomized order for 5 trials. For these problems, the criterion for successfully passing the problem translated to .80 correct for a session. Problems D, E, F and G were included to introduce the subject to problems in which the correct string was non-parallel with respect to the subject's position. These four problems were each presented for one 10-trial session, with each problem form presented for an equal number of trials. There were no errors for these four problems except for time outs (120 sec without a choice).

Beginning with problem I sessions were expanded to include two blocks of trials. For problems I and L, each of which had two string forms, sessions consisted of two 10-trial blocks with each string form presented for 10 trials within a session (20 total trials per

session) and for an equal number of trials within each block. For these problems, the criterion for successful passage at $P = .05$ translated to .75 correct across the session. Problems J and K each had four forms of string arrangements. These problems were presented in sessions that consisted of two blocks of 12 trials each with each string form presented for six trials within a session (24 total trials per session). For these problems the criterion for successfully passing the problem translated into .70 correct for an entire session. One session was conducted per day. In cases where blocks of trials were given there was approximately a 15-min break between blocks.

We set testing limits on a given problem at a maximum of 40 trials for problems that consisted of two forms (problems C, H, I and L), and at 48 trials for problems that consisted of four forms (problems J and K). If the proportion correct criterion was not met within these trial limits we considered that problem failed. In Jet's first testing, the first problem that she failed given these criteria was problem J. She also did not meet criterion on problems K and L. For these last two problems we decided to consider the problem failed but to continue beyond our stated testing limits to determine whether improvements in performance would be observed as a function of extended testing.

For problem K Jet was continued for two additional sessions beyond the two-session criterion. For problem L, as described below, Jet recorded no correct responses during the first session. Given that performance and her earlier failures on problems J and K, we decided that Jet would be continued on problem L under altered conditions of training following her first session. She was continued for a total of 13 sessions beyond the first session. During these additional sessions Jet was first given a prompt in the form of black electrical tape wrapped around the entire length of the correct string for the first five of these sessions. These sessions were followed by three sessions with no tape prompt, one session with full tape, three sessions with one-half of the correct string taped as a form of fading the prompt, and one final session with no tape prompt.

Jet's second testing and Peaches' testing—Jet was tested for a second time one year following the first testing. The apparatus was the same as was used in the first testing and the conditions of testing and food deprivation were the same. Three additional problems, M, N, and O were used in this experiment, each of which was designed to be nonlinear. Problems were again presented in alphabetical order. In this testing all problems were presented to the subjects in two blocks of 10 trials each session, or 20 total trials each session. This procedure represents a change from the first testing in which some sessions consisted of 10 trials, others consisted of 20 trials, and problems that consisted of four forms were presented in blocks of 12 trials. With this procedure, meeting the criterion of $P = .05$ across an entire session consisting of all problem forms required a proportion correct of .75.

We alternated blocks between the two subjects so that Jet received her first block of trials in a session, then Peaches received her first block. Each subject then received their second block of trials in the same order. Both marmosets received a shortened protocol of problems in this second testing. Both subjects received problems A, C and H as introductory problems before proceeding to problem J. Problems with one response choice, problems D, E, F, and G were not used in this second testing, nor was problem I used.

As was true in Jet's first testing, we considered a problem failed if the criterion was not met in two sessions. As was done in Jet's first testing, some problems were continued beyond that point. Additionally, as was done in Jet's first testing, we used a prompt and fade strategy on problem L during extended training.

Results

Jet's first testing

A summary of Jet's and Peaches' performances is presented in Table 1. The columns labeled "proportion correct" reflects performance during the session in which the criterion was met, or the last session of training if the criterion was not met. Additional training sessions beyond the criterion session are not included in Table 1. The only problem among problems C through H that required more than one 10-trial session was problem C, which required four 10-trial sessions to reach criterion. Jet's errors on this problem were equally distributed between the two forms of the problem. The next five problems, D through H, were passed within one 10-trial session.

