ANIMAL BEHAVIOUR, 1999, **57**, 565–582
Article No. anbe.1998.1032, available online at http://www.idealibrary.com on IDEAL

Problem solving and functional design features: experiments on cotton-top tamarins, *Saguinus oedipus oedipus*

MARC DAVID HAUSER*†‡, JERALD KRALIK* & CAREZZA BOTTO-MAHAN†
Departments of *Psychology and †Anthropology, ‡Program in Neurosciences, Harvard University

(Received 4 June 1998; initial acceptance 9 July 1998;
final acceptance 11 September 1998; MS. number: A8095R)

In any problem-solving situation, there are features associated with the problem that are relevant from a functional perspective and other features that are irrelevant. To determine whether animals are sensitive to the distinction between functionally relevant and irrelvant features of a problem, we conducted two main experiments with a New World monkey, the cotton-top tamarin. In the first condition of both experiments, subjects were required to pull a piece of cloth to gain access to a piece of food. The first experiment involved choosing between food that was on the cloth and food that was off the cloth. The second experiment involved choosing between food that was on a connected piece of cloth and food that was on two pieces of cloth separated by a horizontal gap. Having learned to solve either of these two problems, we conducted a series of probe conditions to determine whether the tamarins would generalize to changes in the shape, size, colour, and texture of the cloth and food, the position of the food relative to the cloth, and the type of connection between two pieces of cloth. For most of the probe conditions, the tamarins readily generalized, showing no decrement in performance, even on the first trial. For other conditions, involving apparently more subtle discrimination (e.g. a narrow vertical gap between the two pieces of cloth), explicit training was required. These results indicate that tamarins solve means–end relationships, and that their ability depends on a discrimination between properties that are functionally relevant as opposed to irrelevant.

Problem solving often involves an analysis of means–end relationships. In Kohler's (1925) classic experiments on chimpanzees, for example, an individual was required to place a series of boxes under a hanging banana, stand on the boxes, and then use a stick to draw the banana closer. In this situation, as in any means–end analysis, some features of the problem are relevant to the task at hand, whereas other features are irrelevant. In Kohler's task, the functionally relevant features concern the shapes and sizes of the boxes and sticks. In contrast, the colour of the boxes and sticks are irrelevant. But to what extent are animals sensitive to the distinction between functionally relevant and irrelevant features of a means–end task? This paper addresses this question by testing cotton-top tamarins, *Saguinus oedipus oedipus*, on a series of means–end problems.

Much of the recent work in this area has focused on the causal nature of tool use, and to some extent, the functional properties of tools (McGrew 1992). In particular, Visalberghi and colleagues (Visalberghi & Fragaszy 1991; Visalberghi & Limongelli 1994, 1996; Visalberghi et al. 1995) as well as Westergaard et al. (1997) have demonstrated that when confronted with a problem requiring a tool for its solution, some capuchin monkeys can select among the most appropriate objects out of a set. Their selection appears to be based on the functional properties of such objects, such as whether a stick is sufficiently small to fit inside of a tube and whether it is sufficiently rigid to displace a piece of food; not all subjects passed such tests, however, suggesting important individual variation. In addition to using objects as tools, these tests also show some capacity to solve means–end relationships. Specifically, the capuchins appear to understand that a certain kind of object represents the means to a piece of food; the orang-utans and chimpanzees tested on this test showed comparable competence. Matsuzawa's (1996; Tonooka et al. 1997) work with chimpanzees also shows that in selecting a hammer and anvil for opening up hard-cased fruits, chimpanzees appear to search for appropriate materials among the potential range of natural variation. Thus, they select stones with flat surfaces as anvils, and stones or branches as hammers that are narrow at one end (i.e. for holding) and broad at

Correspondence and present address: M. D. Hauser, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA, 02138, U.S.A. (email: hauser@wjh.harvard.edu). C. Botto-Mahan is at the Departamento de Ciencias, Ecologicas Facultad de Ciencias, Universidad de Chile.

the other (i.e. for making contact with the nut placed on the anvil). In a few cases, Matsuzawa observed chimpanzees placing a second stone underneath the anvil to level the smashing surface. Last, and to highlight the fact that such means–end analyses are not restricted to primates, Hunt's (1996) field observations of the New Caledonian crow, reveal that individuals modify blades of grass into shapes designed to extract different kinds of insect prey. In each of these cases, then, individuals solve a means–end relationship task in the context of a significant ecological problem, and show some sensitivity to the relevant functional features of a potential tool.

The early literature on animal learning indicates that in contrast to Old World monkeys and apes, New World monkeys, and in particular the Callitrichidae, have considerably greater difficulty acquiring learning sets, and thus, fail to generalize from particular instances of a problem to more abstract situations (Harlow 1949; Miles & Meyer 1956; Schrier et al. 1965; Warren 1965). In this sense, even if tamarins and marmosets are capable of solving a means–end task, they might not tolerate featural transformations (e.g. colour or shape) from the originally learned stimulus set. None the less, results from recent laboratory experiments (Hauser 1997) indicate that tamarins spontaneously use an object as a means to accessing a piece of food. Furthermore, having learned the solution to this problem with one object, they readily transferred their knowledge to other novel objects, showing some sensitivity to the relevant functional features. In this experiment, for example, their choice of potential objects indicated that they distinguished between features that were relevant to the object's functionality as a tool (e.g. shape) as opposed to features that were irrelevant (e.g. colour). Although promising, these studies were limited by the relatively small number of features manipulated. Furthermore, to probe the generality of the tamarins' capacity, it is important to test them on different means–end tasks involving different kinds of objects. The experiments presented below therefore build on these studies, and the ones cited above, and add to our general understanding of the organization of object knowledge in animals (Hauser et al. 1995; Hauser 1997, 1998; Uller 1997; Hauser & Carey 1998).

METHODS

Subjects

Cotton-top tamarins are small New World monkeys, native to the rainforests of Colombia, inhabiting upper portions of the rainforest canopy (Savage et al. 1996). Like most other New World monkeys, they live in small social groups consisting of a breeding adult pair and their offspring. As a general rule, females give birth to twins, with singletons and triplets being considerably more rare. Their diet comprises fruits, insects, and exudate from the trees (Rylands 1996; Savage et al. 1996).

Subjects were obtained from the New England Regional Primate Research Center, Southborough, Massachusetts, in 1992. All individuals were born in captivity, and at the time of testing, had no prior experimental experience; the experiments presented here were conducted before those described in the Introduction (Hauser 1997). Several subjects, however, were simultaneously being trained with traditional operant techniques (J. Kralik & M. D. Hauser, unpublished data) involving the discrimination of visual stimuli presented on a monitor and food reinforcement for correct presses on a button panel. Thus, subjects were familiar with some aspects of our protocol (see below).

The 10 tamarins we obtained for study had previously been divided into four social groups, three consisting of a breeding adult male and an adult female, and one consisting of a breeding pair and their offspring. Only nine subjects were used in these experiments, four adult males, three adult females, one subadult female and one subadult male; the one adult female that was not run was sick at the beginning of our experiments. For the three groups consisting of a single breeding pair, their cages measured $1.2 \times 1.2 \times 1.8$ m, whereas the family group of four lived in a cage measuring $2.4 \times 2.4 \times 1.8$ m. Each cage was filled with branches from local trees, wooden platforms and nestboxes. Additionally, subjects were provided with ad libitum access to water and were maintained on a diet of Purina monkey chow, peanuts, sunflower seeds, crickets, mealworms, yogurt and fruit. Weights were maintained at approximately 10% less than would arise from ad libitum feeding under captive conditions; the maintained weights more closely approximated those obtained from animals living in the wild (A. Savage, personal communication).

Experimental Design

The procedure employed with the tamarins was a means–end task modelled after Willatts' (1984, 1990) studies of human infants. For the infants, three towels were presented. On the end of one towel was a toy, and to the side of the other two towels were identical toys. To obtain the toy the child had to pull a towel, and in particular, the towel with a toy located on its surface. By the age of 1 year, infants succeed in this task. For the tamarins, subjects sat in a Plexiglas box with body position relatively centred due to the curvature of the rear wall of the chamber (Fig. 1). The front Plexiglas panel had two rectangular openings (4×6 cm) that provided the tamarins access to a light blue tray where the means–end problem was set up. The tray (18×45 cm) was divided into two equal halves by a 3.0-cm high partition, thus setting up a two-alternative forced choice procedure. On each side of the tray was a potentially useful object (means) and a potentially accessible piece of food (end). Subjects were run on one session per day, with no more than 4 days separating consecutive sessions, and a mode of 1 day between sessions. The first problem (On) presented a choice between food located on the surface of a piece of cloth versus food located off or touching the side of a piece of cloth (Fig. 2). The second problem (Connected) presented a choice between food located on a single connected piece of cloth versus food located on a piece of cloth that was separated from a second piece of cloth by a horizontal gap of 1.5 cm (Fig. 3). Half of the subjects started with the On problem and then moved on

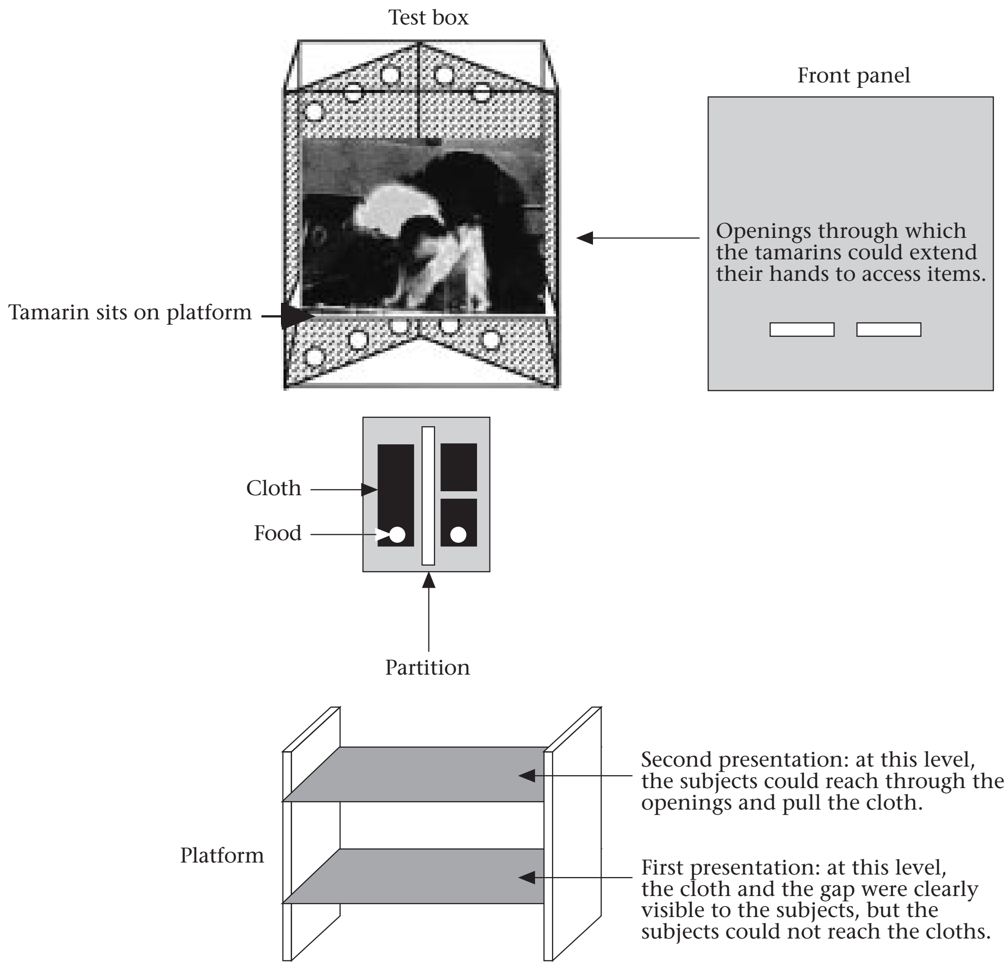

Figure 1. Set-up for testing tamarins. The test chamber provided space for the subjects to sit and reach through one of two openings in the front Plexiglas panel.

