



Published in final edited form as:

*Anim Cogn.* 2013 May ; 16(3): 445–458. doi:10.1007/s10071-012-0585-8.

## Automated cognitive testing of monkeys in social groups yields results comparable to individual laboratory based testing

Regina Paxton Gazes, Emily Kathryn Brown, Benjamin M. Basile, and Robert R. Hampton  
Department of Psychology and Yerkes National Primate Research Center, Emory University

### Abstract

Cognitive abilities likely evolved in response to specific environmental and social challenges and are therefore expected to be specialized for the life history of each species. Specialized cognitive abilities may be most readily engaged under conditions that approximate the natural environment of the species being studied. While naturalistic environments might therefore have advantages over laboratory settings for cognitive research, it is difficult to conduct certain types of cognitive tests in these settings. We implemented methods for automated cognitive testing of monkeys (*Macaca mulatta*) in large social groups (Field station) and compared the performance to that of laboratory housed monkeys (Laboratory). The Field station animals shared access to four touch screen computers in a large naturalistic social group. Each Field station subject had an RFID chip implanted in each arm for computerized identification and individualized assignment of cognitive tests. The Laboratory group was housed and tested in a typical laboratory setting, with individual access to testing computers in their home cages. Monkeys in both groups voluntarily participated at their own pace for food rewards. We evaluated performance in two visual psychophysics tests, a perceptual classification test, a transitive inference test, and a delayed matching to sample memory test. Despite differences in housing, social environment, age, and sex, monkeys in the two groups performed similarly in all tests. Semi-free ranging monkeys living in complex social environments are therefore viable subjects for cognitive testing designed to take advantage of the unique affordances of naturalistic testing environments.

### Keywords

transitive inference; memory; classification; psychophysics; rhesus macaque; social housing; RFID

---

Many animals live complicated lives. They may spend their days foraging, hunting, avoiding predators, socializing, fighting, mating, migrating, defending territories, or caring for offspring. Cognitive abilities have likely evolved in response to specific environmental and social challenges and are therefore expected to be specialized for each species' life history (Shettleworth 2009). Specialized cognitive abilities may be most readily engaged and measured under conditions that approximate critical aspects of the natural environment of a given species. Most studies of cognition are conducted in laboratories, devoid of many of the social and physical challenges present in nature. Laboratory conditions afford excellent control over developmental histories and testing conditions, facilitating studies of processes such as learning that may be more difficult or impossible to document under natural conditions. However, laboratory environments often limit exposure to natural cognitive demands, and may deprive subjects of experiences necessary for normal cognitive

development (Rommeck et al. 2011). Animals raised in more complex and natural physical and social conditions may therefore exhibit a more developed range of cognitive abilities and behaviors (Gersick et al. 2012; White et al. 2010; White et al. 2012).

Creative field experiments have been successfully used to explore complex cognitive process such as communication (Biben and Bernhards 1994; Windfelder 2001; Zuberbuhler et al. 1999), individual recognition (Cheney and Seyfarth 1982; Fischer 2004), social cognition (Bergman et al. 2003; Cheney and Seyfarth 1999), transitive inference (Weiß et al. 2010), and memory (Henderson et al. 2001). However, the ability to conduct field studies can be limited by difficulties controlling the location of subjects, their experience with particular stimuli, attrition, and other factors such a lighting, noise, and the presence of predators or other distractions. Studies of learning, in particular, are difficult to conduct in natural settings because inferences about learning require comprehensive documentation and control of subjects' experience. It would therefore be advantageous to develop study settings that provide the benefits of complex group living while preserving much of the accessibility and experimental control afforded by standard laboratory settings.

Technological advances in radio frequency identification (RFID) and computer technology have made automated cognitive testing of primates living in complex environments possible. Andrews and Rosenblum (1994), Wallen (Hasset et al. 2007), and Fagot (2009) have created cognitive testing systems that use RFID chips to identify individual monkeys as they complete computerized motor, perceptual, and cognitive tasks. Monkeys living in small groups have been found to perform motor and cognitive tasks (motor: Andrews and Rosenblum 1994; visual search: Barbet and Fagot 2011; Bonte et al. 2011; Fagot and Bonte 2010; working memory: Fagot and De Lillo 2011; complex matching: Fagot and Paleressompouille 2009; orthographic processing: Grainger et al. 2012), indicating the feasibility of this method of testing.

However, no studies have compared learning and performance between monkeys living in complex naturalistic environments and monkeys living in traditional laboratory environments. Performance might differ between these groups of monkeys for either central or peripheral reasons. For example, semi-natural environments may foster superior cognitive development, leading to differences in central cognitive capacities and thus performance. Additionally adlibitum testing may lead to improved learning compared to more restricted testing schedules (Fagot and Thompson, 2011). By contrast, differences in housing conditions may affect performance but not central cognitive capacities, due to differences in the frequency of distractions, motivation to participate in testing, and availability of alternative activities. Specifically, monkeys in complex environments have many behavioral options, may be distracted by other individuals, are subject to changing weather and lighting conditions, and access to testing equipment may be limited by competition and dominance (Drea 1998; Drea and Wallen 1999). These factors have the potential to impact the acquisition and validity of cognitive data.

To assess the validity of data gathered from monkeys housed in complex social environments, we compared learning and performance on four cognitive tasks between monkeys living in a laboratory setting and monkeys living in a large, naturalistic social group. We used four tasks from core areas of cognitive research: psychophysical and perceptual classification tasks to assess visual perceptual function, a transitive inference test to assess complex cognitive function, and a delayed-matching-to-sample task to assess memory performance.

## Subjects and apparatus

Subjects were rhesus macaque monkeys (*Macaca mulatta*) housed at the Yerkes National Primate Research Center in Atlanta, GA and Lawrenceville, GA. All procedures were approved by the Institutional Animal Care and Use Committee of Emory University and were in compliance with National Institutes of Health guidelines for the care and use of laboratory animals. Subjects were housed and tested in one of two environments: Laboratory and Field station.

### Laboratory subjects and apparatus

Twenty-four laboratory housed monkeys, aged 4–8 years, participated in this study, six in each of the four experiments. All Laboratory subjects were adult male rhesus monkeys who had been raised by their biological mothers in a large social group until approximately 2.5 years of age. Whenever possible, monkeys were pair-housed, and all monkeys received enrichment consisting of hard rubber toys, puzzle feeders, foraging devices, and destructible items. All monkeys were kept on a 12:12 light:dark cycle with light onset at 7:00 am. Animals received a full ration of food daily, supplemented with fruits and vegetables, and ad libitum access to water. All subjects had extensive histories with computerized cognitive testing (e.g., Adachi et al., 2009; Basile and Hampton, 2010, 2011; Paxton et al., 2010; Templer and Hampton, 2012).

Laboratory testing occurred in the subjects' home cages. Computerized touch-screen test systems, each consisting of a 15-inch LCD color monitor running at a resolution of 1024 X 768 pixels, stereo speakers, two automated food dispensers (Med Associates Inc. St. Albans, VT), and two food cups below the screen, were attached to the front of each monkey's cage (Figure 1). Computer screens were locked to the front of each monkey's cage and the door was raised, giving subjects full visual and tactile access to the screen during testing. Correct responses were rewarded with nutritionally balanced fruit flavored pellets on a majority of trials and miniature chocolate candies on a minority of trials.

During testing, pair-housed monkeys were separated by an opaque plastic divider with holes that allowed visual, auditory, and tactile contact but prevented the monkeys from accessing their partner's equipment. Test sessions were conducted between 10 am and 5 pm, six days per week.

### Field station subjects and apparatus

Field station subjects lived and were tested in a large multi-male multi-female group of 77 individuals (not counting young infants). Monkeys lived primarily outdoors in a 30 × 30 meter enclosure with access to a temperature-moderated indoor housing area. Food and water were available ad libitum and animals received fruits and vegetables daily. All animals had a small RFID microchip (Biomark, Boise ID) implanted in each forearm for automated individual identification.

Four touch screen computer stations were located immediately outside the enclosure. Each station included a 15-inch LCD color monitor (3M, St. Paul, MN) running at a resolution of 1024 X 768 pixels, an automated pellet dispenser (Med Associates Inc. St. Albans, VT), stereo speakers, an RFID chip reader (Biomark, Boise, ID), and a stool for monkeys to sit on while testing (Figure 2). Touch screens were located 15 cm behind a poly panel, in an enclosed area that limited incursion of ambient light. The touch screens could be viewed through a 15 × 20 cm mesh window and could be reached through a 5 cm diameter arm hole that was surrounded by an antenna for reading the RFID chips. Correct responses were reinforced with sucrose or nutritionally-balanced fruit-flavored pellets.

The RFID reader at each testing station was connected to a computer that controlled stimulus presentation and recorded data. When a monkey put its arm through the arm hole to touch the touch screen, the reader identified the monkey, and the computer selected the appropriate task and trial for that subject and recorded responses in a subject-specific data file. In this way, we controlled the tasks presented to each subject and tracked individual performance. When a monkey returned to the testing system the experiment began where the animal left off when it last tested. Incomplete trials were restarted when the monkey returned to the testing system.

The four testing stations were available 24 hours per day, 7 days per week. Testing stations were visually monitored via a remote controlled internet camera (Axis Communications, Lund, Sweden) and data could be accessed via internet at any time to assess progress. At the start of these experiments, subjects had been trained to touch images on the screen and had achieved at least 90% accuracy on a four choice delayed matching-to-sample task with a 200ms delay interval. All subjects completed four experiments in the same order: perceptual classification, visual psychophysics, transitive inference, and delayed matching-to-sample. For each of the four experiments presented in this paper, data from all subjects who had completed that experiment by April 2012 are reported. Thus, different experiments have different subjects and different numbers of subjects. The sex, age, and rank of subjects therefore vary and are reported for each experiment.

## Field station testing demographics

Field station data presented here were collected beginning when the cognitive testing system was installed in May 2010 and continuing through April 2012. During that time 39 of 77 non-infant monkeys completed the initial shaping program and were eligible to participate in the experiments.

A larger proportion of juvenile (under 5 years of age) than adult animals used the testing system, and both males and females participated (Table 1). The four adult male group members never participated.

The dominance hierarchy of the group was established through creation of a matrix of dyadic interactions (Bernstein 1970; De Vries and Appleby 2000) based on observation of dominance and submissive behaviors: displacement, threat, chase, attack, and fear grimace. The 30 adult females were then divided into High, Middle, and Low ranking groups of 10 animals each. Juveniles were assigned the same rank as their mothers. Animals from all three rank groupings used the testing system (Table 1). While a larger sample from each rank group is needed to draw conclusions about the relationship between rank and cognitive performance, preliminary analyses conducted on performance data from the four experiments in this paper revealed no significant differences between rank groups.

## Experiment 1: Visual psychophysics

One concern about outdoor group testing is the lack of environmental control (Fagot and Bonte 2010). For example, whereas laboratory lighting is constant, outdoor lighting is variable, which could alter the appearance of stimuli between trials. Visual psychophysical tests allow sensitive comparisons of perception because they require subjects to discriminate subtle differences in stimuli.