Problem I required two 20-trial sessions to reach criterion. Errors on this problem were made exclusively when the correct string was to the left (form I-a). The proportion of correct choices when the correct string was on the right (form I-b) was 1.0 across the two sessions. The proportion of correct choices when the correct string was on the left were .14 during the first block of 10 trials during session 1. During the second block of 10 trials in session 1, Jet recorded four consecutive time outs. As a result, the session was terminated. The reason for the discrepant performance on the two forms of the problem during the first session was that Jet developed a right-side bias. During the second session on this problem, Jet's right-side bias was less in evidence. Her overall proportion correct was .85 across the two blocks of testing during the second session.

Problem J was the first problem in which Jet did not reach criterion. Forms J-b and J-d, were mastered without trial and error learning. Over the course of two sessions of testing that involved 24 presentations of forms J-b and J-d combined, Jet's performance was errorless. Her responding to forms J-a and J-c revealed a different pattern. Jet was correct on only 2 out of 6 presentations of each form in the second training session. The frequency of these correct choices for J-a and J-c were not different from chance using the sign test ($P > .05$ for each problem form). The reason for the qualitatively different performance on forms J-b and J-d compared to forms J-a and J-c was likely the relationship between the correct string and the location of the reinforcer. Looking at the different forms for this problem presented in Figure 1, it can be seen that for forms J-b and J-d the correct string was closest to an imaginary line running between the reinforcer and the position of the subject. For forms J-a and J-c the opposite was true. As a consequence, forms J-b and J-d were in our terminology linear problems, whereas the remaining two forms were nonlinear problems.

The pattern observed in problem J in which some forms of the problem were far more easily mastered than were others was observed again in problem K. Averaged across all forms of problem K, criterion was not met in two sessions. Jet's inability to master this problem was principally a function of forms K-a and K-d. The combined proportion correct on these two problem forms was .17 during the second session, whereas the proportion correct was .75 for forms K-b and K-c combined. The difference between these two proportions was significant using the binomial test at $P = .05$. We continued testing for two additional sessions past the point where the problem was considered failed to determine if Jet would increase her proficiency with form K-d. Overall, problem K was presented for four sessions. In three of those sessions Jet received 24 trials. During the first 12-trial block in session 2 Jet failed to meet the time criterion on three consecutive trials, so the block was discontinued and Jet was considered to have failed all trials in this block. After a 10-minute break, she received the second 12-trial block, and she completed all 12 trials. Across all four sessions she received 21 trials on each form of problem K in a total of 84 trials. Over the course of these four sessions she did not choose the correct string in form K-d during any trial. It should be added that none of Jet's incorrect choice on form K-d were the result of exceeding time

limits for a trial. Form K-d is the symmetrical complement to form K-a in that the correct string lies further from the imaginary line drawn between the subject and the reinforcer than does the incorrect string. Jet produced correct choices on 6 out of 21 exposures to form K-a across the four sessions. This level of success on form K-a was significantly less than chance using the sign test ($P = .039$).

In problem L, each form of the problem is nonlinear. As earlier noted in the description of methods, in the first session on problem L Jet recorded no correct choices on either form of the problem and both blocks of trials had to be terminated before they were completed due to numerous timeouts. Because of this failure, and in light of her performance on problems with similar characteristics (J-a and J-c, along with K-a and K-d) in which Jet was unable to make correct choices at an above chance, we decided to follow a different strategy in presenting problem L following the first session. We adopted a type of prompt and fade strategy beginning in the second session. The prompt consisted of wrapping the correct string in black tape, thereby making this string distinctive compared to the incorrect string. This altered string was used in the next five testing sessions. Jet required four sessions with this prompt before reaching our acquisition criterion of .75 correct for a session. We continued the tape for one more session (proportion correct = .80) before presenting Jet with three sessions of the original problem, that is, without the prompt. Her performance fell to chance levels (proportion correct = .37). In the following session (session 10) Jet again received a session in which the correct string was wrapped in tape, and her performance once again met criterion. For the next three sessions we partially faded the prompt by removing one-half of the tape from the correct string. Jet's performance met criterion in the last two of these three sessions. On the final session (session 14) we removed the tape entirely from the correct string. Jet's performance once again did not reach criterion (proportion correct = .65).