to the Connected problem, and the other half started with Connected and then moved to On (Table 1). The logic underlying this subgrouping was to determine whether the tamarins would have greater difficulty solving one means–end task when compared with the other. There are several similarities and differences between the On and Connected problems. For example, in both problems, the food must be on and connected to the cloth, rather than merely touching, for the cloth to serve a useful function in bringing the food closer. In the Connected problem, however, the food must also be on and connected to the particular cloth that when pulled, brings the food closer; pulling a piece of cloth that is separated by a gap from the cloth with food fails to achieve the targeted end. In contrast, for all of the On problems, subjects must simply attend to the position of the food pellet relative to the cloth and select the one with a food pellet that is on and connected to the cloth. In this sense, On is likely to represent a simpler means–end task for the tamarins. If so, subjects should solve the On problem faster than the Connected problem.

For the initial conditions of both problems, subjects were required to pull a piece of dark blue cloth (6 cm wide by 3–11 cm long) to gain access to a flat, white, 45-mg food pellet; in no trial was it possible to access the pellet without using the cloth. In most trials, there was only one correct choice for any given pairing (i.e. one cloth on each side of the tray). Moreover, because subjects were allowed only one choice, they were forced to use a forward planning strategy. This forced choice resulted from the fact that the front panel of the test box only permitted a single selection at a time (i.e. they could only reach one arm through one hole in the partition at a time; after reaching through one hole, the tray was removed before they could reach through the second hole or reach across the partition).

In the On problem, food could only be obtained when the pellet was located on a portion of the cloth, and subjects pulled this cloth rather than the alternative (where food was placed Off the cloth). For the Connected task, the relevant problem for the tamarins was to determine which of the two observable pellets was located on a completely connected piece of cloth as opposed to a piece of cloth separated from another by a gap. If, relative to the tamarin's position, the pellet was located on the piece of cloth opposite the gap, then pulling the reachable cloth would fail to advance the pellet.

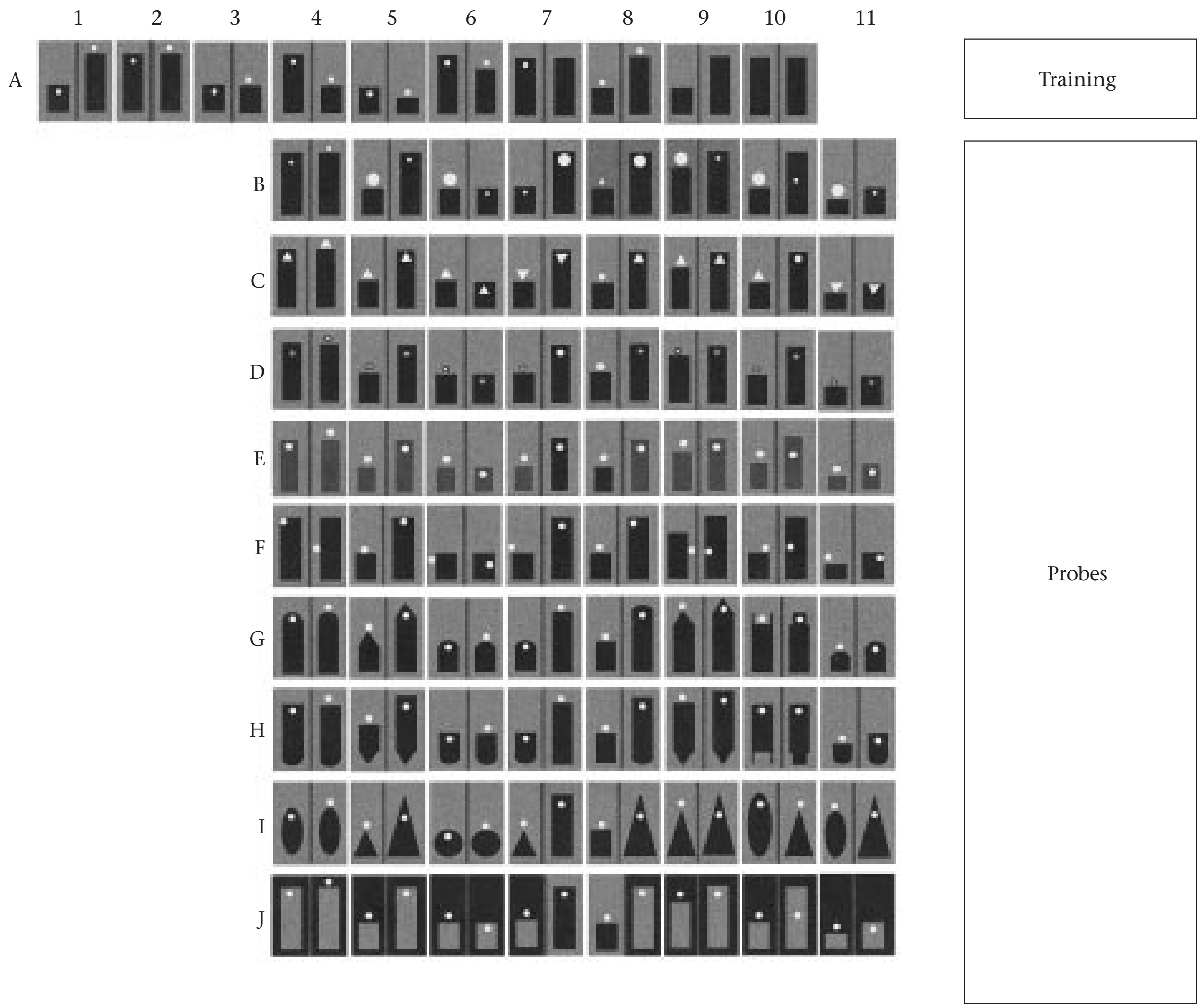

Figure 2. Training test trials (A1–A10) and probes (B–J) for the On problem. Description of each condition is provided in Table 2. All objects are drawn approximately to scale.

The initial training condition (A) for the On test consisted of seven unique trials with only a single correct choice on each trial (A1–A7; Fig. 2). One trial (A8) involved the presentation of two pieces of cloth where the food pellet was unattainable. On two trials (A9, A10) cloths were presented, but without pellets. These latter three trials were run to assess whether tamarins would only pull cloths under conditions when pellets could be accessed, or whether the presence of cloths simply elicited pulling independently of food. Specifically, given the important role that inhibition plays in problem solving tasks (e.g. Diamond 1988; Deacon 1997), we designed these trials to assess the tamarins' capacity for inhibition. In this first condition, each subject received a total of 20 trials per session. Of the 14 trials where food was accessible, half provided food on the right and the other half provided food on the left; presentation was random across trials with the one constraint that there were no more than three left or three right correct (i.e. food accessible) trials in a row. Because food was not accessible for trials A8–A10, accuracies were evaluated only for A1–A7 (i.e. a total of 14 trials per session).

The initial condition for the Connected problem (Fig. 3) involved the presentation of six unique trials (A1–A6) where only one side of the tray yielded a pellet when the cloth was pulled. Conditions A5 and A7, like A8 for the On problem, presented a situation where neither side of the tray provided the means for gaining access to a pellet. Lastly, trials A8 and A9 involved cloths on each side of the tray, but no pellets; the aim of these three trials was, again, to assess the tamarins' capacity for inhibition. A total of 18 trials were run per session. Of the 12 trials where food was accessible, half yielded pellets on the right side and half on the left; as with the On problem, trials were randomized and the same constraint was operative with respect to the number of consecutive correct trials. Accuracies were calculated for trials A1–A6, that is, a total of 12 trials per session. The horizontal gap separating the two disconnected pieces of cloth was held constant at 1.5 cm.