To assess visual perception, we tested the ability of monkeys to choose a target stimulus from among distracters in two tasks, a size discrimination task and a brightness discrimination task. We varied the difficulty of the two tasks and included very difficult trials to ensure that differences in visual perception could be detected if they existed. If

differences between the laboratory and field station environments significantly affect the perception of on-screen stimuli, then the two groups should show different learning rates or patterns of accuracy across different levels of difficulty in the two tasks.

## Subjects

Laboratory subjects were 6 pair-housed 4-year-old male rhesus macaque monkeys with one year of experience with computerized cognitive testing.

Field station subjects were 2 male and 10 female subjects aged 2–4 years ( $M = 2.8$  years). Five subjects were members of the high ranking group, 5 were members of the middle ranking group, and 2 were members of the low ranking group.

## General procedure

Within each task, the target stimulus remained the same across trials, and difficulty was varied on a trial by trial basis by changing the discriminability of three identical distracters from the target. To start a trial, monkeys touched a green start box at the bottom center of the screen. To prevent counting spurious touches as responses, all choices required two consecutive touches (FR2) within the image border of a single stimulus. The target and the three identical distracters then appeared in the four corners of the screen, with location counterbalanced and pseudorandomized (Figure 3). Within each session, the distracters differed from the target by 5 levels of difficulty, each consisting of two different distracter values, one lesser (smaller or darker) and one greater (larger or brighter) in magnitude than the target by equal amounts (Figure 4).

Selection of the target stimulus resulted in positive auditory and food reinforcement. Selection of a distracter stimulus resulted in negative auditory feedback and a black screen time out period. Consecutive trials were separated by a 2-second inter-trial interval during which the screen was black. Each session consisted of 100 trials, with 20 trials from each of the five difficulty levels.

**Size discrimination**—Monkeys were required to select a 100-pixel diameter grey circular target (128,128,128 RGB) from among distracters that matched the target in color and brightness, but differed in size. The five levels of distracter difficulty ranged from easy ( $\pm 40$  pixels difference in diameter from the target) to hard ( $\pm 8$  pixels difference in diameter; Figure 4, *top*).

Subjects were trained on the size discrimination until they had completed at least five 100 trial sessions and had reached criterion of above 85% correct on the easiest difficulty level (level 1) over two consecutive sessions.

**Brightness discrimination**—After they reached criteria on the size discrimination test, subjects were presented with the brightness discrimination. Stimuli were grey  $100 \times 100$  pixel squares that varied in brightness, but were identical along all other dimensions. The target stimulus was medium grey (128 128,128 RGB). The five levels of distracters ranged from easy ( $\pm 64$  RGB value difference) to hard ( $\pm 24$  RGB value difference; Figure 4, *middle*) with two distracters per level, one darker than the target and one lighter.

For all experiments presented in this paper, proportion correct scores were arcsine transformed to best approximate normality (Aron and Aron 1999). An alpha level of .05 and two-tailed tests were used for all analyses. Throughout this paper RMANOVA were used to assess learning and performance differences that involved multiple tests of the same animals. Independent samples t-tests were used to compare performance between the Field

station and Laboratory subjects; one sample t-tests were used to compare performance to chance, and paired samples t-tests were used to compare performance within each condition.

## Results and discussion

There was no significant difference between the Laboratory and Field station groups in the number of errors required to reach criteria (Mean<sub>Lab</sub> + SD = 107.17 + 16.73, Mean<sub>FS</sub> + SD = 156.50 + 89.10; independent samples t-test,  $t_{15} = -1.52$ ,  $p=0.15$ ). However Laboratory monkeys learned the brightness discrimination to criteria with fewer errors than Field station monkeys (Mean<sub>Lab</sub> + SD = 335.83 + 170.00, Mean<sub>FS</sub> + SD = 900.18 + 240.40; independent samples t-test,  $t_{15} = -2.41$ ,  $p=0.03$ ).

Difficulty level affected discrimination performance of the two groups equally. A two way repeated measures ANOVA (Difficulty Level X Group) comparing average accuracy in the two criterial sessions showed a main effect of difficulty level for both the size and brightness discriminations (Size:  $F_{4,60} = 208.39$ ,  $p < 0.01$ ; Brightness:  $F_{4,60} = 80.37$ ,  $p < 0.01$ ). However, there was no main effect of group (Size:  $F_{1,15} = 4.12$ ,  $p = 0.06$ ; Brightness:  $F_{1,15} = 0.85$ ,  $p = 0.37$ ) and no interaction effect for either task (Size:  $F_{4,60} = 0.88$ ,  $p = 0.48$ ; Brightness:  $F_{4,60} = 1.81$ ,  $p = 0.14$ ; Figure 5), indicating that testing environment did not affect perception dramatically.

The Laboratory group was tested in an environment with consistent lighting within and between testing sessions. We attempted to hold the light conditions constant for the Field station group by shielding the video screens from outside light; however, changes in sunlight within and between testing days may have altered the perceived brightness of the stimuli, and could account for the difference in rate of learning this task. There was no significant difference in asymptotic performance on either the size or brightness discrimination between the two groups, indicating that housing and testing environments did not dramatically affect perception of basic stimulus properties. Nonetheless, it should be noted that Laboratory monkeys did tend to perform better than the Field station monkeys on the size discrimination, and that the main effect of group in the size discrimination approached statistical significance. These trends may be due to differences in testing environment, or to differences in experience, age, and sex of subjects. Future studies with larger numbers of subjects will further explore the effect of these demographic factors on performance.

## Experiment 2: Perceptual classification

Laboratory and Field station groups performed similarly on two visual psychophysics tasks in Experiment 1, indicating that they perceive on-screen stimuli similarly despite differences in lighting conditions. In Experiment 2, we required monkeys to classify photographs of birds, fish, flowers, and people based on their shared perceptual features. Like the previous psychophysical tests, this classification task also requires accurate perception of visual stimuli, but with an increased cognitive demand associated with assigning the photographs to categories.

Perceptual classification tasks can be successfully solved using different strategies. At a narrow level, subjects may memorize appropriate responses for each individual image (Schrier et al. 1984) or base discriminative responding on perceptual features that humans would consider irrelevant to category membership (D'Amato and Van Sant 1988). At a broader level, subjects may generalize discriminative responding to a set of perceptual features, or invariants, that accurately describe the category (Schrier and Brady 1987). One can determine whether subjects have learned a narrow or broad rule by requiring them to classify novel images. Subjects that have extracted a broad set of common features will accurately classify novel images that share those broad features, whereas subjects using a

narrow set of features will show poor transfer because novel images are unlikely to conform to a narrow rule. In Experiment 2, we evaluated whether differences between the Laboratory and Field station environments caused differences in acquisition and transfer of a four-choice classification task.

## Subjects

Laboratory subjects were six pair-housed 4–5 year old male rhesus macaque monkeys with a 1 year history with computerized cognitive testing. They were housed and tested as previously described for all Laboratory subjects, but were not the same subjects who participated in Experiment 1.

Field station subjects were 2 male and 11 female subjects aged 2–4 years ( $M = 2.75$  years) housed at the Yerkes Field Station. Five subjects were members of the high ranking group, 6 were members of the middle ranking group, and 2 were members of the low ranking group. Twelve of the 13 subjects also participated in Experiment 1.

## Stimuli and procedure

Training stimuli were 400 color photographs 100 from each of four categories: fish, flowers, birds, and people. All images were gathered from the online photo repository Flickr (Yahoo!, Sunnyvale, CA) and duplicates were eliminated using DupDetector (Prismatic Software, Anaheim, CA) and visual inspection. Images were cropped to 400×300 pixels using Adobe Photoshop (Adobe, San Jose, CA). Each image contained at least one representative of its assigned category, but varied widely in other perceptual features (e.g. fish included goldfish and sharks, alone and in schools, in the ocean and on a plate). Images were screened manually to ensure that they did not contain representatives from more than one category.

To start a trial, monkeys touched a green start box at the bottom center of the screen (FR2). A sample image from one of the four categories appeared in the center of the screen. When the subject touched the sample (FR2), four 100 X 100 pixel classification icons appeared in the corners of the screen. Each of the four categories was associated with one of the icons, which always appeared in the same location on the screen (Figure 6). Selection of the correct category icon for the sample was reinforced with a positive auditory reinforcer and a food reward. Selection of an incorrect icon resulted in a negative sound and a 5 second black screen time out period, followed by a correction trial which repeated the original trial exactly. A correct choice on the correction trial resulted in a food reward and auditory reinforcer. An incorrect choice resulted in a negative sound, a 5-second time out, and a second correction trial. On the second correction trial, the start box and the sample were presented in the same way as a normal trial, but at test only the correct classification icon was present. When the monkey touched this icon he was rewarded with a positive auditory reinforcer and a food reward. Only performance on the first iteration of each trial was used in analyses.

A 3-second black screen inter-trial interval separated consecutive trials. Training sessions consisted of 400 trials 100 from each category. Subjects were trained until they classified images to criterion level. Laboratory and Field station monkeys were trained on this task before the comparison presented here was planned. Accidentally they were trained to different criteria. For Laboratory monkey criterion was overall accuracy above 75% for one session, for Field station monkeys it was accuracy over 80% on each category on one session.

To assess whether subjects learned to categorize images using a broad or narrow set of features, we conducted a transfer test in which subjects saw 50 novel images from each

category intermixed with the 400 training images. If performance on these transfer images did not differ significantly from the criterion level reached on training images, it would indicate that subjects based their choices on a category specific array of perceptual features instead of a narrow selection of simple features. We conducted one 600-trial transfer session which was run in the same way as training sessions.

## Results and discussion

The Laboratory and Field station groups did not differ in the numbers of errors made before reaching the Laboratory criterion of above 75% performance (Mean<sub>Lab</sub>+SD= 1837.00 + 342.86, Mean<sub>FS</sub>+SD=1976.00 + 639.23 ; independent samples t-test,  $t_{17}=0.43$   $p=0.67$ ).

Analysis of performance on the 400 trained and 200 novel transfer images presented during the single transfer session indicated that performance did not differ from criterion level for either training or transfer images in the Laboratory (one sample t-tests compared to 75%: *transfer images*:  $t_5=.15$ ,  $p=0.89$ ; *training images*:  $t_5=.73$ ,  $p=0.50$ ; Figure 7, *left*) or Field station group (one sample t-test compared to 80%: *transfer images*:  $t_{12}=48$ ,  $p=0.14$ ; *training images*:  $t_{12}=7.08$ ,  $p<0.01$ ). The Laboratory group showed no difference in performance between trained and transfer images (paired samples t-test,  $t_5=1.5$ ,  $p=0.30$ ), while the Field station monkeys showed a small (6%) but significant decrease in performance on the transfer images (paired samples t-test,  $t_{12}=6.17$ ,  $p<0.01$ ) resulting from *above* criterion performance on training images (Figure 7, *right*).

The two groups learned to classify images at the same rate, indicating that the testing environment had no effect on acquisition. Additionally, both groups performed at criterion level on their first, and only, exposure to transfer images, which indicates that monkeys likely classified items based on a broad set of perceptual features shared by all category members, not based on memorized responses to specific stimuli. Monkeys in the two groups performed similarly, showing that testing environment did not have a substantial effect on perceptual classification.