Jet's second testing and Peaches' testing

Jet's second testing was conducted one year after the first testing. In this second testing we sought to retest Jet on the problems she experienced earlier to determine whether there were savings evident in her performance. We also presented her with additional problems, problems, M, N, and O, which were developed to test further her ability to solve nonlinear problems. We also tested her cagemate Peaches, using the same order of problems that was given to Jet in this second testing

Jet received one session of problem A as an adaptation session followed by problem C. Unlike her performance in the first testing in which she required four sessions (40 trials) to reach criterion, Jet's proportion correct choices on problem C met criterion in the first session. She also met criterion in one session on problem H, as she had done during the first testing. Jet met criterion on problem J in the second session in this testing. In the first testing she had failed to reach criterion within two sessions. In that first testing Jet had difficulty with forms J-a and J-c. In the second testing Jet was at chance levels on these two forms in the first session (proportion correct = .50, averaged across both forms). In her second session Jet averaged .90 on these two forms ($P < .05$ on both J-a and J-c). As was true in first testing, Jet did not record any incorrect choices to forms J-b and J-d across the two sessions of testing.

Jet did not meet criterion within two sessions of testing on problem K. We continued her for three additional sessions. In the first testing Jet did not reach criterion on this problem after four sessions of testing. In that first testing she showed evidence of significant difficulty with form K-d, and to a lesser extent form K-a. That same difficulty with form K-d was again evident in this testing. Jet did not record any correct choice to form K-d across the five sessions of testing. She met criterion in the fifth session because her choice on forms K-b

and K-c were nearly errorless (only one incorrect response to form K-c across the five sessions) and her performance on form K-a was sufficiently high to produce an overall proportion correct of .75 in the fifth session.

In the first testing Jet had demonstrated considerable difficulty on problem L. Her difficulties with problem L were equally divided between the two forms of the problem. In her second testing Jet again failed to reach criterion. We continued her for a total of five sessions without the tape prompt. The distribution of her errors was different than was observed in the first testing. Jet adopted a “choose right” bias, the result of which was that across the first five sessions of testing Jet was correct in 49 out of 50 choices of form L-a, in which pulling the right-side string was correct. She was incorrect in 48 out of 50 choices on form L-b, when the left-hand choice was correct. Over the last three sessions of this phase of problem L, Jet also recorded 22 time outs on form L-b during this phase. She recorded only two time-outs on all other problems in this second testing

Following these five sessions, we reinstituted the full tape prompt for two sessions during which Jet produced errorless performance on both forms of the problem. When the prompt was removed Jet met criterion in her first session. We continued her for one more session without the prompt, during which Jet produced errorless performance on both forms of the problem.

As can be seen in Table 1, Jet met criterion on problems M, N, and O in the first session. These problems were not used in the first testing and each problem was presented in a nonlinear form.

Peaches

Peaches received problems A, C and H before being run on the first of the nonlinear problems, problem J. As can be seen in Table 1 Peaches reached criterion in her first session of both problems C and H. Peaches did not reach criterion in two sessions on problem J. Similar to Jet’s first testing results, Peaches’ performance on this problem was composed of nearly errorless performance on forms J-b and J-d, chance performance on form J-a, and less than chance performance on form J-c ($P = .010$).

Like Jet, Peaches also did not meet criterion on problem K. She produced no correct choice to form K-a ($P = .001$) and she was at chance performance on the complementary problem form K-d. We continued her for four sessions beyond her second session. In these four additional sessions she remained significantly below chance on forms K-a and K-d ($P = .01$ on each form), and she did not meet criterion during any session. Her errors on these problem forms were principally attempted self-corrections in which she chose the incorrect string first and moved it towards her, then stopped and chose the correct string. She recorded 17 self-corrections across the six sessions of exposure to problem K, 37 percent of her total self-corrections across all sessions of testing. As earlier noted, these self-corrections were treated as errors. All but one of these self-corrections occurred on problem forms K-a and K-d.