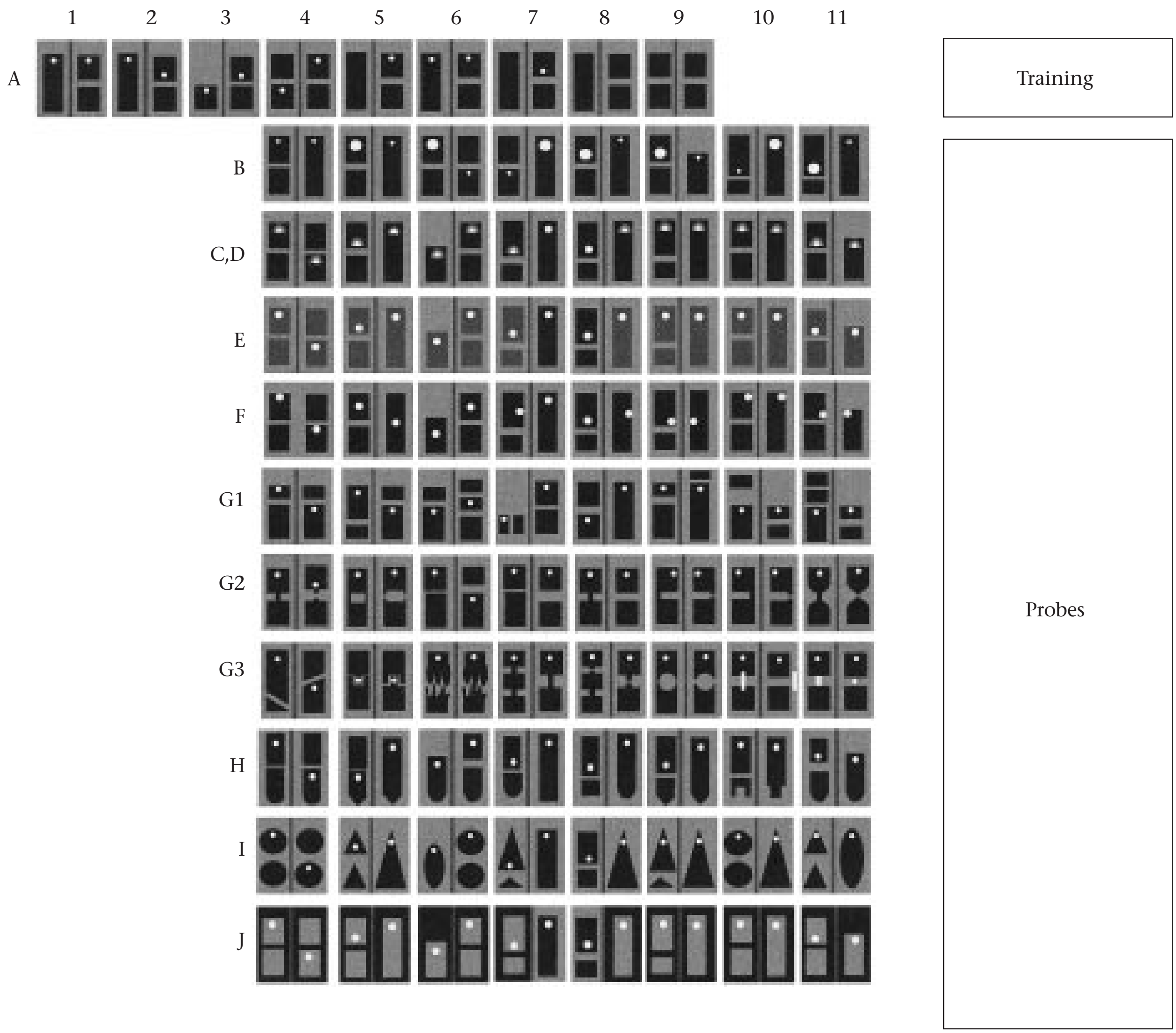

Figure 3. Training test trials (A1–A9) and probes (B–J) for the Connected problem. Description of each condition is provided in Table 2. All objects are drawn approximately to scale.

Table 1. Subjects and test problem received first

Subject (ID)	Group	Sex	Problem received first
Bellugi (B)	L1	F	On
Chomsky (C)	L1	M	On
Fernald (F)	L2	F	On
Locke (L)	L2	M	On
Nagel (N)	P	M	On
Millikan (M)	P	F	Connected
Quine (Q)	P	M	Connected
Churchland (H)	P	F	Connected
Pinker (P)	L3	M	Connected

During the initial training trials, for both the On and the Connected problems, we systematically altered only two features of the presentation: distance of the food and length of the cloth. Then, after the initial conditions for both On and Connected problems, we conducted a series of probe conditions to determine what dimensions of the stimuli were being responded to. In the probe conditions, the following parameters were systematically varied: (1) cloth colour, shape, texture; (2) food type, size, shape, colour, position relative to the cloth; (3) tray colour; (4) gap shape; (5) type of connection between two or more pieces of cloth or other material. Table 2 provides a verbal description of each condition within the On and Connected problems and a classification of featural changes into functionally relevant and irrelevant. If subjects had difficulty with a probe condition, we provided them with further experience (training) before moving on to the next condition (see below).

By systematically altering the parameters in the above list, we would be able to determine the featural changes that the tamarins could tolerate without any effects on performance, as well as the featural changes that significantly affect performance. This design provides an opportunity, therefore, to assess the kinds of features that the

Table 2. Description of featural changes for On and Connected problems

Problem/Condition	Functionality of change?	Featural change
On problem		
Condition		
A	Relevant	Cloth length and distance to food: cloths varied in size and food position varied with respect to its distance from the base of the cloth.
B	Irrelevant	Food size: large piece of food was four times larger than the small piece. In some trials, the large piece was closer to the subject and in other trials it was further away.
C	Irrelevant	Shape of food: from circular to triangular.
D	Irrelevant	Colour of food: from white to red.
E	Irrelevant	Cloth colour: from dark blue to pink.
F	Relevant	Food position On and Off the cloth: the first version of this condition provided only a few relatively easy manipulations of food position. The second version was more difficult and is presented in Fig. 2.
G	Irrelevant	Shape of the upper portion of the cloth (i.e. part furthest from subject).
H	Irrelevant	Shape of the lower portion of the cloth (i.e. part closest to subject).
I	Irrelevant	Overall shape of the cloth.
J	Irrelevant	Reverse colours of tray and cloth: tray is dark blue and cloths are light blue.
Connected problem		
Condition		
A	Relevant	Cloth length and distance to food. Cloths varied in size and food position varied with respect to its distance from the base of the cloth. Gap width (1.5 cm) was held constant.
B	Irrelevant	Food size: the large piece of food was four times larger than the small piece. In some tests, the large piece was closer to the subject and on some trials it was further away.
C,D	Irrelevant	Food shape and colour: change from circular and white to triangular and red.
E	Irrelevant	Cloth colour: change from dark blue to pink.
F	Relevant	Food position of Connected and Disconnected cloth.
G1–G3	Relevant	Changes in position, shape and orientation of the gap between each cloth, from relatively simple (G1) to relatively complex changes (G3), including reduction in gap width to 0.8 cm.
H	Irrelevant	Shape of the lower portion of the cloth (i.e. part closest to subject).
I	Irrelevant	Overall shape of the cloth.
J	Irrelevant	Reverse colours of tray and cloth: tray is dark blue and cloths are light blue.
K	Relevant	Changes in the type of connection and changes in the material used to access the food.

tamarins perceive as relevant or irrelevant to the means–end task, and the extent to which their competence at solving the On and Connected problems goes beyond the limited set of stimuli in the initial condition (see Figs 2 and 3). Specifically, we manipulated features that were functionally irrelevant to the task (e.g. colour of the cloth and food) as well as functionally relevant (e.g. type of connection between the cloths) to assess whether the tamarins were sensitive to the functionality of the cloth as a means for obtaining food (i.e. ends; see Table 2). Both means–end tasks were set up in such a way that the relationship between performance and the functional relevance of the feature could be explored. Thus, imagine a 2×2 table with the functionality of the feature (irrelevant versus relevant) on the Y axis (rows) and performance (stays at criterion versus drops below criterion) on the X axis (columns). If performance stays above criterion for changes to irrelevant and relevant features, then the tamarins have acquired a general, abstract solution to the problem. If performance drops in response to changes in irrelevant or relevant features, then the tamarins' ability to solve these problems is limited, and depends on these features being held constant. For example, if the tamarins use cloth colour to solve the means–end task, then changing colour should cause a drop in performance. Similarly, if the tamarins solve the problem by using a particular type of connection between two pieces of cloth, then changing this connection (e.g. from a piece of cloth to a wooden dowel; condition K below) should cause a drop in performance.

Test sessions proceeded as follows. First, the subject was removed from its home cage and transported in a Plexiglas box to a visually isolated testing room. Subjects had been trained to move out of their home cages and into the transport boxes without handling; they readily make such moves several times a day and do so in the absence of behavioural signs of stress. Second, subjects were transferred to the test box (Fig. 1). Third, a two-tiered stand was placed in view. An experimenter then prepared the tray for each trial, out of view of the test subject. Fourth, the prepared tray was presented to the subject on the first tier of the stand. Here, subjects could readily see the positioning of the cloths and pellets, but could not gain access to them; the tray was presented and left at the first tier until the subject looked at the set-up for at least three consecutive seconds. During this presentation, the experimenter's hand was placed at the midpoint of the tray (i.e. the point where the middle partition connected with the back wall of the tray), making sure that the tamarin attended to the display. Fifth, the tray was moved to the second tier of the stand and the subject was given one chance to pull a piece of cloth. The experimenter never looked directly at the tamarin; the experimenter monitored the tamarin's response out of the corner of his or her eye, while looking straight down at the floor. The tray was removed after the tamarins pulled one cloth, or if they touched the cloth on the incorrect side, or if they failed to pull a cloth within 10 s. This procedure imposed a cost for either making a mistake or for failing to pull within a short window of opportunity. As soon as a choice had been made, the tray was reloaded and the next trial initiated. The order of trials within each condition was derived from a random number generator for each session.

Prior to moving on to subsequent conditions, subjects were allowed to make only one mistake out of the total number of trials in conditions A and B, and were required to maintain this level of accuracy (92–100%) on five consecutive sessions. We demanded such high accuracies at this early stage to make sure that performance was stable in the transition to subsequent probe conditions. This was important because we wanted to evaluate performance on the first trials of each new condition, in addition to assessing the rapidity with which subjects learned new conditions (i.e. potential featural transformations of the initial problem). Thus, if a subject's performance on the original condition was relatively low, we would be in a weak position with regard to evaluating their performance on first trials of the second condition. Failure on the first trial of a new condition may indicate that they are incapable of generalizing to new conditions (i.e. a constraint imposed by their species-typical cognitive ability) or may indicate that performance is unstable due to motivational or attentional factors that fluctuate somewhat unpredictably from session to session. We also repeated the first three trials from condition A within each new condition; these were randomly presented within a session. This provided us with a baseline performance level for judging success in each new condition. Thus, if subjects performed accurately on the old trials, but failed on the new trials, we could more confidently conclude that the imposed transformations were responsible for the observed performance problems.

Following conditions A and B, we required subjects to attain an accuracy of at least 75% on two consecutive sessions before moving on to a new condition. We changed the criterion because as the experiment advanced, some conditions proved extremely difficult for the tamarins, even after training; this criterion is, however, consistent with many primate experiments using operant procedures (see for example Schrier et al. 1965). Consequently, more stringent criteria may have prevented some subjects from moving through all of the conditions in the experiment. As described below, some individuals failed to pass a condition even with a criterion of 75%. Finally, if subjects were below criterion and not improving over two sessions, we modified specific properties of the problem presented. For example, in some of the Connected conditions, the width of the gap and the connection between the two pieces of cloth were increased by 1.0 cm. Furthermore, if subjects were having particular problems with a given set of tests, but not others, we eliminated the easy tests and increased the number of trials per session of the difficult ones. All of these modifications were designed to increase the subject's opportunity to learn a particular aspect of the problem and then move on to a new generalization phase. In summary, then, we were looking for both spontaneous transfer from a learning set as well as the capacity to learn a problem through training if subjects failed to transfer.

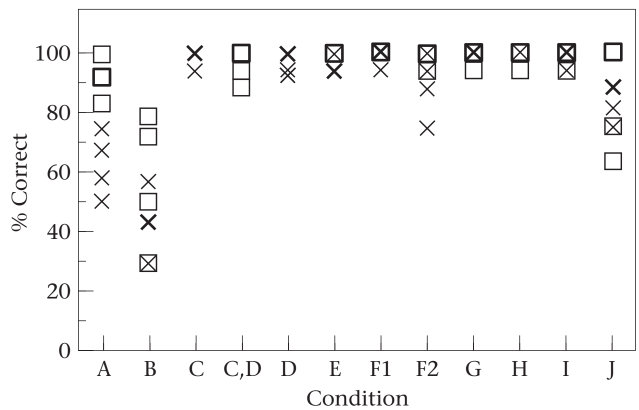

Figure 4. Scores (% correct) on first session for each condition of the On problem for subjects that started with On (×) and those that started with Connected (□). Thicker lines represent multiple data points (i.e. >1). Subjects that started with the Connected problem received a combination of conditions C and D, rather than each separately.