## Experiment 3: Transitive inference

In Experiments 1 and 2, Laboratory and Field station monkeys performed similarly on three perceptual tasks, indicating that differences in lighting or distractions did not interfere with cognitive testing in the Field station environment. However, it is still possible that the distractions and behavioral options present in the Field station environment may compromise performance on more cognitively demanding tasks where the basis for correct responses is not perceptually available. Transitive inference (TI) requires subjects to infer non-perceptual relations between items based on their shared relations to other items. For example, if you know that Jane is taller than Sue and Sue is taller than Mary, you can infer that Jane is taller than Mary without perceiving Jane and Mary together. TI has long been considered a prototypically cognitive process (McGonigle and Chalmers 1977; Piaget 1960), as the correct answer is not observable, but must be based on mentally combining previously learned information.

In the laboratory, TI is often studied using pairs of overlapping stimulus discriminations, such that a subject is trained that A is rewarded when paired with B (A+B-), B is rewarded when paired with C (B+C-), and so on through item G (C+D-; D+E-; E+F-; F+G-). Never before seen non-adjacent pairs of stimuli are then presented to test whether subjects infer relations between these stimuli (e.g. BD, CF). If subjects correctly select the higher ranked item (e.g. B in BD, or C in CF), it suggests that they have a representation of the implied order of the stimuli.



Many species solve TI tasks (corvids: Bond et al. 2010; fish: Grosenick et al. 2007; crows: Lazareva et al. 2004; lemurs: MacLean et al. 2007; monkeys: McGonigle and Chalmers 1977; rats: Roberts and Phelps 1994; pigeons: Von Fersen et al. 1991). Additionally, most studies report a symbolic distance effect (SDE), such that performance increases and response latency decreases when the distance in the implied order between the two images in a test pair is larger. The presence of this effect is often taken to indicate that subjects formed an ordered cognitive representation of the TI stimuli.

The only study to examine TI performance in animals living in a natural social environment found that greylag geese performed above chance on critical non-adjacent test trials. However there are no data from this same task in captive geese, so we cannot assess whether they performed similarly or relied on the same cognitive mechanisms to solve the task. We assessed whether testing environment influences learning and performance on a transitive inference task in monkeys. In addition to performance on internal test trials, we examined learning rates and the symbolic distance effect to determine if subjects were relying on the same cognitive mechanism to solve the task.

## Subjects

Laboratory subjects were six 8-year-old male rhesus macaque monkeys with a 5 year history with computerized cognitive testing. They were housed and tested as described for all Laboratory subjects, but did not participate in any of the previous experiments. Four subjects were pair housed, two were separated by a slotted panel that allowed limited physical and visual contact.

Field station subjects were 2 male and 10 female subjects aged 2–4 years ( $M = 2.9$  years). Four subjects were members of the high-ranking group, 6 were members of the middle-ranking group, and 2 were members of the low-ranking group. All subjects participated in the previous 2 experiments.

## Stimuli and procedure

Seven 300 X 300 pixel color clip art images were presented in six overlapping adjacent pairs (AB, BC, CD, DE, EF, FG). There were two sets of distinct clip art images used in this experiment, with half of the subjects in each training environment trained on each set. Neither group of subjects had prior experience with transitive inference tasks.

A green box appeared at the bottom of the screen and remained until the monkey touched it (FR2) to start a trial. Two adjacent clip art images from the training set appeared on the right and left sides of the screen (counterbalanced over trials), and monkeys were required to touch one of the two images (FR2). Selection of the correct image always resulted in an auditory reinforcer that was coupled with a food reward on 80% of trials, while selection of the incorrect image in the pair resulted in a negative auditory stimulus and a five-second black screen time out. A 3-second inter-trial interval (ITI) separated each trial.

Training proceeded from the adjacent pair at the bottom of the order (FG) to the pair at the top of the order (AB, Figure 8, *left*). Each training pair was introduced in 25-trial sessions consisting of just that pair until subjects reached 80% correct. Then 25 trials of that pair were intermixed with 25 trials of each of the previously trained pairs, until performance on all of the presented pairings within a session was above 80% (Treichler and Van Tilburg 1996). This pattern continued until all 6 adjacent training pairs had been presented and learned.

Once subjects reached 80% or better on all 6 training pairs intermixed in one session, one trial of each never before seen non-adjacent test pair was added into the session, for a total

of 15 test trials pseudo-randomly intermixed with the original 150 training trials (25 of each adjacent training pair type). To prevent subjects from learning on the test trials, every test trial response was reinforced with a positive auditory reinforcer only, whether correct or not. Auditory only reinforcement was consistent with the pattern of reinforcement monkeys had come to expect for correct responses during training, which included 20% auditory only reinforcement trials. Subjects received 4 sessions of these test trials. Critical non-adjacent test trials were internal pairs (BD, BE, BF, CD, CF, DF) that did not contain the first or last image in the list (A or G), as these had either been always or never reinforced, respectively (Figure 8, *right*).

## Results and discussion

Housing environment had no effect on the rate of learning the adjacent pairs, and monkeys performed similarly on the probe test pairs. There was no significant difference in the total number of errors made before reaching criteria by monkeys in the Laboratory and Field station groups (Mean<sub>Lab</sub> + SD = 785.33 + 283.23, Mean<sub>FS</sub> + SEM = 933.67 + 385.70; independent samples t-test:  $t_{16} = 0.82$ ,  $p = .47$ ). Both the Laboratory and Field station groups selected the higher ranked image significantly more often than expected by chance on internal test trials (Mean<sub>Lab</sub> + SEM = 70.83 + 4.93%; one sample t-test,  $t_5 = 3.93$ ,  $p = 0.01$ ; Mean<sub>FS</sub> + SEM = 60.71 + 4.26%; one sample t-test,  $t_{11} = 2.50$ ,  $p = 0.03$ ), and there was no significant difference in internal test trial performance between the two groups (independent samples t-test:  $t_{16} = -1.46$ ,  $p = 0.16$ ).

Subjects in both groups displayed the SDE for test trial accuracy, showing increased accuracy as the symbolic distance between the images increased, and there was no significant difference in this effect between the two groups (RMANOVA group X symbolic distance: group  $F_{1,64} = .004$ ,  $p = .95$ ; symbolic distance  $F_{4,64} = 39.23$ ,  $p = .00$ ; interaction  $F_{4,64} = 1.40$ ,  $p = .25$ ; Figure 9, *left*). Neither group showed the SDE for median response latency, which did not differ as a function of group, symbolic distance, or the interaction of those two factors (RMANOVA group X symbolic distance: group  $F_{1,64} = .09$ ,  $p = 0.77$ ; symbolic distance  $F_{4,64} = 1.15$ ,  $p = 0.34$ ; interaction  $F_{4,64} = 2.02$ ,  $p = 0.10$ ; Figure 9, *right*). Results from both groups of subjects are consistent with previous TI performance by rhesus monkeys (Gazes et al., 2012; Rapp et al., 1996; Treichler and Raghanti, 2010).

Learning and performance did not differ between the two groups on this task, suggesting that monkeys in both housing conditions relied on the same cognitive mechanism to solve the TI task. Additionally, these results indicate that the distractions of the Field station testing environment did not inhibit monkeys' ability to perform complex cognitive tasks.

## Experiment 4: Memory

In Experiments 1 through 3, Field station monkeys learned and performed comparably to Laboratory monkeys on psychophysical, perceptual classification, and transitive inference tasks. However, all of these tasks required only a short period of attention on each trial. It is possible that Field station monkeys will be unwilling or unable to complete long trials containing delays due to the distractions in their environment. This is the first experiment to test performance of monkeys in naturalistic groups in memory tests with significant delay intervals.

In delayed matching-to-sample tests, subjects see an image, and after a delay must select it from among several distracter images. Visual interference during the delay impairs memory on these types of tasks in humans and monkeys, even when the interfering information is only passively viewed (Logie 1986; Phillips and Christie 1977; Washburn and Astur 1998). Because the Field station monkeys live in a more complex environment than Laboratory

monkeys, there may be more interference from the environment, which could produce poorer performance than would be obtained under more controlled conditions. The effect of interference is strongest when the to-be-remembered images are from a small, frequently repeating image set (Basile and Hampton, in press). Therefore, to maximize the possibility of discovering limitations of conducting memory tasks in the field station environment, monkeys were presented with a delayed matching to sample task with a small image set.

## Subjects

Laboratory subjects were six 6–7 year old male rhesus macaque monkeys with a 6-month history with computerized cognitive testing. They were housed and tested as described for all Laboratory subjects, but did not participate in the other experiments. Five subjects were pair housed and one was housed with protected contact.

Field station subjects were 2 male and 5 female subjects aged 3–5 years (Mean = 3.7 years). Two subjects were members of the high-ranking group, 4 were members of the middleranking group, and 1 was a member of the low-ranking group. All subjects participated in the previous 3 experiments.

Subjects in both groups had extensive experience on matching-to-sample tasks at various delays with varying image set sizes at the start of the reported experiment.

## Stimuli and procedure

A set of four session-unique clip art images were used in each 100-trial test session, such that on the first trial of a session all images were novel, but by the end of the session all images were familiar to the subjects. To start a trial, monkeys touched (FR2) a green box at the bottom of the screen (Figure 10). A sample image then appeared in the center of the screen. Touches (FR2) to this image resulted in a blank screen delay, and then four images appeared in the four corners of the screen. A touch to the image identical to the sample was rewarded with a food pellet and a positive auditory reinforcer, whereas touches to any of the three incorrect comparison images resulted in a negative auditory stimulus and a black screen time out (15 seconds). Delays of 0.2, 6 12 24, and 48 seconds were counterbalanced and pseudorandomized within 100-trial sessions. A 3-second (Field station) or 30 second (Laboratory) inter-trial interval (ITI) separated each trial.

Monkeys in the Field station environment had the option to walk away from a trial or to move to a different computer during the delay. If a monkey moved to a new computer or there were no touches to the screen for 30 seconds after the delay, the trial aborted and the same trial was repeated when the monkey returned. The number of times a trial was aborted was recorded, but to better equate the task between the two groups, comparison analyses excluded repeated trials.

## Results and discussion

Monkeys in the two groups performed similarly. In the final 100 trial session, monkeys in both groups showed a significant decrease in performance at longer delays (RM ANOVA:  $F_{4,36} = 46.35$ ,  $p < 0.001$ ; Figure 11), but there was no difference between the two groups ( $F_{1,36} = 0.06$ ,  $p < 0.81$ ) or interaction between group and delay ( $F_{4,36} = 0.09$ ,  $p < 0.98$ ). Despite many possibilities for visual and auditory distraction during delays in the Field station environment, monkeys' performance across the delays did not differ from Laboratory subjects working in a more controlled environment. This indicates that animals in complex environments are viable research subjects even for tasks requiring attention over long delays. Field station monkeys were more likely to abort trials with longer delay lengths (RM ANOVA:  $F_{4,36} = 13.68$ ,  $p < 0.001$ ; Figure 11). Accuracy did not differ across delays when

repeated aborted trials were included or excluded from the data (RMANOVA: Main effect of repeated trials:  $F_{4,24} = 0.13, p < 0.73$ ; Interaction:  $F_{4,24} = 0.47, p < 0.76$ ). These results support previous findings that rhesus monkeys solve recognition memory tasks with long delays (e.g., Mishkin and Murray, 1998; Tu and Hampton, 2012), and that animals in complex environments perform well on tasks that require sustained attention (Fagot and De Lillo, 2011).

