Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________ ______________
Regina Paxton Date

Evolution of social cognition and the cognitive bases of transitive inference in monkeys

By

Regina Paxton

M.A., Emory University, 2007

B.S., Bucknell University, 2004

Department of Psychology

Robert Hampton
Advisor

Jocelyne Bachevalier
Committee Member

Harold Gouzoules
Committee Member

Philippe Rochat
Committee Member

Mark Wilson
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____________________ Date

Evolution of social cognition and the cognitive bases of transitive inference in monkeys

By

Regina Paxton

M.A. Emory University, 2007

B.S., Bucknell University, 2004

Advisor: Robert Hampton, PhD.

An abstract of a dissertation submitted to the Faculty
of the James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Psychology

2012

Abstract

Evolution of social cognition and the cognitive bases of transitive inference in monkeys

By

Regina Paxton

The social intelligence hypothesis posits that life in complex social groups created selection pressures that resulted in the evolution of sophisticated cognitive adaptations specialized to each species' life history. These cognitive abilities may be most readily engaged and measured under conditions that approximate a species' natural environment. I validated a novel method for cognitive testing in animals housed in a semi-natural social group by showing that these subjects learn and perform similarly to laboratory animals on a battery of cognitive tasks including perceptual discrimination, classification, memory, and transitive inference (TI). Transitive inference is a cognitive ability theorized to have been under selection pressures related to learning dominance relations. For example, if animal *A* is dominant to *B (A>B)* and *B* is dominant to *C (B>C)*, the relation between animals *A* and *C* can be logically inferred (*A>C*). TI can also serve non-social purposes; many species solve nonsocial laboratory TI tasks in which the relations between items are defined by reinforcement contingencies (if *A+B-*, and *B+C-*, then *A+C-*). However, because the relations are trained by reinforcement, it is unclear whether non-social TI performance is controlled by inference or by associative values accrued to individual stimuli. TI based on associative values would not support learning dominance relationships, because observation of social interactions is not followed by explicit reinforcement. Rhesus monkeys were presented with a series of tests to determine the contributions of associative values and logical inference to TI performance. We found that associative value can influence TI choice in extreme circumstances, but the inferred order of stimuli better explains TI choice under normal testing conditions. Measurement of the associative values of TI stimuli showed that these values did not correspond to performance on TI tests. Monkeys linked two previously learned 7-item TI lists into one 14 item list after training on just one linking pair, performance that cannot be explained by associative value alone. The viability of socially housed animals as cognitive test subjects will allow for future studies that examine the extent to which monkeys use TI to learn natural dominance hierarchies.

Evolution of social cognition and the cognitive bases of transitive inference in monkeys

By

Regina Paxton

M.A. Emory University, 2007

B.S., Bucknell University, 2004

Adviser: Robert Hampton, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Psychology

2012

Table of Contents

Figures and Tables

Chapter 1 Introduction

The social intelligence hypothesis posits that life in complex social groups created a suite of selection pressures that resulted in the evolution of sophisticated cognitive adaptations (Byrne & Whiten, 1997; Emery et al., 2007; Humphrey, 1976; Kamil, 2004). 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 a species natural environment. Most studies of cognition are conducted in laboratories that lack many of the social and physical challenges present in nature. Laboratory conditions afford excellent control over animals' experiences, and facilitate studies of processes such as learning that may be difficult or impossible to explore through observation of wild animals. However, laboratory environments also limit subjects' exposure to natural cognitive demands such as complex social interactions and foraging, and may deprive them of experiences necessary for normal cognitive development (Rommeck et al., 2011). Animals raised in more complex natural physical and social conditions may therefore exhibit a broader and more developed range of cognitive abilities. It would be advantageous to develop study settings that provide semi-natural living conditions while preserving the experimental control afforded by standard laboratory settings.