RESULTS

On Problem

For condition A, performance on the first session was consistently lower for subjects that started with the On problem than for subjects that started with the Connected problem (Mann–Whitney *U* test: $U=0$, $P<0.01$; Fig. 4). Subjects starting with the Connected problem had more experience, as well as more varied experience, than those starting with the On problem. As revealed below, however, the amount of time on a particular problem did not appear to account for the variation. For all other conditions, there were no statistically significant differences ($P>0.05$) between groups in the percentage of correct responses on the first session (Fig. 4). Thus, type of experience, rather than the amount of time on the Connected problem seemed to benefit performance on the On problem.

Restricting the analyses to first trial performance only, and breaking down the data by subject and condition, it is clear that most subjects were correct on the very first trial of most conditions (Table 3). Specifically, subjects were correct on 76 out of 85 possible first trials (89%). Although subjects starting with Connected performed better on the first trial of condition A of On than did subjects starting with On, by condition B, this difference disappeared.

In condition A, food was unattainable on three trials (A8–A10). For these trials, two out of the five subjects starting with the On problem consistently inhibited their response to pull the cloth. Specifically, over five consecutive sessions, they pulled a cloth on no more than four out of 18 trials. Two out of the four subjects starting with the Connected problem also inhibited their response to pull the cloth. The other subjects occasionally refrained from pulling the cloths, even though they were never rewarded for such behaviour.

Conditions A and B were, relative to the other conditions, more difficult (Fig. 4). Having passed conditions A and B (see below), subjects consistently performed well on the first session of each new condition (as well as the first trial of each condition; see Table 3), showing that changes in the colour, shape and relative positions of the food and cloth had little effect on performance; subjects readily generalized without training. The fact that subjects performed poorly on condition B, after reaching criterion on condition A, is significant because the difference between these conditions is small at one level, and apparently large at another. Specifically, the only difference between A and B (see Fig. 2) is that in some trials, subjects were required to select a small accessible piece of food over an inaccessible, but substantially larger piece of food. This posed a problem because subjects appeared to be highly motivated to obtain the large piece of food. As discussed below, subjects required several sessions to overcome this motivation.

In addition to obtaining lower scores on the first session, subjects starting with the On problem also took longer to reach criterion for conditions A and B (Fig. 5),

Table 3. Performance on the first trial of each condition of the On problem

	Subjects starting with On					Subjects starting with Connected				
Condition	N	B	L	C	F	M	P	Q	H	% Correct
A	I	I	C	C	C	C	C	C	C	78
B*	I	I	C	C	I	I	C	I	I	33
C or C,D	C	C	C	C	C	C	C	C	C	100
D	C	C	C	C	C	C	C	C	C	100
E	C	C	C	C	C	C	C	C	C	100
F	C	C	C	C	C	C	C	C	C	100
G	C	C	C	C	C	C	C	C	C	100
H	C	C	C	C	C	C	C	C	C	100
I	C	C	C	C	C	C	C	C	C	100
J	C	C	C	C	I	C	C	C	C	89

C: Correct choice; I: incorrect choice.
*Conditions that proved difficult for the tamarins, as revealed by the number of sessions required to reach criterion (see Fig. 5).

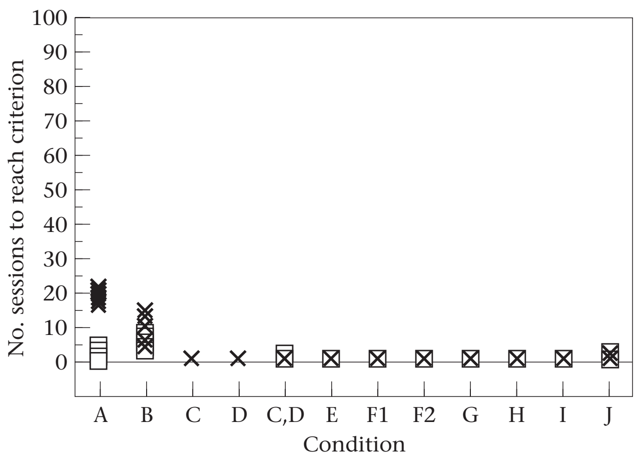

Figure 5. Number of sessions to reach criterion for each condition of the On problem for subjects that started with On (×) and those that started with Connected (□). Thicker lines represent multiple data points (i.e. >1).

but only for condition A was this difference statistically significant (Mann–Whitney *U* test: $U=0$, $P<0.01$). Again, performance on condition B is of considerable interest. As the data reveal, although subjects moved to condition B only after scoring a minimum of 13 of 14 on five consecutive sessions of condition A, most subjects experienced a significant drop in performance on condition B, with scores far below chance on the first few sessions; some individuals took an additional 10 or more sessions to reach criterion (Fig. 6). For several subjects, performance appeared to rebound from below chance to criterion in just a few sessions. Overall, having passed condition B, subjects in both groups passed subsequent conditions on the first or second session, and in most cases, were correct on the first trial of the session (Table 3). Thus, they readily transferred their knowledge from conditions A and B to the subsequent conditions, illustrating that these featural transformations (e.g. cloth and food shape, colour and size; see Table 2) had little to no effect on the tamarins' success in the means–end problem. In fact, although our measure of reaction time to pull the cloth was only assessed in terms of 'immediate' or 'delayed' designations, it was our impression that subjects pulled as fast on each new condition as on the previous conditions, with no noticeable change in reaction time across sessions. In fact, most subjects appeared to make a decision when the tray was placed on the first tier of the stand, and then simply lined up to one side once the tray was in position on the second tier.

Connected Problem

In contrast to results from the On problem, there were no statistically significant differences between groups on the proportion of correct responses in any of the Connected conditions (Fig. 7). In general, scores on the first session of each Connected condition were lower than for those of the On condition. Subjects showed the same difficulty with condition B of Connected as they did for condition B of On: after reaching criterion on condition A, scores dropped significantly for condition B even though the only difference was in the relative size of food. Furthermore, the same subjects that inhibited cloth pulling for the On test trials, did so for trials A7–A9 of the Connected test trials.

Restricting our analyses to first trial performance only, it is clear that most subjects were correct on the very first trial in many of the conditions (Table 4), with the exception of condition K (functionally relevant: change in type of connection and material used to access food). Specifically, out of 108 possible first trials, 97 were correct (90%) and there were only minor differences between test groups. Note that although there was a positive relationship between the overall difficulty of the condition with regard to session scores and first trial performance, first trial performance did not completely account for performance on a given condition. For example, although conditions G2 and G3 (functionally relevant: change in position, shape and orientation of gaps between cloths) were clearly more difficult for the tamarins, and required several training sessions before a criterion score was attained, a high proportion of the subjects pulled the correct cloth on the first trial.

Paralleling results on scores for the first session of each Connected condition, there were also no statistically significant test group differences in the number of sessions to criterion (Fig. 8). Overall, subjects had considerably more difficulty with the featural transformations in the Connected problem set than they did in the On problem set. There were, however, similarities between the On and Connected problems. For example, subjects required a relatively large number of sessions to pass condition B (Fig. 8), the large food condition; and as demonstrated for condition B of the On problem, several subjects showed a rapid change in performance from significantly below chance to criterion (Fig. 9). Subjects also had little difficulty with changes in cloth or food colour, shape or position, as evidenced by their first trial and session accuracies. The most significant effects on performance in the Connected problem were created by changes in the shape of the gap separating each piece of cloth (conditions G2 and G3), as well as in changes in the type of connection (condition K). Condition G2 was extremely difficult for the tamarins (Fig. 8), and for some individuals, we were forced to simplify the set-up by increasing the gap distance as well as the thickness of the connection (see Methods; Table 2). Having been trained to pass such simplified versions of G2 tests, we then returned to the original tests involving narrower gaps and thinner connections between the pieces of cloth. Ultimately, subjects passed this condition, but required considerable training. Condition G3 represented a second round of changes to gap shape, and the subjects had somewhat less difficulty in this condition than in G2 (Fig. 8).

Overall, then, it appeared that the Connected problem proved more difficult for the tamarins than did the On problem. Furthermore, if one matches up similar featural transformations (e.g. change in cloth colour, condition E) for each of the two means–end problems, there were no

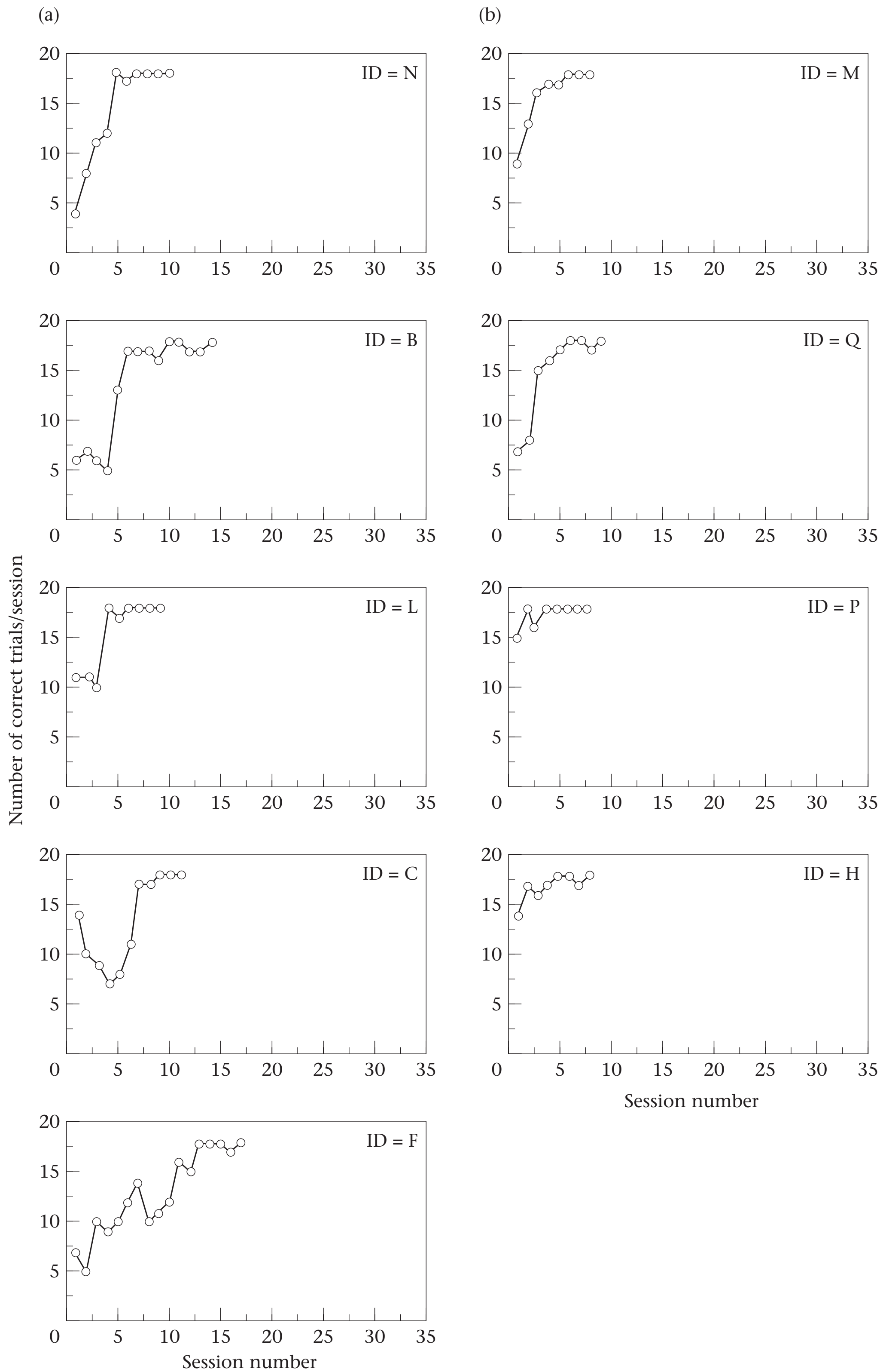

Figure 6. Performance on condition B (changes in food size) of the On problem by subjects starting with (a) the On problem and (b) the Connected problem. N=18 trials/session.