One of the factors limiting performance in matching-to-sample tests with small image sets is proactive interference (e.g. Wright et al. 1986). One of the ways proactive interference can occur in these tests is that subjects remember distracters used in the current trial from seeing them as samples on previous trials. It can therefore be difficult for the subjects to determine which image was seen most recently as the sample. Longer inter-trial-intervals (ITIs) can reduce this interference because longer ITIs cause forgetting of information from previous trials. Laboratory monkeys were trained on this matching-to-sample task for an unrelated memory study for which a 30s ITI was used. To decrease the probability that Field Station monkeys would lose interest and leave the testing area between trials, we used a shorter 3s ITI at the field station. This could have inflated the performance of the Laboratory group relative to the Field station group. However, monkeys in both environments worked at their own pace. Field station monkeys performed fewer trials per day, on average ( $M = 9.59$ ), than did Laboratory monkeys ( $M = 60.00$ ), thereby probably mitigating any effect of differences in ITI. It is never possible to precisely control the pace at which subjects complete trials, and this problem is exacerbated by the Field station testing environment.

## General discussion

Despite dramatic differences in housing, testing environment, sex, and ages of the subjects, Field station and Laboratory monkeys learned and performed similarly on psychophysical, classification, transitive inference, and memory tasks. This provides strong evidence that animals living in complex environments can be productive research subjects in a broad range of cognitive domains, and that data collected in this setting can be considered comparable to that obtained from laboratory housed monkeys, at least in many cases.

High ranking monkeys did not dominate the field station testing apparatus as might be expected in a despotic species like rhesus macaques (Drea 1998; Drea and Wallen 1999). Subjects from low, medium, and high ranking families were represented in all experiments. While subject numbers in the present experiment were too low to compare performance as a function of rank, future studies will use this testing system to test for rank related differences in learning rates, cognitive abilities, and social knowledge.

While monkeys from all age groups completed some trials on the Field station testing system, all the subjects that advanced far enough to participate in the four experiments presented here were juveniles and therefore did not have infants. As young animals are less engaged in mating, child rearing, and grooming than adult animals, the large number of young subjects in this study may be because these animals have more time available for testing. Additionally, young animals often show more interest in novel objects and are more likely to solve novel problems than adults, which may account for the higher rates of use of the testing system by juveniles (Biondi et al. 2010; Glickman and Sroges 1966; Morand-Ferron et al. 2011).

We have shown that testing socially housed monkeys in their home group produces data similar to that obtained in a laboratory environment in both perceptual and cognitive tasks. Social housing is often more cost effective and, in highly social species like rhesus monkeys, may produce experimental results that are more consistent with the natural

abilities of the research subjects. However, we did not see superior performance by the Field station subjects housed in a complex environment compared to the Laboratory subjects on the four tasks in this set of experiments. All the monkeys used in these studies had been raised in the Field station environment to at least 2.5 years of age, and this may be long enough for critical environment-dependent cognitive development to occur. Additionally, our laboratory testing conditions are not entirely typical. Our animals are tested in their home cages with familiar individuals around them, instead of being removed and isolated in a testing room. So these results should be generalized to other laboratory rearing and testing environments with caution.

While our data show that socially housed animals can be viable research subjects for cognitive testing, it is probably neither desirable nor feasible for testing in social settings to replace standard laboratory testing. One hundred percent of Laboratory subjects participated in our studies, while only 51% of non-infant Field station monkeys participated at all, and only 15% produced usable amounts of data in the cognitive tasks presented here. We were unable to collect data from even a single adult male monkey at the field station. While we did not find large differences in performance between our test groups as a result of such self selection, the current testing configuration does not allow for the study of very young animals or adult males. Identical experiments took much longer to conduct in the field station environment because subjects completed many fewer trials each day. The comparatively slow rate of data acquisition, combined with the comparatively low percentage of field station animals that participated, limits the efficiency of animal use in the Field station setting. Finally, we note that the type of social housing used here is not suitable for many climates or for research centers with more limited space. Laboratory studies are likely to remain the most efficient method for studying certain cognitive abilities. However, socially housed animals may be most appropriate for studying certain cognitive abilities, such as social cognition.

Our findings suggest that living in complex environments provides no advantage or disadvantage in the general cognitive processes tested here. However previous studies have found that life in complex social environments confers behavioral and reproductive advantages specifically in the social domain (Gersick, et al. 2012; Schwandt et al. 2010; White, et al. 2010). Positive effects of social housing on cognitive abilities may only be evident in tests of social cognition, which were not conducted in this study. Like wild animals, socially housed animals may possess social knowledge about individuals and relations among individuals in the group that is not present in monkeys living in less social settings. Field experiments have long relied on naturalistic stimuli and tasks to successfully study complex cognitive abilities that are difficult or impossible to study in a traditional laboratory environment due to subjects' limited social experience (Bergman et al. 2003; Biben and Bernhards 1994; Cheney and Seyfarth 1982; Cheney and Seyfarth 1999; Fischer 2004; Windfelder 2001; Zuberbuhler et al. 1999). The Field station testing environment will allow for cognitive studies that use similarly naturalistic stimuli and tasks. Future Field station studies will utilize this unique resource to explore social cognition, and to further probe cognitive differences between animals housed in different environments.

## Acknowledgments

The Yerkes National Primate Research Center is fully accredited by the American Association for Accreditation of Laboratory Animal Care. This project was funded by the National Center for Research Resources P51RR165, currently supported by the Office of Research Infrastructure Programs/OD P51OD11132, the Center for Behavioral Neuroscience under the STC Program of the NSF under Agreement IBN-9876754, a grant from the James S. McDonnell Foundation, NIH grant 1R01MH082819, NSF grants 0745573 and 1146316. Regina Paxton was supported in part by National Science Foundation grant DGE- 0231900.

We thank Dina P. Chou, Steven R.L. Sherrin, Jacey Jones, and Angela Tripp for help testing subjects. We thank Tom Hassett for help collecting dominance hierarchy data on Field station subjects, and the colony management and veterinary staff of the Yerkes Field Station for their gracious support of this work.