Studies of social cognition and related cognitive processes in monkeys would benefit particularly from the use of socially housed subjects. The natural social lives of rhesus monkeys are complex, and behavior is governed by a stable linear dominance hierarchy in which animal *A* dominates all others, *B* dominates all except *A*, etc., such that $A>B>C>D$ (Appleby, 1983; Barchas & Mendoza, 1984; Chase, 1974, 1980; Jackson & Winnegrad, 1988). The ability of group members to recognize their position in a social hierarchy and behave accordingly maintains stable relations and increases individual fitness. Appropriate submission and aggression are critical for the acquisition of social status, reproductive success, and survival in Old World primates (de Waal & Luttrell, 1985; Dittus, 1980; Fairbanks & McGuire, 1984; Silk et al., 2003). Physical cues such as size and age often do not reliably predict rank (Barchas & Mendoza, 1984; Chapais, 1988; Chase, 1982; Holekamp et al., 2007), suggesting that stereotyped appearances and behaviors are not sufficient for rapid learning of a hierarchy (Chapais, 1983; Cheney & Seyfarth, 1990). Socially housed animals must therefore learn the unique dominance hierarchy of their specific social group through experience and observation.

A linearly organized hierarchy results in dominance relationships that are transitive, such that if *A* is dominant to *B*, and *B* is dominant to *C*, then *A* will be dominant to *C* (Chase, 1982, 1986). Selection would favor cognitive mechanisms for rapid, safe, and flexible learning of dominance hierarchies, particularly when hierarchies are large and the cost of aggression high, as in rhesus monkeys. One useful cognitive mechanism for learning this type of social hierarchy is transitive inference (TI). TI is the ability to infer the relation between two items or individuals in a linear hierarchy based on relations with a third item or individual. TI could be used to construct an entire dominance hierarchy from the third party perspective through observation, avoiding many costly aggressive encounters (Cheney & Seyfarth, 1990). An animal capable of such inference could place itself within the hierarchy after only a small number of first party interactions.

In rhesus monkeys, pressure to rapidly learn new dominance relationships may have been particularly strong on males, who leave their natal group in early puberty and later join an established group (Thierry, 2007). The ability to infer social relationships from a small set of observations would be particularly useful for rapid learning of large numbers of new dominance relationships when entering a new group. A female rhesus monkey remains in her natal group and after puberty will rank near her mother and sisters (Chapais, 1988; Janus, 1992). Therefore she can learn her own place in the dominance hierarchy without inference through social experiences gained while closely following her mother in the first years of life. However, females do need to respond flexibly to changes in the normally stable hierarchy order that occur during their lifetime. The most frequent hierarchy changes in macaque social groups occur after matrilinial overthrows, where one family group takes over the rank of another family group without altering the individual ranks within a matriline (Figure 1, right; Cheney & Seyfarth, 1990; Ehardt & Bernstein, 1986; Gouzoules, 1980). Therefore, a monkey could potentially correctly infer the new ranks of all members of the affected matrilines from observing only a small set of interactions between individuals from those families. Rhesus monkeys of both sexes may have been under selection pressure to rapidly infer new dominance relationships and the results of major hierarchy shifts.

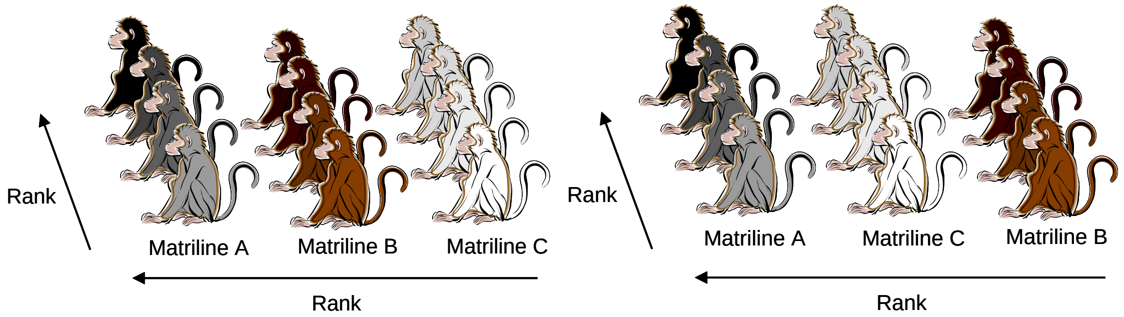

Figure 1. *Left*. Three matrilines in a typical rhesus monkey group. Arrows indicate increasing rank. *Right*. The same three matrilines after matriline C overthrows B. Within a matriline all members can maintain their relative rank after an overthrow, but every member of C now outranks every member of B.