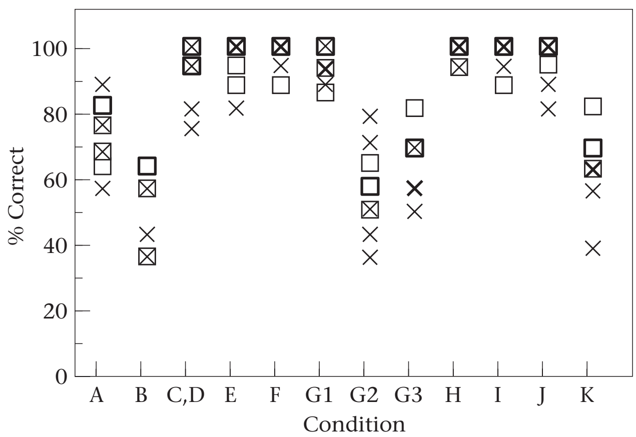

Figure 7. Scores (% correct) on first session for each condition of the Connected problem for subjects that started with On (×) and those that started with Connected (□). Thicker lines represent multiple data points (i.e. >1). Conditions C and D were combined.

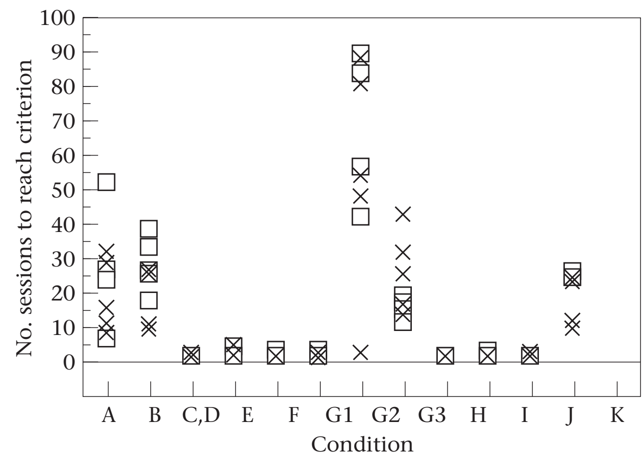

Figure 8. Number of sessions to reach criterion for each condition of the Connected problem for subjects that started with the On problem (×) and those that started with the Connected problem (□). Thicker lines represent multiple data points (i.e. >1).

test group differences in the number of sessions to criterion, except that subjects required more sessions to pass conditions A and B of Connected than conditions A and B of On (Mann–Whitney *U* test: $P<0.01$).

Condition K of the Connected problem represented a different kind of test from the others because it set up a new problem: having learned to make a distinction between pieces of cloth that were or were not connected, subjects were now required to make a distinction between cloths with a strong versus weak connection, or with material that differed fundamentally from the original cloth (Fig. 10). Subjects starting with On performed significantly better on the first session than those starting with Connected (Mann–Whitney *U* test: $U=1.0$, $P<0.02$; Table 4). There was, however, no difference between test groups on first trial performance, with two of the five On-group subjects missing the first trial and two of the four Connected-group subjects missing the first trial (Table 4).

There was considerable variation between individuals with regard to performance on individual tests within condition K (Fig. 11). Some subjects passed certain tests on their first try but completely failed on others, and some subjects never passed some tests; we stopped running subjects after 24 sessions. The failure to pass some tests occurred even though we attempted to facilitate the problem by, for example, increasing gap size or increasing the width of the connection by 1.0 cm and reducing the trials within a session to only those tests that proved difficult. Only one subject (N) passed all of the tests. With the exception of subject C, however, all of the other individuals passed at least six out of the eight tests in condition K.

Finale: Mixed conditions of On and Connected

To determine the extent of the tamarin's comprehension of the On and Connected problems, we ran one

Table 4. Performance on the first trial of each condition of the Connected problem

	Subjects starting with On					Subjects starting with Connected				
Condition	N	B	L	C	F	M	P	Q	H	% Correct
A	C	C	C	C	C	C	C	C	C	100
B*	I	C	C	I	C	C	C	C	C	78
C or C,D	C	C	C	C	C	C	C	C	C	100
E	C	C	C	C	C	C	C	C	C	100
F	C	C	C	C	C	C	C	C	C	100
G1	I	C	C	C	C	C	I	C	C	78
G2*	I	C	I	C	C	C	C	C	C	78
G3*	I	C	I	C	C	C	C	C	C	78
H	C	C	C	C	I	C	C	C	C	89
I	C	C	C	C	C	C	C	C	C	100
J	C	C	C	C	C	C	C	C	C	100
K	C	I	C	C	I	C	C	I	I	56

C: Correct choice; I: incorrect choice.
*Conditions that proved difficult for the tamarins, as revealed by the number of sessions required to reach criterion (see Fig. 8).

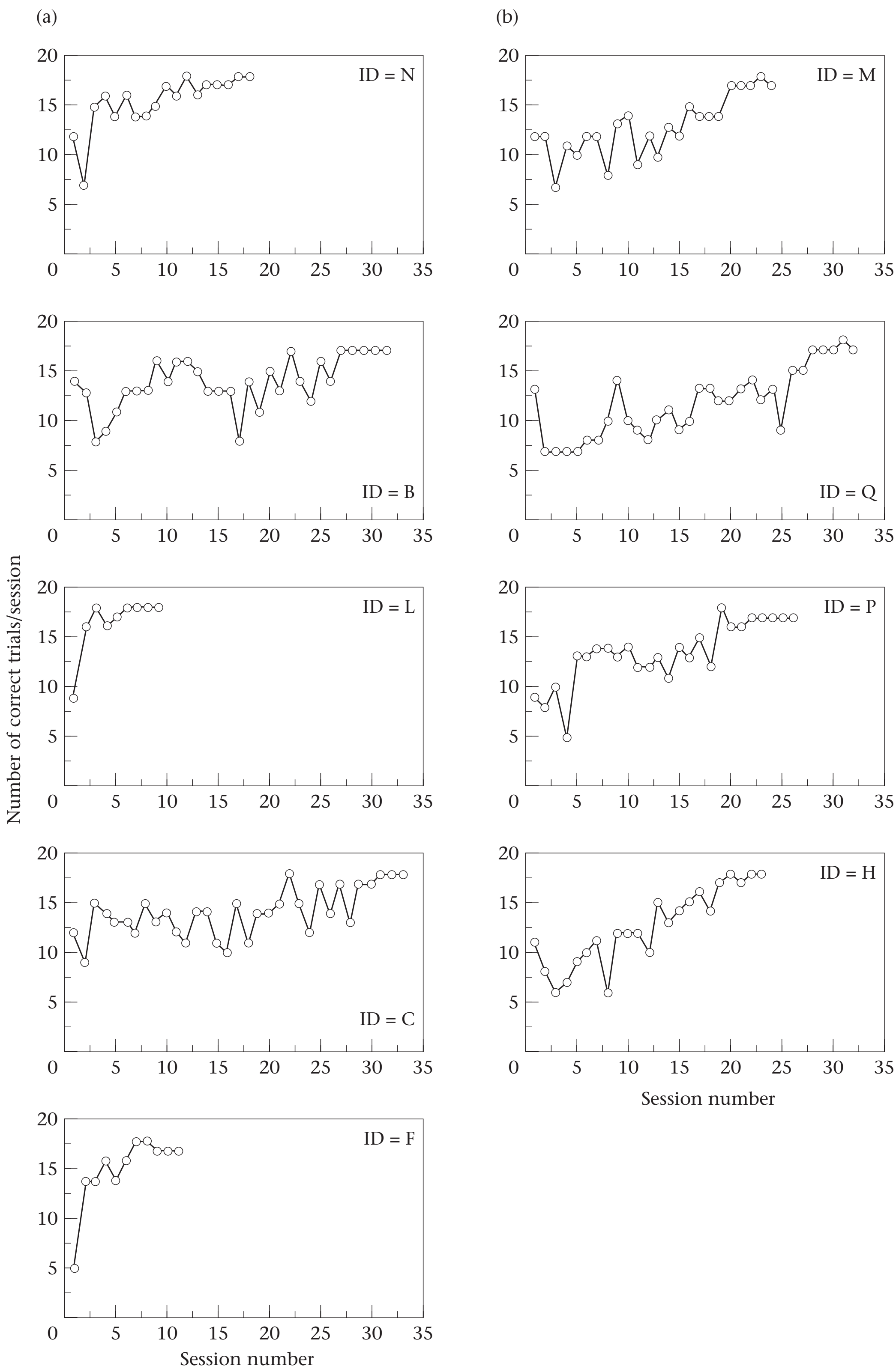

Figure 9. Performance on condition B (changes in food size) of the Connected problem by subjects starting with (a) the On problem and (b) the Connected problem. N=18 trials/session.