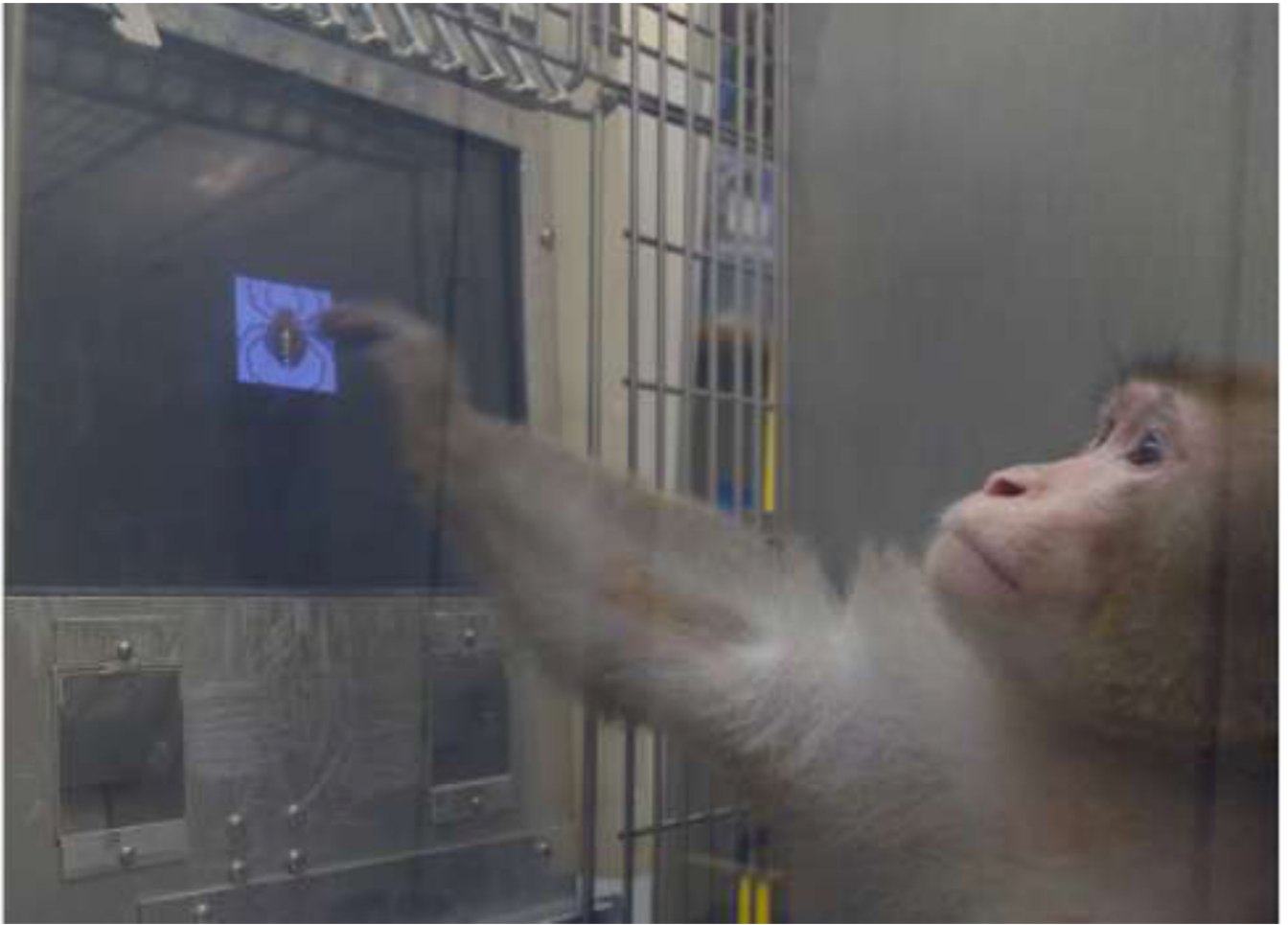
## References

- Adachi I, Chou DP, Hampton RR. Thatcher effect in monkeys demonstrates conservation of face perception across primates. *Curr Biol.* 2009; 19(15):1270–1273. [PubMed: 19559613]
- Andrews MW, Rosenblum LA. Automated recording of individual performance and hand preference during joystick task acquisition in group-living bonnet macaques (*Macaca radiata*). *J Comp Psychol.* 1994; 108:358–362. [PubMed: 7813194]
- Aron, A.; Aron, E. *Statistics for Psychology*. Upper Saddle River, NJ: Prentice Hall; 1999.
- Barbet I, Fagot J. Processing of contour closure by baboons (*Papio papio*). *J of Exp Psychol-Anim Behav Process.* 2011; 3:407–419. [PubMed: 22003965]
- Basile BM, Hampton RR. Rhesus monkeys (*Macaca mulatta*) show robust primacy and recency in memory for lists from small, but not large, image sets. *Beh Proc.* 2010; 83:183–190.
- Basile BM, Hampton RR. Monkeys recall and reproduce simple shapes from memory. *Curr Biol.* 2011; 21(9):774–778. [PubMed: 21530257]
- Basile BM, Hampton RR. Dissociation of active working memory and passive recognition in rhesus monkeys. *Cogn.* (in press).
- Bergman TJ, Beehner JC, Cheney DL, Seyfarth RM. Hierarchical classification by rank and kinship in baboons. *Sci.* 2003; 302:1234–1236.
- Bernstein, IS. Primate status hierarchies. In: Rosenblum, LA., editor. *Primate Behav.* New York: Academic Press; 1970.
- Biben M, Bernhards D. Naive recognition of chuck calls in squirrel monkeys (*Saimiri sciureus macrodon*). *Lang, Commun.* 1994; 14:167–181.
- Biondi LM, Bo MS, Vassallo AI. Inter-individual and age differences in exploration, neophobia and problem-solving ability in a Neotropical raptor (*Milvago chimango*). *Anim Cogn.* 2010; 15:701–710. [PubMed: 20300791]
- Bond A, Wei CA, Kamil AC. Cognitive representation in transitive inference: A comparison of four corvid species. *Behav Process.* 2010; 85:283–292.
- Bonte E, Flemming T, Fagot J. Executive control of perceptual features and abstract relations by baboons (*Papio papio*). *Behav Brain Res.* 2011; 222:176–182. [PubMed: 21435356]
- Cheney DL, Seyfarth RM. Recognition of individuals within and between groups of freeranging vervet monkeys. *Am Zool.* 1982; 22:519–529.
- Cheney DL, Seyfarth RM. Recognition of other individuals' social relationships by female baboons. *Anim Behav.* 1999; 58:67–75. [PubMed: 10413542]
- D'Amato MR, Van Sant P. The person concept in monkeys (*Cebus apella*). *J Exper Psychol-Anim Behav Process.* 1988; 14:43–55.
- De Vries H, Appleby MC. Finding an appropriate order for a hierarchy: a comparison of the I&SI and the BBS methods. *Anim Behav.* 2000; 59:239–245. [PubMed: 10640386]
- Drea CM. Social context affects how rhesus monkeys explore their environment. *Am J Primatol.* 1998; 44:205–214. [PubMed: 9519240]
- Drea CM, Wallen K. Low-status monkeys “play dumb” when learning in mixed social groups. *Proc Nat Acad Sci.* 1999; 96:12965–12969. [PubMed: 10536031]
- Fagot J, Bonte E. Automated testing of cognitive performance in monkeys: Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (*Papio papio*). *Behav Res Methods.* 2010; 42:507–516. [PubMed: 20479182]
- Fagot J, De Lillo C. A computerized study of working memory: Immediate serial spatial recall in baboons (*Papio papio*) and humans. *Neuropsychologia.* 2011; 49:3870–3880. [PubMed: 22015260]
- Fagot J, Thompson RKR. Generalized relational matching by guinea baboons (*Papio papio*) in two-by-two-item analogy problems. *Psychol Sci.* 2011; 22(10):1304–1309. [PubMed: 21934135]

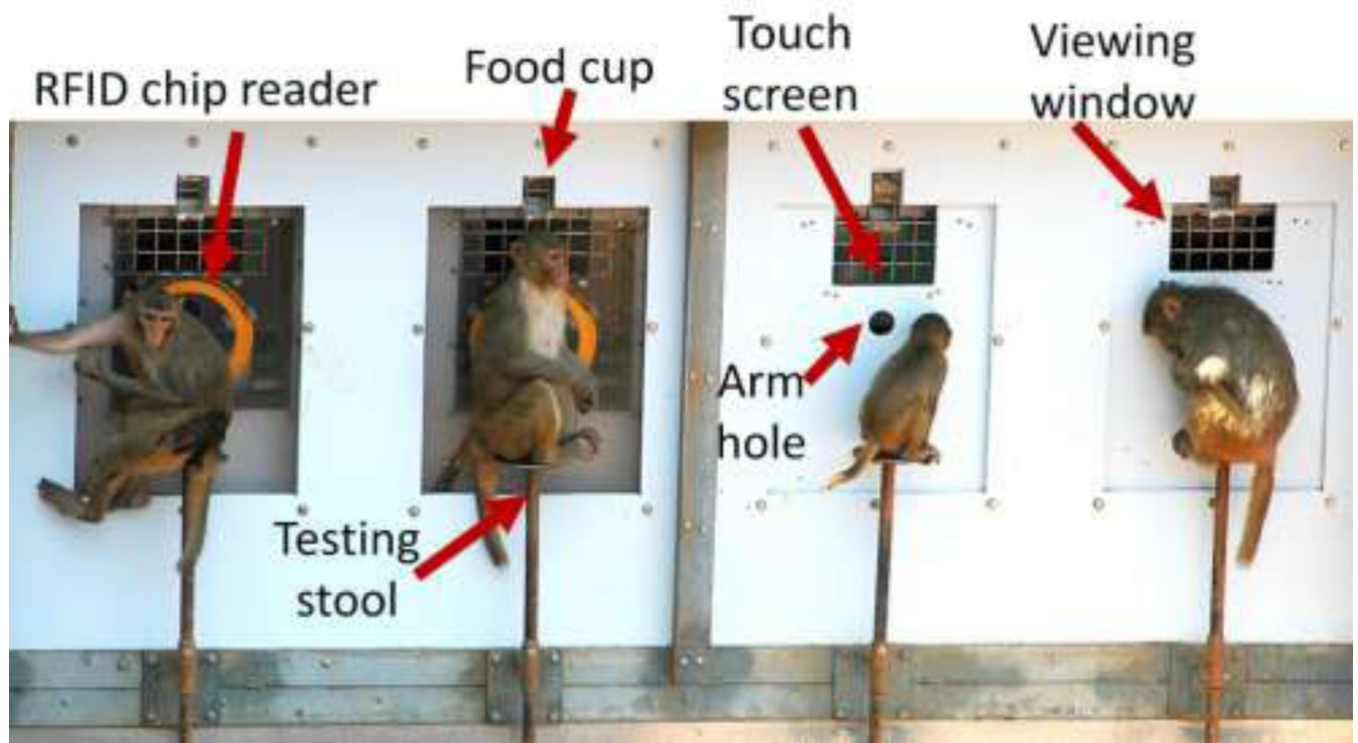
- Fagot J, Paleressompouille D. Automatic testing of cognitive performance in baboons maintained in social groups. *Behav Res Methods*. 2009; 41:396–404. [PubMed: 19363180]
- Fischer J. Emergence of individual recognition in young macaques. *Anim Behav*. 2004; 67:655–661.
- Gazes RP, Hampton RR, Chee NW. Cognitive mechanisms for transitive inference performance in rhesus monkeys: Measuring the influence of associative strength and inferred order. *J Exp Psychol: Anim Behav Process*. 2012; 38(4):331–345. [PubMed: 23066978]
- Gersick AS, Snyder-Mackler N, White DJ. Ontogeny of social skills: social complexity improves mating and competitive strategies in male brown-headed cowbirds. *Anim Behav*. 2012; 83:1171–1177.
- Glickman SE, Sroges RW. Curiosity in zoo animals. *Behav*. 1966; 26:151–188.
- Grainger J, Dufau S, Montant M, Ziegler JC, Fagot J. Orthographic processing in baboons (*Papio papio*). *Sci*. 2012; 336:245–248.
- Grosenick L, Clement TS, Fernald RD. Fish can infer social rank by observation alone. *Nat*. 2007; 445:429–432.
- Hasset, JM.; Martin-Malivel, J.; Lange, H.; Fischer, A.; Wallen, K. Age and rank influences on access to an automated system for cognitive testing in socially housed rhesus monkeys. Paper presented at the International Conference on Comp Cogn; Melbourne, FL. 2007.
- Henderson J, Hurly TA, Healy SD. Rufous hummingbirds' memory for flower location. *Anim Behav*. 2001; 61:981–986.
- Lazareva OF, Smirnova AA, Bagozkaja MS, Zorina ZA, Rayevsky VV, Wasserman EA. Transitive responding in hooded crows requires linearly ordered stimuli. *J Exp Analysis Behav*. 2004; 82:1–19.
- Logie RH. Visuo-spatial processing in working memory. *Q J Exp Psychol Sec A*. 1986; 38:229–247.
- MacLean EL, Merritt DJ, Brannon EM. Transitive inference in two species of prosimian primates. *Am J of Primatol*. 2007; 69:102–102.
- McGonigle BO, Chalmers M. Are monkeys logical. *Nat*. 1977; 267:694–696.
- Mishkin M, Murray E. Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *J Neurosci*. 1998; 18(16):6568–6582. [PubMed: 9698344]
- Morand-Ferron J, Cole EF, Rawles JEC, Quinn JL. Who are the innovators? A field experiment with 2 passerine species. *Behav Ecol*. 2011; 22:1241–1248.
- Paxton R, Basile BM, Adachi I, Suzuki WA, Wilson M, Hampton RR. Rhesus monkeys (*Macaca mulatta*) rapidly learn to select dominant individuals in videos of artificial social interactions between unfamiliar conspecifics. *J Comp Psychol*. 2010; 124(4):395–401. [PubMed: 20853947]
- Phillips WA, Christie DFM. Interference with visualization. *Q J Exp Psychol*. 1977; 29:637–650. [PubMed: 601182]
- Piaget, J. *Logic and Psychology*. New York: Basic Books, Inc; 1960.
- Rapp PR, Kansky MT, Eichenbaum H. Learning and memory for hierarchical relationships in the monkey: Effects of aging. *Behav Neurosci*. 1996; 110(5):887–897. [PubMed: 8918992]
- Roberts WA, Phelps MT. Transitive inference in rats: A test of the spatial coding hypothesis. *Psychol Sci*. 1994; 5:368–374.
- Rommeck I, Capitanio JP, Strand SC, McCowan B. Early social experience affects behavioral and physiological responsiveness to stressful conditions in infant rhesus macaques (*Macaca mulatta*). *Am J of Primatol*. 2011; 73:692–701. [PubMed: 21462233]
- Schrier AM, Angarella R, Povar ML. Studies of concept formation by stump-tailed monkeys: Concepts humans, monkeys, and letter A. *J Exp Psychol: Anim Behav Process*. 1984; 10:564–584.
- Schrier AM, Brady PM. Categorization of natural stimuli by monkeys (*Macaca mulatta*): Effects of stimulus set size and modification of exemplars. *J Exp Psychol: Anim Behav Process*. 1987; 13:136–143. [PubMed: 3572306]
- Schwandt ML, Lindell SG, Sjöberg RL, Chisholm KL, Higley JD, Suomi SJ, et al. Gene-environment interactions and response to social intrusion in male and female rhesus macaques. *Biol Psychiatry*. 2010; 67:323–330. [PubMed: 20015482]
- Shettleworth, SJ. *Cognition, Evolution, and Behavior*. Second Edition. New York: Oxford University Press; 2009.