As was true for Jet in both testings, Peaches was significantly below chance in her first two sessions of exposure to problem L. We continued her for a third session. She did not show evidence of a position preference, but her choices were incorrect in 59 out of the first 60 choices made on this problem across the three sessions ($P = .001$). After the third session we gave Peaches the same full-tape prompt that we used with Jet for five sessions. Peaches showed improvement during each prompt session and produced errorless performance during the last two sessions. When returned to the no-prompt condition, she met criterion in

her third session. We continued her for one additional session. In that last session she produced errorless performance.

We presented Peaches with the three problems that were added in the second testing for Jet, problems M, N, and O. As can be seen in Table 1, Peaches reached criterion in one session on problem M and in two sessions on problem O. She did not reach criterion on problem N. We continued her for two additional sessions on problem N, and she met criterion during the fourth session. Over the four sessions on problem N, Peaches produced a significant proportion of correct choices to form N-a ($P = .001$) and chance performance to form N-b.

Discussion

The principal issue that motivated this work was whether performance on patterned string problems might be characterized as revealing evidence for the presence of an understanding of physical causality. If this issue is narrowly framed by asking whether there is evidence that marmosets can use strings as a tool to obtain reinforcement without the necessity of extensive trial and error training, the answer is positive. The marmosets quickly mastered single string problems (problems A, D, E, F, and G) and the linear problem H was mastered within one session by both Jet and Peaches (see Halsey et al., 2006, for evidence that wild marmosets can also master a linear two-string problem, though not without trial and error learning). Both Jet and Peaches also demonstrated nearly-errorless performance on linear problem forms such as J-b, J-d, K-b and K-c.

It should be added in this regard that there was clear evidence of savings in Jet's performance during the second testing, particularly in her performance on problem J which she had failed during her initial exposure to this problem. She also mastered problem C within one session in her second testing compared to the four sessions that were required in the first testing. The presence of savings, that is, improved relearning of material previously learned, cannot be taken as evidence of a generalized problem-solving strategy, but it does indicate some memory of prior learning that is available at the time of relearning. (See Monk et al., 1996, for other evidence of savings in a nonhuman primate.)

If the issue is framed more broadly as whether there is evidence for the emergence of correct performance without extensive training across a variety of problem types, particularly across non-linear problem types, the answer is less encouraging. Problem L is the most evident example of this problem, along with the nonlinear problem forms J-a and J-c, K-a and K-d. It should be noted that problem L is one that has been used with a number of species ranging from great apes to avians, and it has often proven to be of considerable difficulty (dogs: Osthaus et al., 2005; cats: Whitt et al., 2009; crows: Taylor et al., 2010). To our knowledge no previous study has engaged in the amount of extended testing that we used for this problem, nor has any study to our knowledge used a prompt and fade strategy to encourage improved performance. Jet received 14 sessions of problem L in her first testing, including several sessions in which a prompt and fade strategy was used. Despite this extensive training, Jet was unable to meet criterion. In her second exposure to this problem Jet failed the problem for the first five sessions, and only following another prompt and fade intervention was she able to meet criterion. At that point Jet had received 23 sessions of exposure to problem L. Peaches required 11 sessions before meeting criterion on problem L. She failed the problem during the first three sessions. She then required four sessions that included a full-tape prompt before she met criterion with the prompt. She required three additional sessions without a prompt before meeting criterion.