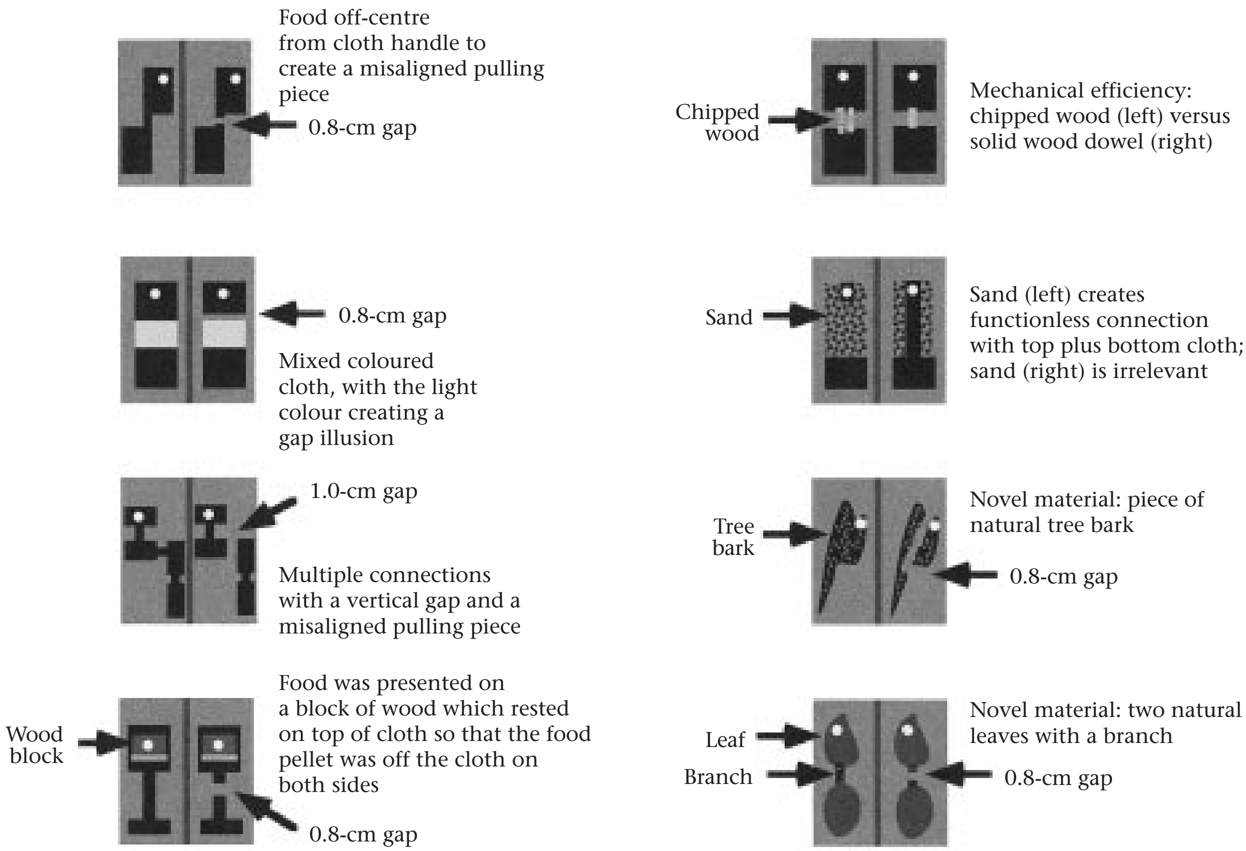

Figure 10. Different tests for condition K of the Connected problem. Tests focused on differences in gap position, type of connection, and the material properties of the objects presented.

final condition: a mixture of On and Connected tests from a sampling of previous trials. More specifically, we extracted all of the most difficult tests within the various conditions (i.e. trials with early failures in prior test sessions) and combined them into a final condition, representing 22 trials (Fig. 12). Each subject was run four times on this condition, with sessions starting immediately after they finished the last On or Connected condition. Three out of five of the On-group subjects chose correctly on the first trial of session 1, whereas three of four of the Connected-group subjects chose correctly on the first trial of session 1 (Fig. 13). Four subjects obtained accuracy scores of 75% or better on the first day, whereas the others scored below criterion. By the second day, there was general improvement that was maintained until day 4. Four subjects ended this condition with accuracies of 80% or better, whereas the other five ended with accuracies of 60–75%.

DISCUSSION

In the Introduction to this paper, we suggested that few studies have been able to specify, systematically, whether animals that solve means–end tasks do so by discriminating functionally relevant and irrelevant features of the task. Our specific aim was to determine whether the cotton-top tamarin could solve a means–end task, attending to the functional features of the problem, disregarding irrelevant features. Results from our experiments show that the cotton-top tamarin can indeed solve means–end tasks and do so by attending to the functional features of the problem (also see Hauser 1997). We now summarize the key findings from our experiments and discuss how they relate to other experimental procedures and findings.

Subjects starting with the On problem finished all conditions and moved to the Connected problem before any of the subjects starting with the Connected problem had finished their conditions. None the less, there was only one significant effect of initial problem (i.e. On versus Connected) on performance: all subjects starting with the Connected problem reached criterion on condition A of the On problem more rapidly than subjects starting with the On problem. These patterns suggest that the tamarins may have developed a general learning set in the Connected problem, which allowed them to generalize to the On problem.

Having reached criterion on condition A, all subsequent conditions except B were designed to assess whether subjects had learned about the functionally relevant and irrelevant features of the means–end task. Thus, for example, the first condition of On could be partially solved by looking for something white (i.e. the food pellet) on a dark blue background (i.e. the cloth).

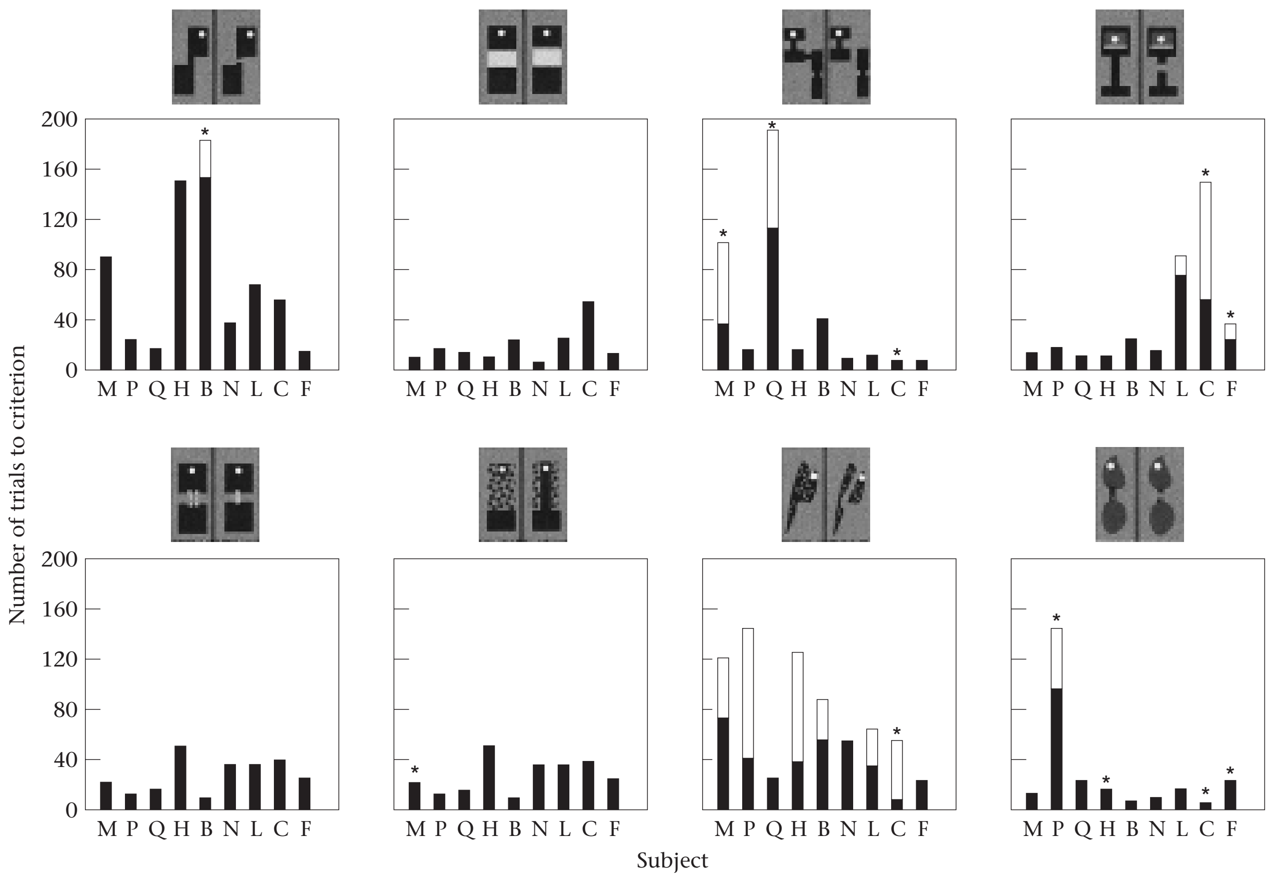

Figure 11. Performance on each test within the K condition. Above each separate histogram is an illustration of the test (see Fig. 10 for description). ■: Scores on the original test. □: Scores on tests that were modified (e.g. by increasing gap distance and the width of the connection). An asterisk above a histogram indicates that the subject never passed the test.

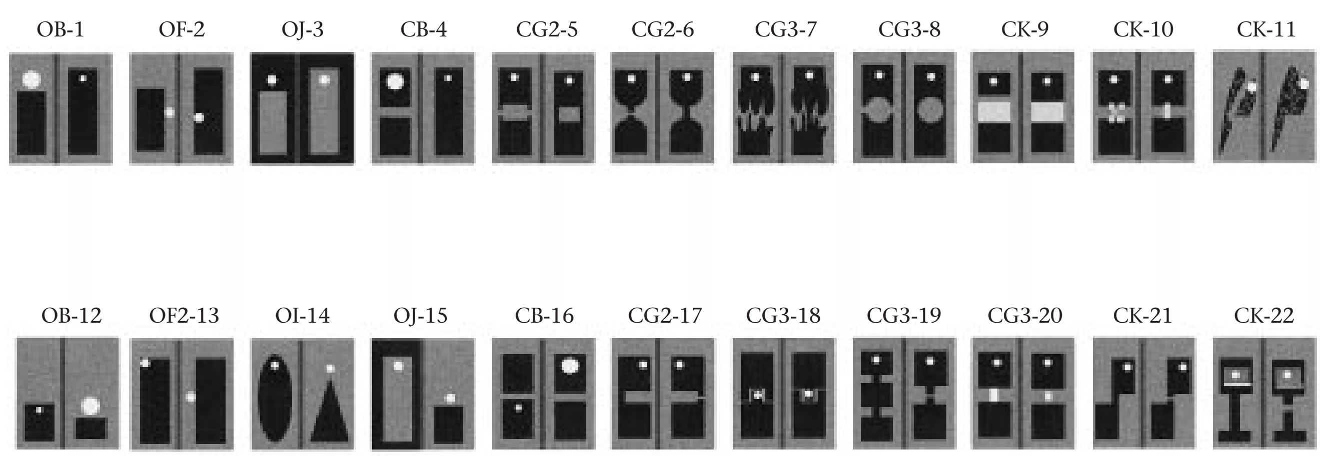

Figure 12. Tests for the final condition. Each trial represents one of the more difficult tests from an earlier On or Connected condition. Numbers associated with each test refer to corresponding trials presented in Figs 2, 3 and 10.