- Templer VL, Hampton RR. Rhesus monkeys (*Macaca mulatta*) show robust evidence for memory awareness across multiple generalization tests. *Anim Cogn.* 2012; 15:409–415. [PubMed: 22183204]
- Treichler FR, Raghanti MA. Serial list combination by monkeys (*Macaca mulatta*): test cues and linking. *Anim Cogn.* 2010; 13:121–131. [PubMed: 19533183]
- Treichler FR, Van Tilburg D. Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List linking. *J Exper Psychol: Anim Behav Proc.* 1996; 22:105–117.
- Tu HW, Hampton RR. One-trial memory and habit contribute independently to matching-to-sample performance in rhesus monkeys (*Macaca mulatta*). *J Comp Psychol*, Advance online publication. 2012
- Von Fersen L, Wynne CDL, Delius JD, Staddon JER. Transitive inference formation in pigeons. *J of Exp Psychol-Anim Behav Process.* 1991; 17:334–341.
- Washburn DA, Astur RS. Nonverbal working memory of humans and monkeys: Rehearsal in the sketchpad? *Mem, Cogn.* 1998; 26:277–286.
- Weiß BM, Kehmeier S, Schloegl C. Transitive inference in free-living greylag geese, *Anser anser*. *Anim Behav.* 2010; 79:1277–1283.
- White DJ, Gersick AS, Freed-Brown G, Snyder-Mackler N. The ontogeny of social skills: Experimental increases in social complexity enhance reproductive success in adult cowbirds. *Anim Behav.* 2010; 79:385–390.
- White DJ, Gersick AS, Snyder-Mackler N. Social networks and the development of social skills in cowbirds. *Philos Trans R Soc B-Biol Sci.* 2012; 367:1892–1900.
- Windfelder TL. Interspecific communication in mixed-species groups of tamarins: Evidence from playback experiments. *Anim Behav.* 2001; 61:1193–1201.
- Wright, AA.; Urcuioli, RJ.; Sands, SF. Proactive interference in animal memory. In: Kendrick, DF.; Riling, M.; Denny, R., editors. *Theories of Animal Memory*. Hillsdale, NJ: Earlbaum; 1986. p. 101-125.
- Zuberbuhler K, Cheney DL, Seyfarth RM. Conceptual semantics in a nonhuman primate. *J of Comp Psychol.* 1999; 113:33–42.



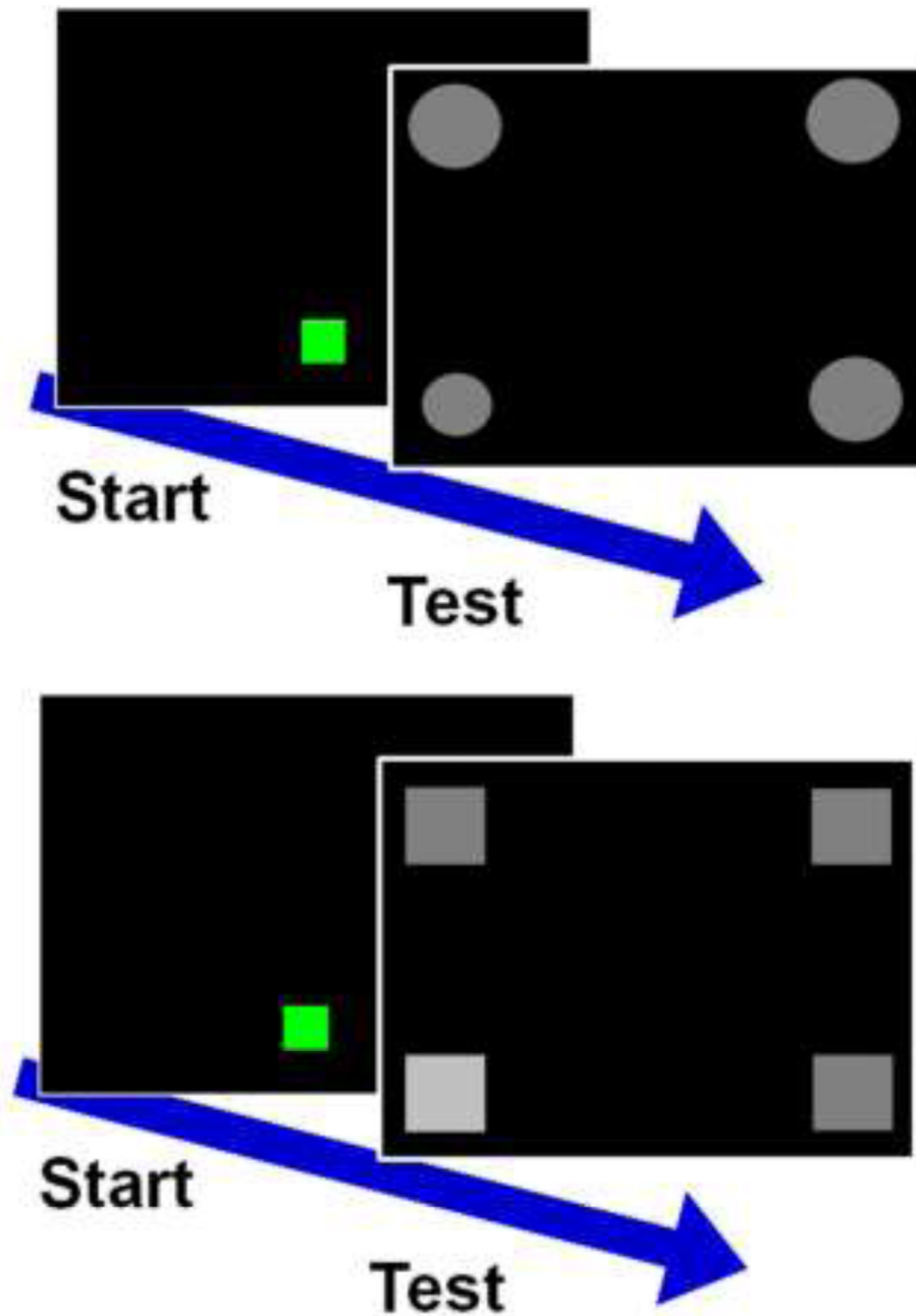


**Fig 1.**  
The cognitive testing system used with Laboratory monkeys. A touch screen computer was hung and locked to the front of an animal's home cage and the door was raised. Food rewards for correct responses were dispensed in cups below the screen.

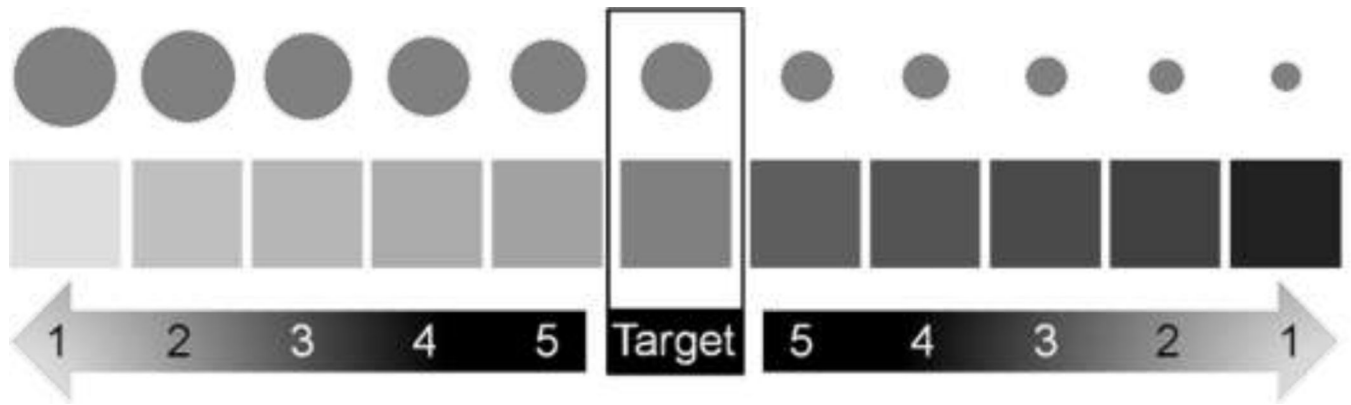


**Fig 2.**

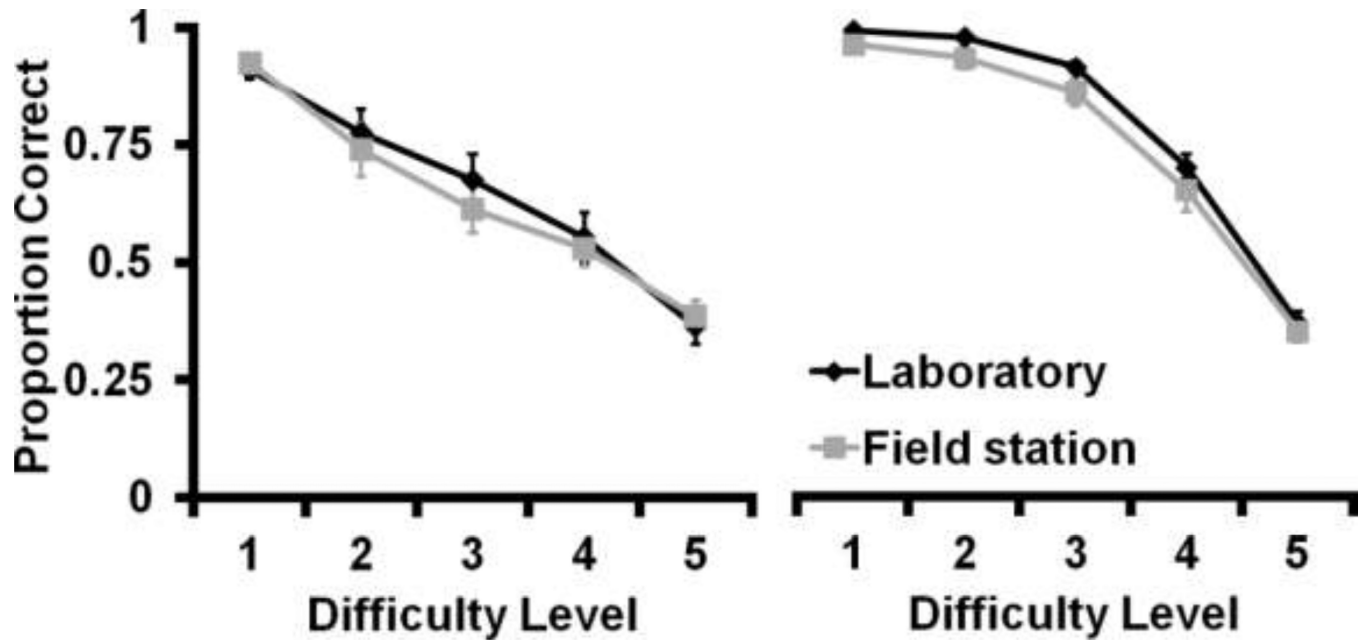
The four testing stations used by the field station subjects. Monkeys sat on a stool and could view the touch screen through the mesh window. When monkeys reached through the arm hole to touch the screen, the RFID chip reader sent a unique code for that monkey to the connected computer. The computer selected the appropriate test trial for the subject, and images appeared on the touch screen. Correct responses were reinforced with a distinctive sound and a food reward. Food rewards were dispensed into a receptacle above the viewing window.