The results for problem L indicate that extensive associative learning was necessary to produce correct performance. They also suggest that there may be limited success in using

interventions in the form of prompt and fade. That conclusion should not be taken to mean, however, that there are no alterations of these problems that may facilitate improved performance. There is evidence that some alterations can result in dramatic improvements. Hood et al. (1999) demonstrated that cotton-top tamarins, a New World primate that occupies a similar ecological niche to marmosets, have difficulty in solving gravity problems, problems in which food is dropped down a chimney connected by an opaque tube to one of three containers. Much like young children, there was a significant tendency for tamarins to search in the container directly beneath the chimney where the food was dropped on a given trial, hence the name gravity problems. However, in a later study Hauser et al. (2001) demonstrated that if gravity could be eliminated as a cue by presenting the problem horizontally rather than vertically to the tamarins, improvements in performance followed. This particular manipulation is probably not relevant to the present work in that the string problems were presented horizontally (in Hauser et al.'s terms) to subjects, but this work does suggest that one should not rule out the possibility that perhaps other alterations of string problems might prove effective.

The finding that the use of prompts was of limited value, coupled with the need for extensive training on a number of problems, provides little evidence for the presence of some generalized understanding of the constituents of effective performance. Given this pattern, it should be asked: What were the marmosets learning during testing? They did not appear to be attending to the relationship between the string-as-tool and the reinforcer, and, as a result, learning a generalized rule about that relationship that might be characterized: "Pull the string that can be traced back to the reinforcer." Instead, they appeared to be tracking the location of the reinforcer and then behaving as though they were asking: "Which string is closest to a line drawn between the reinforcer and me?" In this manner, their performance when successful might be characterized as resulting from the invocation of a simple spatial proximity rule in which the chosen string is closest to an imaginary line drawn between the reinforcer and the subject.

It is of interest that for Jet the three new problems in the second testing were so quickly mastered. Her performance may represent the emergence of a more efficacious performance strategy that by-passes the spatial proximity rule. But it should be added that we have no way to determine if these three new nonlinear problems were of the same difficulty level as the original nonlinear problems that were used in each testing. It is tempting to imagine what the interpretation of Jet's performance on problems M, N and O would be for an observer who had not witnessed Jet's training on other patterned string problems. Jet recorded only one error on problem M, four errors on problem N (all to form N-b), and four errors to problem O, out of 20 trials for each problem. It would not be surprising if such an observer concluded that Jet's performance was evidence for an interpretation based on an understanding of physical causality. Yet, if that same observer had witnessed the entire breadth of Jet's training, her hard-won gains on earlier nonlinear problems and the extensive use of prompt and fade strategies for problem L, the interpretation might be markedly different.

Performance on the three new nonlinear problems in the second testing may offer evidence of the development of a more effective learning strategy, but the overall impression garnered from the results considered as a whole must also emphasize the perseveration of incorrect choices over long periods of testing on most of the nonlinear problem forms. This kind of perseveration is certainly not unique to string pulling problems. There is a long history in comparative psychology of instances in which incorrect responding perseverates in the face of contingencies that should encourage change (e. g., detour problems in dogs, Osthaus et al., 2010; object discrimination in Japanese monkeys, Itoh et al., 2001; A-not-B type choice problems in human infants, Diedrich et al., 2000; maze learning, Carr & Watson, 1908). In

the present case this perseveration was most evident precisely for those problems in which successful performance would have provided the clearest evidence for the presence of a flexible understanding of physical causality.

Acknowledgments

This work was supported by grant RR00168 from the National Institutes of Health to the New England Primate Research Center (NEPRC), Southborough, MA. We thank the staff at NEPRC for its support of this work.