This algorithm would work every time in the first condition but would only show that the tamarins had learned something highly specific and functionally irrelevant from this condition. Similarly, for condition A of Connected, subjects could potentially solve each trial by looking for the side with a light blue strip (corresponding to the colour of the tray) in front of the pellet, and then choosing the cloth on the alternative side. Although in this case the light blue strip corresponded to the gap between each piece of cloth; if subjects were merely responding to colour differences, they would not be discriminating the relevant functional properties of the means–end task. Subsequent conditions were thus designed to determine the extent to which subjects were discriminating the relevant functional properties for solving both means–end tasks.

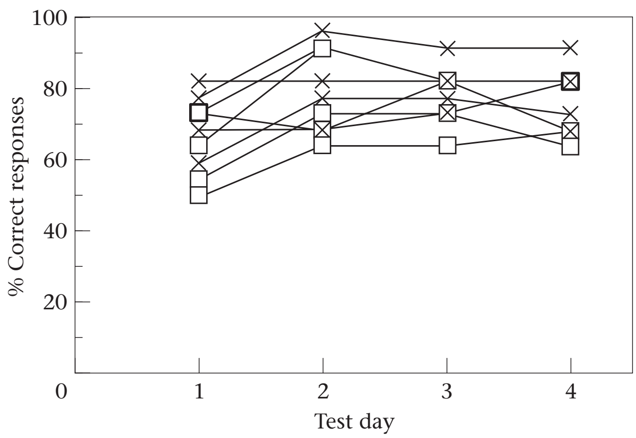

Figure 13. Performance (% correct responses) over four sessions (test days) on the final mixed-condition test. All subjects are presented, with symbols corresponding to the test they started on (×: subjects starting with the On problem; □: subjects starting with the Connected problem).

Leaving aside condition B, the tamarins' performance in the On problem allows us to make two points. First, changes in the colour or shape of the cloth and food, cloth size, relative position of the food to the cloth and food type, had no significant effect on performance. We interpret this pattern as evidence that subjects learned to attend to the functional properties of the problem at an abstract level, tolerating all featural transformations. Second, because of their extremely high accuracies, and the fact that when they made mistakes, these were not more likely to occur at the beginning of the session, performance in the On problem cannot be accounted for by training effects within the probe conditions. The results of the Connected problem were comparable to those of the On problem, with the exception of conditions G2, G3 and K. In parallel with our interpretation of the On problem, we can also make two points. First, subjects tolerated several featural transformations showing that they essentially ignored functionally irrelevant features. Second, for most of the conditions, subjects were correct on the first trial and performed with high accuracies on the first session. Therefore, training cannot account for their performance in these probe conditions. In the three probe conditions that proved difficult, however, training clearly contributed to the subject's performance. Results from these particular transformations illustrate that the tamarins acquired a more limited understanding of Connectedness. In particular, although the tamarins were attending to the functionally relevant features of connection shape, gap shape and type of connection, their responses were nevertheless guided by more specific factors such as gap orientation (e.g. all of the initial training stimuli involved broad, horizontally positioned gaps, whereas G2 and G3 presented narrow, vertically positioned gaps), thereby limiting their capacity to generalize to some novel stimuli. The tamarins therefore required explicit training in these three probe conditions in order to attend to the more appropriate features of the Connected problem.

Having completed all conditions from On and Connected, each subject was then presented with a combination of the most difficult tests within the On and Connected conditions. In general, subjects performed above chance on the first session of this final mixed condition, and by the fourth session, six out of nine subjects obtained accuracies of 75% or better. Thus, the results from this condition suggest that subjects did not simply memorize separate, specific algorithms for solving each individual condition. Rather, they demonstrated a fairly robust understanding of functionally relevant features and a general understanding of the means–end tasks.

Summarizing, then, changes in food colour, food type, cloth colour, local and global shapes of the cloth, cloth size, and the position of the food relative to the cloth had little effect on performance for either the On or Connected problem. It therefore appears that the learned representations or general learning sets (Harlow 1949; Miles 1965) acquired by the tamarins for the On and Connected problems were largely based on the functionality of the cloth. In the On problem we obtained no evidence that the tamarins responded to any other property other than functionality. In the Connected problem we obtained some evidence that the tamarins had acquired a more rigid, less general learning set for the cloth gap or connection. Note, however, that even in the Connected problem, some subjects readily generalized to changes in gap shape, object material (e.g. leaves) and the efficiency of the connection (e.g. chipped wood versus a wooden dowel), showing that they were attending to the functionally relevant features of the task at a relatively abstract level. Furthermore, those animals that had difficulty with such transformations, nevertheless were attending to the functional properties, but at a more specific level. These animals then received training on the difficult transformations, ultimately moved on, and when retested in the final mixed condition, showed that they had maintained their general understanding of the Connected problem.

Condition B (change in food size) of the On and the Connected problem was designed to assess whether tamarins could inhibit their response to obtain a presumably more desirable but inaccessible food for a less desirable but accessible food. Specifically, by pitting a large piece of unattainable food against a small piece of attainable food, we put the tamarins in a situation that required them to inhibit what we predicted would be a highly motivated response. If they recognized which features of the task played a functional role in the problem, then they would not attempt to obtain target items that were off or disconnected (unattainable). As our results show, almost all subjects performed well below chance on the first few sessions of condition B. For several subjects, however, performance increased rapidly, and within a few sessions, was well above chance. Thus, the tamarins were capable of inhibiting attempts to attain the larger piece of food, although for some individuals, this proved exceedingly difficult (i.e. 20–35 sessions to reach criterion).

The change in performance between conditions A and B is all the more impressive when one considers the fact that subjects moved on to condition B only after performing with near perfection on condition A for five consecutive sessions; and the only difference between conditions A and B was food size. Thus, although subjects appeared to solve the means–end task, the apparent possibility of attaining a larger piece of food was difficult to overcome. This difficulty can be interpreted in several ways. For example, the subjects' ability to solve the On or Connected problems may have been restricted to the context of small food size. That is, changing to larger food changed the problem, thereby causing a decline in performance. Alternatively, the subjects' poor performance on condition B could be evidence for an interaction between problem-solving ability and motivational state. Specifically, the tamarin's capacity to solve a problem is restricted to a particular motivational level. When motivational state is changed by increasing desirability of the food reward, one effectively blocks problem-solving knowledge, perhaps by shifting the focus of attention away from the critical factors required to solve the task. None the less, based on the fact that subjects ultimately passed condition B, and were able to refrain from pulling cloths when the food was inaccessible for both alternatives under condition A of On and Connected, we conclude that tamarins have some capacity to inhibit responses even when there is a strong motivational drive towards an inappropriate response.

The capacity to inhibit action is critical to problem solving, and appears in many tasks that have been presented to nonhuman animals and human children (reviewed in Diamond 1988; Deacon 1997). For example, in studies by Boysen (1996), two highly trained chimpanzees were presented with a centrally placed tray. On each side of the tray was a well. On each trial, one well had more food than the other. One chimpanzee played the role of selector, the other the role of receiver. The selector's task was to point to one well of food. The amount of food in the selected well went to the receiver and the food in the remaining well went to the selector. Assuming selfish behaviour on the part of the selector, the game is to pick the well that you do not want; said differently, inhibit pointing to the desired well. Over the course of many trials, with roles changing, neither chimpanzee solved the problem. They consistently picked the larger quantity of food and thus, obtained the smaller amount as a reward. In essence, neither chimpanzee was capable of inhibiting their desire for the larger food quantity. However, one of the chimpanzees had previously been taught the Arabic numerals for object quantities, and consequently, passed the test when numbers were used instead of the food. In other words, when the motivational pull of seeing the large quantity of food was detached in space and time, this one chimpanzee performed successfully.

What these studies show is that an important impediment to problem solving is the inability to inhibit particular actions. Insights into this problem come from comparative neuroanatomical studies revealing that the dorsolateral prefrontal cortex plays a critical role in inhibition (Goldman-Rakic 1987; Yamatani et al. 1990; Deacon 1997). In particular, developmental studies reveal that up until the age of ca. 9 months, human infants have severe difficulties inhibiting reaching responses (Diamond 1988). In parallel with such motor patterns is an ontogenetic change in the prefrontal cortex, which remains relatively immature until the age of 9 months. Similarly, when rhesus monkeys are tested in inhibition experiments that match those conducted with human infants, they show the same deficits when they are between the ages of 2–4 months or if they are adults with the dorsolateral prefrontal cortex lesioned.

Relative to humans over the age of 9 months, apes, most Old World monkeys (e.g. macaques, baboons), and some New World monkeys (e.g. capuchins), cotton-top tamarins have a smaller brain size:body weight ratio (Stephan et al. 1981; Sawaguchi & Kudo 1990; Barton & Dunbar 1997). At present, however, little is known about the prefrontal cortex of tamarins. What is clear is that in an inhibition task that 1-year-old human infants and adult rhesus monkeys pass, adult cotton-top tamarins fail (Santos et al., in press). What these results suggest is that there are important differences in the connectivity of the prefrontal cortex of adult monkey species and these differences need to be explored (Goldman-Rakic 1990; Robin & Holyoak 1995; Deacon 1997). Here then, as in many early neuroethological studies such as bird song (Konishi 1985; Marler & Nelson 1992), barn owl sound localization (Konishi 1993) and bat sonar (Suga 1988; Simmons 1989), the results from condition B and related tests of inhibition demonstrate that behaviour provides important insights into the underlying neural mechanisms.

In some ways, our experiments are comparable to those carried out by Visalberghi and colleagues (Visalberghi & Limongelli 1994; Visalberghi et al. 1995) to explore the capuchin monkey's understanding of cause–effect relations. In most of their tasks, capuchins were required to use an object to gain access to food. Having acquired such skills, these individuals served as demonstrators for naïve subjects that were given the opportunity to learn by imitation. In one experiment, capuchins were presented with a set of potential objects to extract a piece of food located inside of a tube. To be effective, some of the objects required modification. For example, given three short sticks, the capuchins had to place at least two sticks inside the tube, one behind the other, in order to displace the food reward. A subset of the capuchins tested solved this problem, although there was no evidence of imitative learning by naïve individuals. Some individuals experienced grave difficulties in solving the task, attempting, for example, to place fat sticks inside thin tubes. None the less, because several subjects were able to use a variety of objects to solve this problem, Visalberghi and colleagues concluded that capuchins have some understanding of cause–effect relations.

Our tamarin studies differ from those on other primates, and especially capuchins, in two significant ways. First, capuchins, like the other primates where tool use has been studied, are highly dexterous and there are numerous reports from captivity and the wild indicating that they spontaneously use tools (Anderson 1990;

Visalberghi 1990; Westergaard et al. 1997; see McGrew 1992). Tamarins, in contrast, are relatively less dexterous and have never been observed using or manufacturing tools, and infrequently manipulate objects with the exception of food. Second, our experiments were explicitly designed to show whether tamarins distinguish between functionally relevant and irrelevant features of the object used in the means–end task. For example, both here and in our other studies (Hauser 1997), we have shown that tamarins consider colour to be an irrelevant feature of the object's functionality. In contrast, for at least some parts of the object, the shape is highly relevant to functionality. Interestingly, in a series of ongoing experiments with tamarins and rhesus macaques, *Macaca mulatta*, we have some evidence that in evaluating whether something is or is not food, individuals attend to colour as a significantly more salient feature than shape (L. Santos, M. D. Hauser & E. Spelke, unpublished data). If this is confirmed, it suggests that in distinguishing among objects, one feature may be relevant to one class of objects but irrelevant to another (Keil 1994). In the example just given, colour is an irrelevant feature of artefacts, but a highly relevant feature of food.