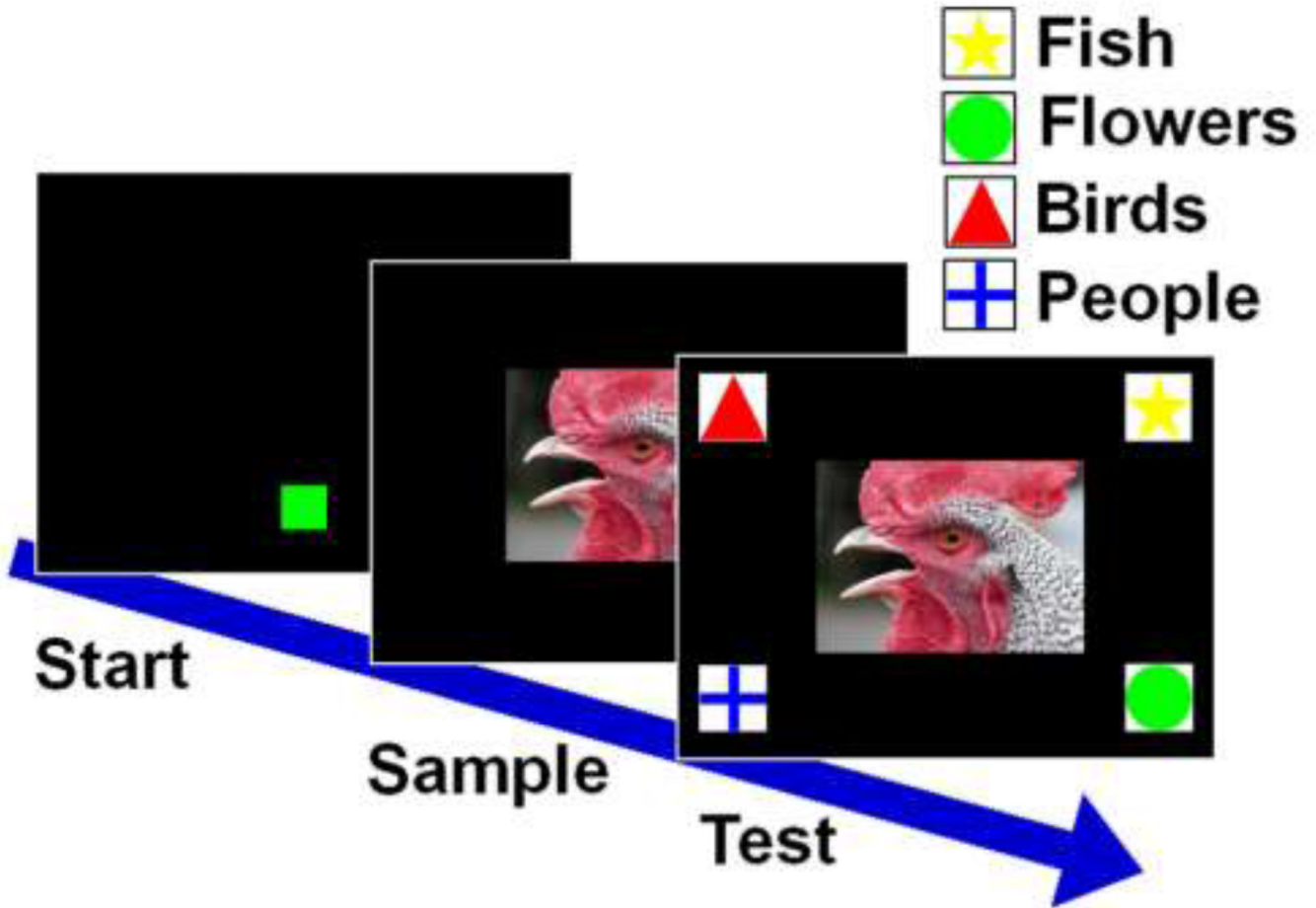
**Fig 3.** Trial progression during the size discrimination (*top*) and brightness discrimination (*bottom*). Monkeys were required to select the target stimulus from among three identical distracters that differed in either size (size discrimination task) or brightness (brightness discrimination task) from the target. In both of these examples, the target stimulus appears in the lower left corner of the screen. In the task, the target location varied pseudo-randomly among all 4 corners.



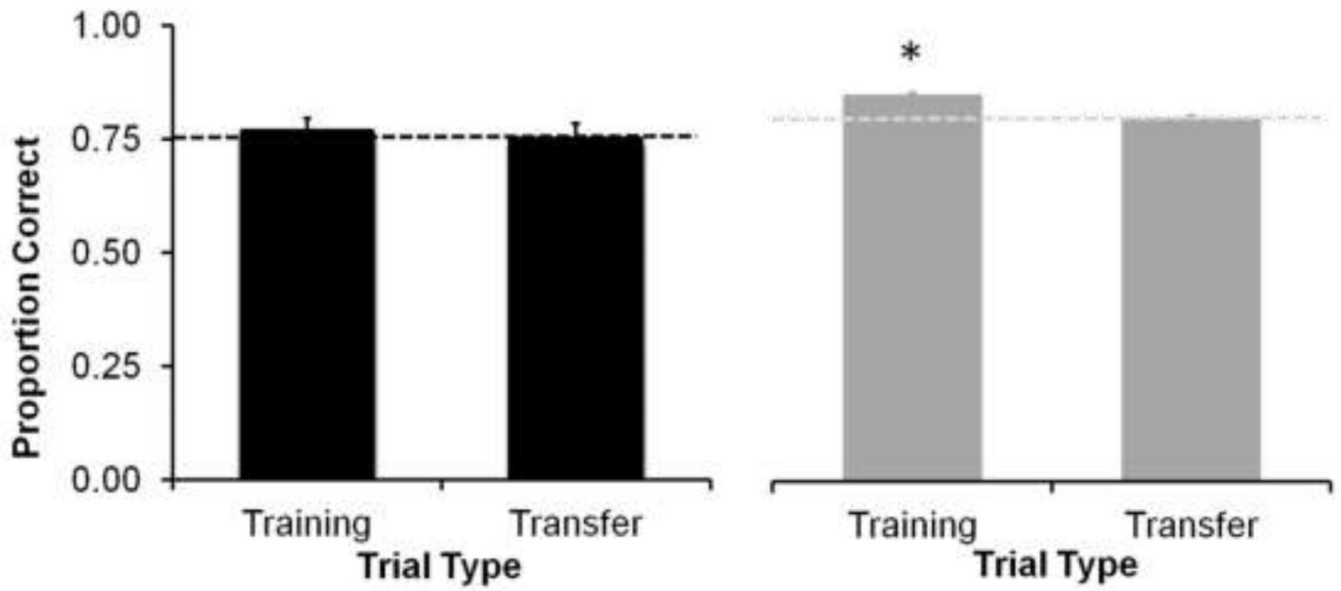
**Fig 4.** Stimuli used in the size discrimination (*top row*) and brightness discrimination (*middle row*). The target stimulus remained the same throughout each task (*center column*). Distracter stimuli differed from the target in two directions along five levels of difficulty (*bottom*). The distracters associated with the higher values (e.g. Level 5) were most difficult to discriminate from the target. Pictured brightness and sizes are not actual stimuli used in the experiment.



**Fig 5.** Mean proportion correct on the two criterion sessions across the five discrimination difficulty levels in the size discrimination (*left*) and the brightness discrimination (*right*) by the Laboratory (black line) and Field station (grey line) groups. Error bars indicate  $\pm$  standard error of the mean.

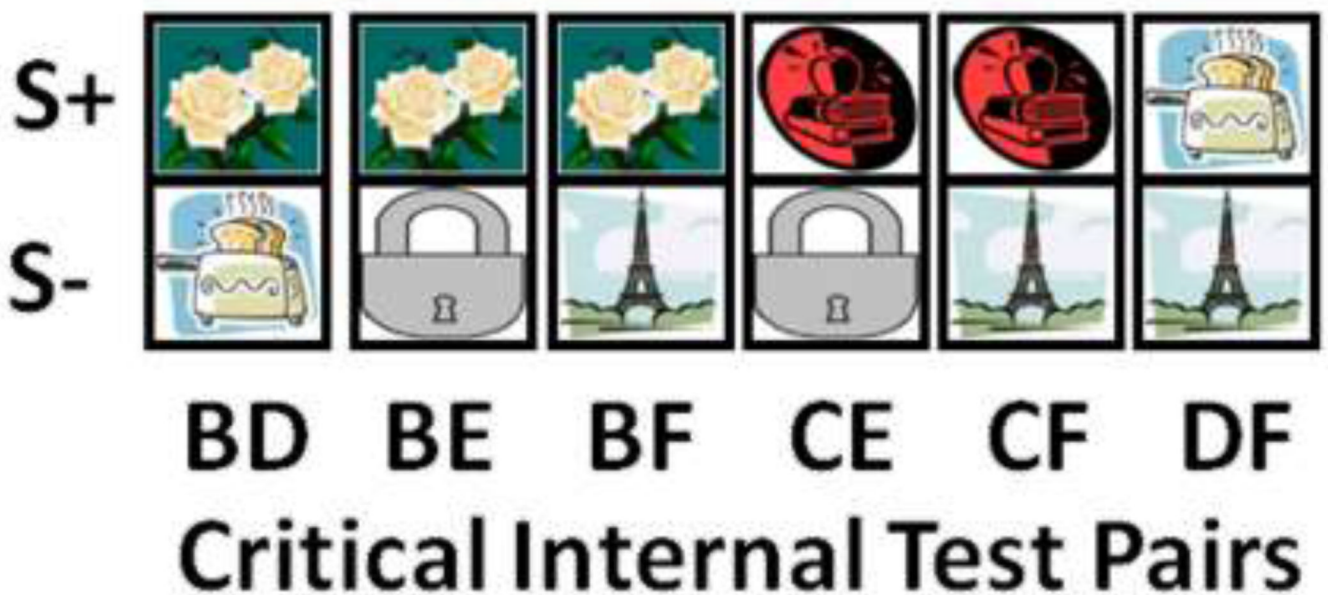
**Fig 6.**

Example trial from the perceptual classification experiment. Subjects began a trial by pressing the green start box. A sample image from one of the four categories (birds, fish, flowers, people) appeared in the center of the screen. After the sample was touched, the four classification icons appeared in the four corners of the screen. Selection of the correct icon, in this case the triangle, was rewarded.



**Fig 7.** Mean proportion correct during the one transfer session by the Laboratory (black bars) and Field station (grey bars) groups. Dotted lines indicate criterion level. Performance did not differ from criterion level on transfer images in either group, but the Field station group performed significantly above criterion with trained images in the transfer session (\* indicates above criterion performance). Error bars indicate + standard error of the mean.