References

- Alberto, PA.; Troutman, AC. Applied behavior analysis for teachers. 7. Pearson Education; Upper Saddle River, NJ: 2006.
- Beck BB. A study of problem solving by gibbons. *Behaviour*. 1967; 28:95–109. [PubMed: 6016362]
- Birch HG. The role of motivational factors in insightful problem-solving. *J Comp Psychol*. 1945; 38 : 295–317. [PubMed: 21005111]
- Carr H, Watson JB. Orientation in the White Rat. *J Comp Neurol Psychol*. 1908; 18:27–44.
- Diedrich FJ, Thelen E, Smith LB, Corbetta D. Motor memory is a factor in infant perseverative errors. *Dev Sci*. 2000; 3:479–494.
- Epstein R, Kirshni CE, Lanza RP, Rubin LC. ‘Insight’ in the pigeon: Antecedents and determinants of an intelligent performance. *Nature*. 1984; 308:61–62. [PubMed: 6700713]
- Gould, JL.; Gould, CG. The animal mind. Scientific American Library/Scientific American Books; New York: 1994.
- Halsey LG, Bezerra BM, Souto AS. Can wild common marmosets (*Callithrix jacchus*) solve the parallel strings task? *Anim Cogn*. 2006; 9:229–233. [PubMed: 16541239]
- Hauser MD, Kralik J, Botto-Mahan C. Problem solving and functional design features: Experiments on cotton-top tamarins, *Saguinus oedipus*. *Anim Behav*. 1999; 57:565–582. [PubMed: 10196046]
- Hauser MD, Pearson H, Seelig D. Ontogeny of tool use in cottontop tamarins, *Saguinus oedipus* : Innate recognition of functionally relevant features. *Anim Behav*. 2002; 64:299–311.
- Hauser MD, Williams T, Kralik JD, Moskovitz D. What guides a search for food that has disappeared? Experiments on cotton-top tamarins (*Saguinus oedipus*). *J Comp Psychol*. 2001; 115:140–151. [PubMed: 11459161]
- Heinrich B, Bugnyar T. Testing problem solving in ravens: String-pulling to reach food. *Ethology*. 2005; 111:962–976.
- Hood BM, Hauser MD, Anderson L, Santos L. Gravity biases in a non-human primate? *Dev Sci*. 1999; 2:35–41.
- Huber L, Gajdon GK. Technical intelligence in animals: The kea model. *Anim Cogn*. 2006; 9 :295–305. [PubMed: 16909237]
- Itoh K, Izumi A, Kojima S. Object discrimination learning in aged Japanese monkeys. *Behav Neurosci*. 2001; 115:259–270. [PubMed: 11345953]
- Köhler, W. The mentality of apes. Winter, E., editor. Liveright, Oxford; 1925/1976.
- Monk CS, Gunderson VM, Gran KS, Mechling JL. A demonstration of the memory savings effect in infant monkeys. *Dev Psychol*. 1996; 32:1051–1055.
- Osthaus B, Lea SG, Slate AM. Dogs (*Canis lupus familiaris*) fail to show understanding of means-end connections in a string-pulling task. *Anim Cogn*. 2005; 8:37–47. [PubMed: 15338446]
- Osthaus B, Marlow D, Ducat P. Minding the gap: Spatial perseveration error in dogs. *Anim Cogn*. 2010; 13:881–885. [PubMed: 20563828]
- Pepperberg IM. ‘Insightful’ string-pulling in Grey parrots (*Psittacus erithacus*) is affected by vocal competence. *Anim Cogn*. 2004; 7:263–266. [PubMed: 15045620]
- Schiller, PH. Innate motor action as a basis of Learning. In: Schiller, CH., editor. *Instinctive behavior; the development of a modern concept*. International University Press; Oxford: 1957. p. 264-287.
- Seibt U, Wickler W. Individuality in problem solving: String pulling in two *Carduelis* Species (Aves: Passeriformes). *Ethology*. 2006; 112:493–502.

- Taylor AH, Medina FS, Holzhaider JC, Hearne LJ, Hunt GR, Gray RD. An investigation into the cognition behind spontaneous string pulling in New Caledonian crows. PLoS One. 2010; 5:e9345. [PubMed: 20179759]
- Thorpe, WH. Learning and Instinct in animals. Methuen; London: 1964.
- Tomasello, M.; Call, J. Primate cognition. Oxford; New York: 1997.
- Vince MA. String-pulling in birds: III. The successful response in greenfinches and canaries. Behaviour. 1961; 17:103–129.
- Whitt E, Douglas M, Osthaus B, Hocking I. Domestic cats (*Felis catus*) do not show causal understanding in a string-pulling task. Anim Cogn. 2009; 12:739–743. [PubMed: 19449193]

Figure 1.