In conclusion, our experiments allow us to draw four conclusions. First, cotton-top tamarins readily solve means–end tasks. Second, in solving such tasks, the tamarins attend to the functionally relevant features. Featural transformations that do not affect an object's functionality were readily tolerated with regard to performance in the means–end task. This shows that tamarins can discriminate between objects that show signs of good design and those that do not, and can use this knowledge to select an appropriate object for solving a problem. Third, some featural transformations that were clearly relevant to the functionality of the task proved difficult for the tamarins and required explicit training. Thus, although the tamarins appear to have a greater capacity for creating learning sets than revealed by some early experiments on the closely related marmoset (Miles & Meyer 1956; see also Rumbaugh et al. 1996), there are limitations. At present, it is not clear whether these limitations arise due to their relatively poor dexterity, to problems in making fine-grained perceptual discriminations (e.g. narrow gaps in the connected problem), to conceptually mediated problems associated with understanding means–end problems (e.g. understanding abstract–relational concepts), or to some combination of these factors. Further research is needed to tease apart these possibilities. Fourth, although the tamarins were able to inhibit some actions, some problems proved difficult because of their inability to inhibit a reaching response. Future work must assess the degree to which problems of inhibition override a species' capacity to solve problems conceptually and the extent to which interspecific differences in the prefrontal cortex contribute to interspecific differences in problem-solving ability.

Acknowledgments

For help in running the experiments, we thank the following superb undergraduates and graduate students at Harvard University: Linda Anderson, Elizabeth Dorash, Brian Ericson, Maria Garrett, Charles Kapelke, Jon Sakata, Laurie Santos and Karissa Schecter. For comments on the data and manuscript, we thank Susan Carey, David Premack, Robert Seyfarth, David Sherry, Charles Snowdon and Elizabeth Spelke. I thank the New England Regional Primate Research Center for supporting our tamarin colony (PHSP51RR00168-36). Support for this project was provided by funds to MDH from a National Science Foundation Young Investigator Award and Harvard University. The research presented here was described in Animal Research Protocol No. 92-16 approved on 11 June 1997 by the Institutional Animal Care and Use Committee of Harvard University.

References

Anderson, J. 1990. Use of objects as hammers to open nuts by capuchin monkeys (*Cebus apella*). *Folia primatologica*, **54**, 138–145.

Barton, R. A. & Dunbar, R. I. M. 1997. Evolution of the social brain. In: *Machiavellian Intelligence. Vol. 2* (Ed. by A. Whiten & R. Byrne), pp. 240–263. Cambridge: Cambridge University Press.

Boysen, S. T. 1996. ''More is less'': The distribution of rule-governed resource distribution in chimpanzees. In: *Reaching into Thought: The Minds of the Great Apes* (Ed. by A. E. Russon, K. A. Bard & S. T. Parker), pp. 177–189. Cambridge: Cambridge University Press.

Deacon, T. W. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain.* New York: Norton.

Diamond, A. 1988. Differences between adult and infant cognition: is the crucial variable presence or absence of language? In: *Thought Without Language* (Ed. by L. Weiskrantz), pp. 337–370. Oxford: Clarendon Press.

Goldman-Rakic, P. S. 1987. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: *Handbook of Physiology: The Nervous System* (Ed. by F. Dean), pp. 373–417. Bethesda, Maryland: American Physiological Society.

Goldman-Rakic, P. S. 1990. Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. In: *Progress in Brain Research 85: The Prefrontal Cortex: Its Structure, Function and Pathology* (Ed. by H. B. M. Uylings, C. G. Van Eden, J. P. De Bruin, M. A. Corner & M. G. Feenstra), pp. 325–336. New York: Elsevier.

Harlow, H. F. 1949. The formation of learning sets. *Psychological Review*, **56**, 51–65.

Hauser, M. D. 1997. Artifactual kinds and functional design features: what a primate understands without language. *Cognition*, **64**, 285–308.

Hauser, M. D. 1998. Expectations about object motion and spatial location: experiments with a nonhuman primate. *Developmental Science*, **1**, 31–38.

Hauser, M. D. & Carey, S. 1998. Building a cognitive creature from a set of primitives: evolutionary and developmental insights. In: *The Evolution of Mind* (Ed. by C. Allen & D. Cummins), pp. 51–106. Oxford: Oxford University Press.

Hauser, M. D., Kralik, J., Botto, C., Garrett, M. & Oser, J. 1995. Self-recognition in primates: phylogeny and the salience of species-typical traits. *Proceedings of the National Academy of Sciences, U.S.A.*, **92**, 10 811–10 814.

Hunt, G. R. 1996. Manufacture of hook-tools by New Caledonian crows. *Nature*, **379**, 249–251.

Keil, F. C. 1994. The birth and nurturance of concepts by domains: the origins of concepts of living things. In: *Mapping the Mind: Domain Specificity in Cognition and Culture* (Ed. by L. A. Hirschfield & S. A. Gelman), pp. 234–254. Cambridge: Cambridge University Press.

Kohler, W. 1925. *The Mentality of Apes*. London: Routledge & Kegan Paul.

Konishi, M. 1985. Birdsong: from behavior to neuron. *Annual Review of Neuroscience*, **8,** 125–170.

Konishi, M. 1993. Listening with two ears. *Scientific American*, **268,** 66–73.

McGrew, W. C. 1992. *Chimpanzee Material Culture*. Cambridge: Cambridge University Press.

Marler, P. & Nelson, D. A. 1992. Neuroselection and song learning in birds: species universals in a culturally transmitted behavior. *Seminars in Neuroscience*, **4,** 415–423.

Matsuzawa, T. 1996. Chimpanzee intelligence in nature and in captivity: isomorphism of symbol use and tool use. In: *Great Ape Societies* (Ed. by W. C. McGrew, L. F. Marchant & T. Nishida), pp. 196–209. Cambridge: Cambridge University Press.

Miles, R. C. 1965. Discrimination-learning sets. In: *Behavior of Nonhuman Primates. Modern Research Trends* (Ed. by A. M. Schrier, H. F. Harlow & F. Stollnitz), pp. 51–96. New York: Academic Press.

Miles, R. C. & Meyer, D. R. 1956. Learning sets in marmosets. *Journal of Comparative and Physiological Psychology*, **49,** 219–222.

Robin, N. & Holyoak, K. J. 1995. Relational complexity and the functions of prefrontal cortex. In: *The Cognitive Neurosciences* (Ed. by M. Gazzaniga), pp. 987–998. Cambridge, Massachusetts: MIT Press.

Rumbaugh, D. M., Savage-Rumbaugh, E. S. & Washburn, D. A. 1996. Toward a new outlook on primate learning and behavior: complex learning and emergent processes in comparative perspective. *Japanese Psychological Research*, **38,** 113–125.

Rylands, A. B. 1996. Habitat and the evolution of social and reproductive behavior in the Callitrichidae. *American Journal of Primatology*, **38,** 5–18.

Santos, L. R., Ericson, B. & Hauser, M. D. In press. Constraints on problem solving and inhibition: object retrieval in cotton-top tamarins. *Journal of Comparative Psychology.*

Savage, A., Giraldo, L. H., Soto, L. H. & Snowdon, C. T. 1996. Demography, group composition, and dispersal in wild cotton-top tamarin (*Saguinus oedipus*). *American Journal of Primatology*, **38,** 85–100.

Sawaguchi, T. & Kudo, H. 1990. Neocortical development and social structure in primates. *Primates*, **31,** 283–290.

Schrier, A. M., Harlow, H. F. & Stollnitz, F. 1965. *Behavior of Nonhuman Primates*. New York: Academic Press.

Simmons, J. A. 1989. A view of the world through the bat's ear: the formation of acoustic images in echolocation. *Cognition*, **33,** 155–199.

Stephan, H., Frahm, H. D. & Baron, G. 1981. New and revised data on volumes of brain structures in insectivores and primates. *Folia primatologica*, **35,** 1–29.

Suga, N. 1988. Auditory neuroethology and speech processing: complex sound processing by combination-sensitive neurons. In: *Auditory Function* (Ed. by G. M. Edelman, W. E. Gall & W. M. Cowan), pp. 679–720. New York: Wiley Liss Press.

Tonooka, R., Tomonaga, M. & Matsuzawa, T. 1997. Acquisition and transmission of tool making and use for drinking juice in a group of captive chimpanzees (*Pan troglodytes*). *Japanese Psychological Research*, **39,** 253–265.

Uller, C. 1997. Origins of numerical concepts: a comparative study of human infants and nonhuman primates. Ph.D. thesis, Massachusetts Institute of Technology.

Visalberghi, E. 1990. Tool use in *Cebus. Folia primatologica*, **54,** 146–154.

Visalberghi, E. & Fragaszy, D. 1991. Do monkeys ape? In: *''Language'' and Intelligence in Monkeys and Apes* (Ed. by S. T. Parker & K. R. Gibson), pp. 247–273. Cambridge, Massachusetts: Cambridge University Press.

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

Visalberghi, E. & Limongelli, L. 1996. Acting and understanding: tool use revisited through the minds of capuchin monkeys. In: *Reaching into Thought: The Minds of the Great Apes* (Ed. by A. E. Russon, K. A. Bard & S. T. Parker), pp. 57–79. Cambridge: Cambridge University Press.

Visalberghi, E., Fragaszy, D. M. & Savage-Rumbaugh, E. S. 1995. Performance in a tool-using task by common chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), an orangutan (*Pongo pygmaeus*), and capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, **109,** 52–60.

Warren, J. M. 1965. Primate learning in comparative perspective. In: *Behavior of Nonhuman Primates. Modern Research Trends* (Ed. by A. M. Schrier, H. F. Harlow & F. Stollnitz), pp. 249–282. New York: Academic Press.

Westergaard, G. C., Lundquist, A. L., Kuhn, H. E. & Suomi, S. 1997. Ant-gathering with tools by captive tufted capuchins (*Cebus apella*). *International Journal of Primatology*, **18,** 95–103.

Willatts, P. 1984. The Stage-IV infant's solution of problems requiring the use of supports. *Infant Behavior and Development*, **7,** 125–134.

Willatts, P. 1990. Development of problem-solving strategies in infancy. In: *Children's Strategies: Contemporary Views of Cognitive Development* (Ed. by D. F. Bjorklund), pp. 143–182. Hillsdale, New Jersey: Lawrence-Erlbaum.

Yamatani, K., Taketoshi, O., Nishijo, H. & Takaku, A. 1990. Activity and distribution of learning-related neurons in monkey *Macaca fuscata* prefrontal cortex. *Behavioral Neuroscience*, **104,** 503–531.