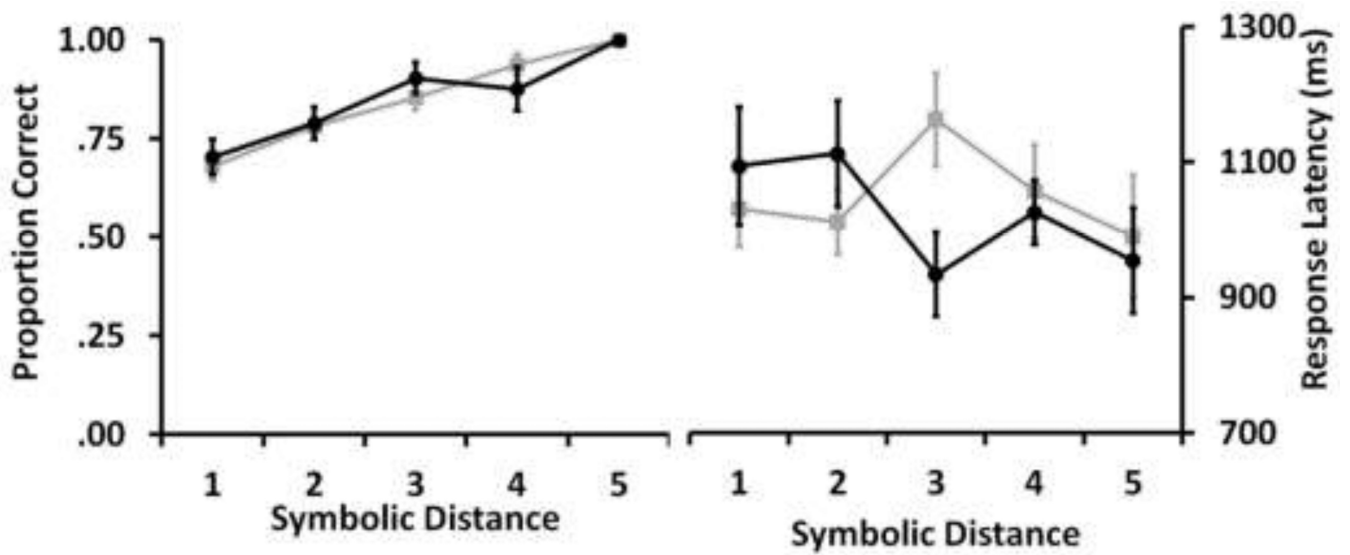
## Adjacent Training Pairs



**Fig 8.**

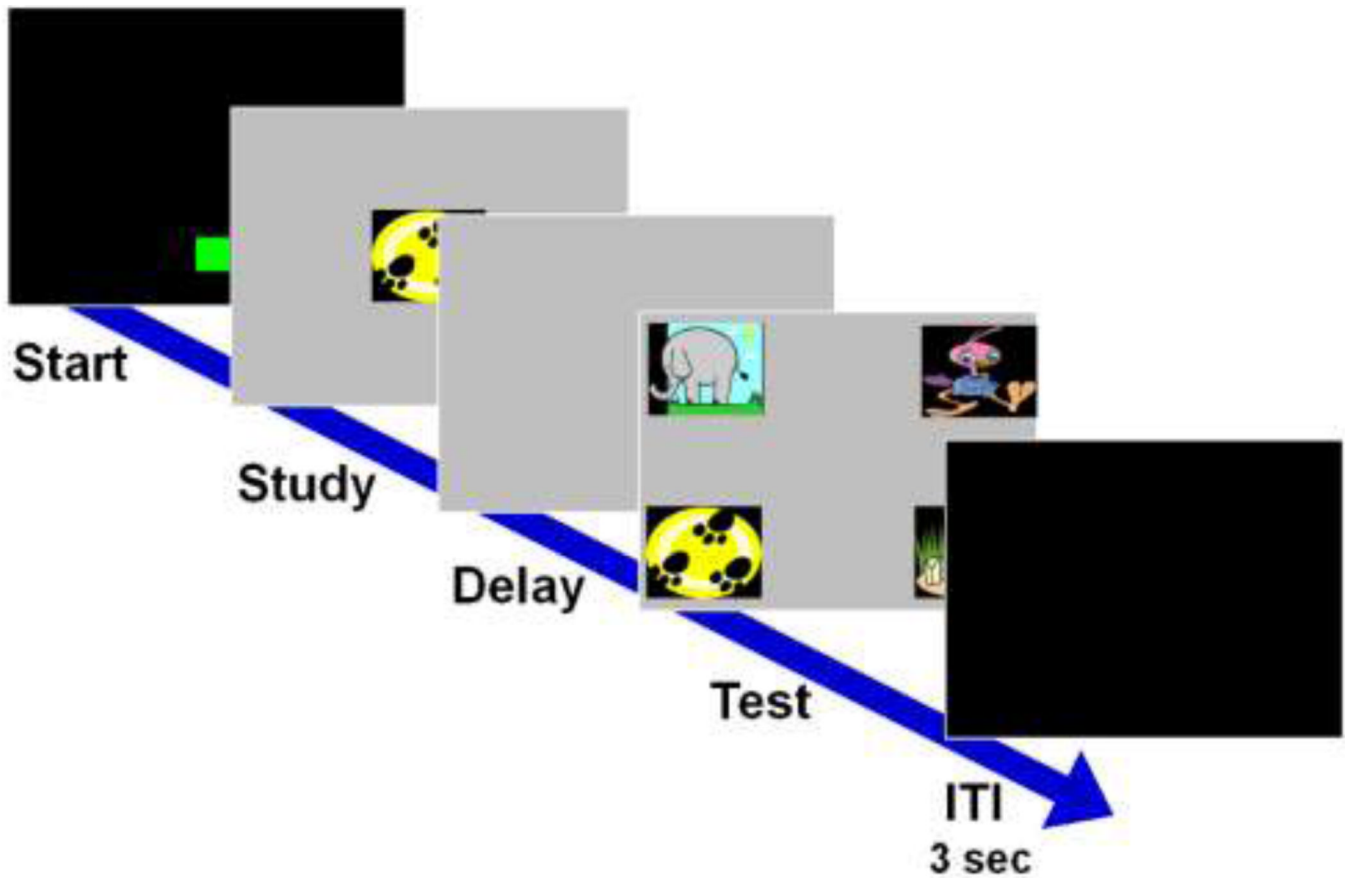
Adjacent training pairs (*top*) and critical non-adjacent internal test pairs (*bottom*) used in the transitive inference experiment. During training, correct selection of the S+ in a given pair resulted in an auditory reinforcer paired with a food reinforcer on 80% of trials. On trials containing test pairs all choices resulted in an auditory reinforcer only.





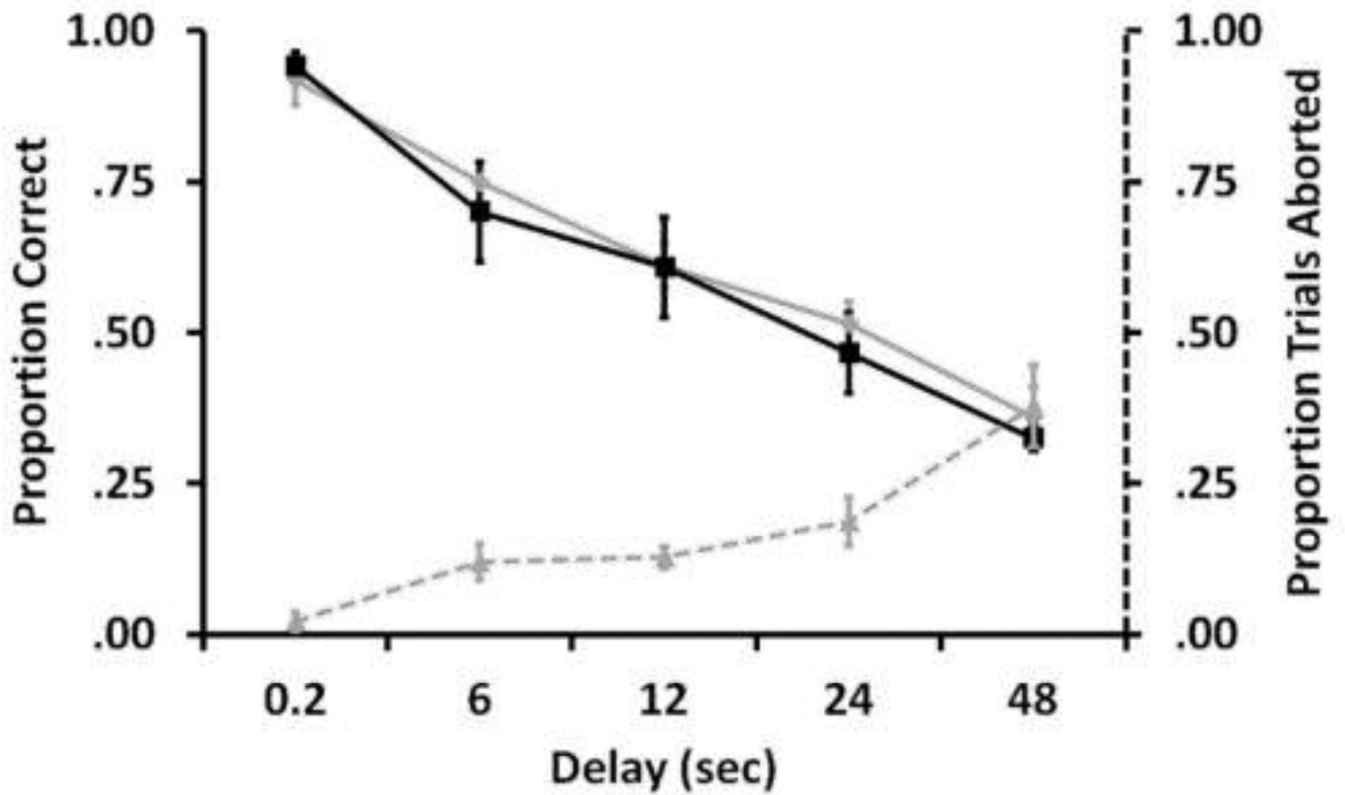
**Fig 9.**

*Left.* Accuracy was higher when the symbolic distance between the two images in the test pair was larger for both the Laboratory (black lines) and Field station (grey lines) subjects. *Right.* Median response latency on correct trials did not vary systematically with symbolic distance for subjects in either group. Error bars indicate  $\pm$  standard error.



**Fig 10.**

Example trial from the memory experiment. Subjects began a trial by pressing the green start box. One of the four possible sample images appeared in the center of the screen. After the sample was touched, a blank screen appeared for the programmed memory interval. The four choice images then appeared in the four corners of the screen. Selection of the item identical to the sample image was rewarded with an auditory and food reinforcer.



**Fig 11.** Accuracy on a small set delayed matching to sample task by subjects in the Laboratory (black line) and Field station (grey line) groups across the 5 delays tested. Dashed grey line indicates proportion of trials aborted across delays by animals in the Field Station group (right axis). Error bars indicate  $\pm$  standard error.

**Table 1**

Proportion of individuals in each age, sex, and rank group that worked on the touch screen system at the field station. The total number of individuals in each demographic group is shown italicized in parentheses.

	Juveniles				Adults			
	High	Middle	Low	High	Middle	Low	Total	
	<i>.60 (43)</i>				<i>.35 (34)</i>			
Male	<i>.5 (4)</i>	<i>1.0 (3)</i>	<i>0.0 (2)</i>	<i>0.0 (2)</i>	---	<i>0.0 (2)</i>	<i>.38 (13)</i>	
Female	<i>.64 (14)</i>	<i>.80 (10)</i>	<i>.40 (10)</i>	<i>.30 (10)</i>	<i>.40 (10)</i>	<i>.50 (10)</i>	<i>.51 (77)</i>	