Table 1

Training durations and proportion correct for all patterned string problems

Problem	Jet – first testing		Jet – second testing		Peaches	
	Sessions (trials)	Proportion correct	Sessions (trials)	Proportion correct	Sessions (trials)	Proportion correct
A summary	1(3) ^a	.77 [*] (A and B combined)	1 (20)	1.00 ^{**}	1 (20)	1.00 ^{**}
B summary	1 (10)					
C summary	4 (40)	.90 [*]	1 (20)	.85 ^{**}	1(20)	.80 [*]
C-a	4 (20)	1.00 ^{**}	1 (10)	.90 ^{**}	1 (10)	.90 ^{**}
C-b	4 (20)	.80	1 (10)	.80 ^{**}	1 (10)	.70
D summary	1 (10)	.90 ^{**}				
E summary	1 (10)	1.00 ^{**}				
F summary	1 (10)	1.00 ^{**}				
G summary	1 (10)	1.00 ^{**}				
H summary	1 (10)	.90 ^{**}	1 (20)	.95 ^{**}	1 (20)	1.00 ^{**}
H-a	1 (5)	1.00 ^{**}	1 (10)	1.00 ^{**}	1 (10)	1.00 ^{**}
H-b	1 (5)	.80	1 (10)	.90 ^{**}	1 (10)	1.00 ^{**}
I summary	2 (30)	.85 [*]	-	-	-	-
I-a	2 (15)	.70	-	-	-	-
I-b	2 (15)	1.00 ^{**}	-	-	-	-
J summary	2 (48)	.67	2 (40)	.95 ^{**}	2 (40)	.58
J-a	2 (12)	.33	2 (10)	.80	2 (10)	.40
J-b	2 (12)	1.00 ^{**}	2 (10)	1.00 [*]	2 (10)	1.00 ^{**}
J-c	2 (12)	.33	2 (10)	1.00 [*]	2 (10)	.00
J-d	2 (12)	1.00 ^{**}	2 (10)	1.00 [*]	2 (10)	.90 ^{**}
K summary	2 (48)	.54	2 (40)	.60	2 (40)	.53
K-a	2 (12)	.33	2 (10)	.40	2 (10)	.00
K-b	2 (12)	.83	2 (10)	1.00 ^{**}	2 (10)	.80 [*]
K-c	2 (12)	.67	2 (10)	1.00 ^{**}	2 (10)	1.00 [*]

Problem	Jet – first testing		Jet – second testing		Peaches	
	Sessions (trials)	Proportion correct	Sessions (trials)	Proportion correct	Sessions (trials)	Proportion correct
K-d	2 (12)	.00	2 (10)	.00	2 (10)	.30
L summary	1 (20)	.00	2 (40)	.50	2 (40)	.00
L-a	1 (10)	.00	2 (20)	1.00**	2 (20)	.00
L-b	1 (10)	.00	2 (20)	0.00	2 (20)	.00
M summary	-	-	1 (20)	.95**	1 (20)	.80**
M-a	-	-	1 (10)	1.00**	1 (10)	.80*
M-b	-	-	1 (10)	.90**	1 (10)	.80*
N summary	-	-	1 (20)	.80*	2 (40)	.60
N-a	-	-	1 (10)	1.00**	2 (20)	.70
N-b	-	-	1 (10)	.60	2 (20)	.50
O summary	-	-	1 (20)	.80*	2 (40)	.75*
O-a	-	-	1 (10)	.70	2 (20)	.60
O-b	-	-	1 (10)	.90**	2 (20)	.90**

Note:

* = P .05,

**

= P .01. The proportion correct reflects performance during the session in which the criterion was met, or in cases where the problem was failed, the proportion correct reflects performance during the second session. The number of sessions and trials reflects the amount of training up to meeting the criterion or total training until the problem was failed

^a Problems A and B were considered adaptation problems that were both presented within Jet's first session of training, 3 trials for problem A and 10 trials for problem B. The combined performance across these two problems was 10/13 (.77) correct, P = .04.