TI may also be employed in non-social contexts. In non-social laboratory tests of TI, subjects learn overlapping adjacent premise pairs of stimuli (*AB*, *BC*, *CD*, *DE*) such that stimulus *A* is rewarded when paired with *B*, *B* is rewarded when paired with *C*, etc., creating an ordered list of stimuli (*A*>*B*>*C*>*D*>*E*). Critical tests use novel non-adjacent pairs (e.g. *AC, AD, AE, BD, BE, CE*). Because subjects have not seen these items together before, correct choice on such trials requires novel combination of knowledge acquired during training. Monkeys, pigeons, geese, rats, crows, humans, and many other species solve non-social laboratory TI tasks by correctly choosing the inferred higher ranked stimulus on these non-adjacent test trials (Bryant & Trabasso, 1971; Davis, 1992a; McGonigle & Chalmers, 1977; Roberts & Phelps, 1994; Weiβ et al., 2010; Wynne, 1995).

However all studies of non-social TI in nonhumans have used food reinforcement during training (Vasconcelos, 2008). Computational models suggest that the associative values accrued to the individual stimuli during training can control responding on non-adjacent tests and generate a TI-like pattern (Vasconcelos, 2008; Von Fersen et al., 1991; Wynne, 1997; Zentall & Sherburne, 1998). In these models the associative values of each TI stimulus increases and decreases as the items are reinforced and non-reinforced during training. On test trials, subjects select the item with the higher associative value. Due to differences in the number of errors committed during acquisition of the adjacent pairs during training, stimuli may gain values that follow the same order as would be determined through inference, making it difficult to distinguish between use of logical inference and associative values in TI test performance based solely on test trial performance in nonhuman animals.

Associative value mechanisms would be unlikely to support dominance hierarchy learning in nature, where observed third-party dominance interactions are not followed by explicit reinforcement. Elegant studies using controlled, live social stimuli and no food reinforcement have found that cichlid fish (*A. burtoni*), pinyon jays (*G. cyanocephalus*), and chickens (*G.domesticus*) apply TI to dominance relations, reacting appropriately on the first encounter with an individual they had only observed engaging in dominance interactions (Grosenick et al., 2007; Hogue et al., 1996; Paz-y-Miño et al., 2004). Chickens and jays appropriately display subordinate behavior during their first interaction with an individual they observed dominating an individual known to be dominant to themselves (Hogue, et al., 1996; Paz-y-Miño, et al., 2004). Cichlid fish that have a natural preference for interacting with lower ranking individuals, choose to interact with an individual inferred to be lower ranking in a hierarchy constructed from an observed series of 4 dyadic social interactions (Grosenick, et al., 2007). While observing high ranking individuals may be rewarding (Deaner et al., 2005), dominant behaviors can also be treated as unpleasant or negative and may be punishing (Marsh et al., 2009). The values at play in social situations are likely much more complex than those in nonsocial laboratory tests, making it difficult for the current associative value based accounts to explain these results.

Many arguments for associative accounts of TI have been advanced on the basis of mathematical modeling. There is extensive modeling evidence that associative values can account for TI performance in pigeons (Siemann, Delius, & Wright, 1996; Steirn et al., 1995; Von Fersen, et al., 1991; Wynne, 1997), but these models are unable to account for TI performance in all species. Associative value models fail to predict performance on non-adjacent test trials in four corvid species tested, but partially account for adjacent pair performance patterns in the non-social Scrub jays (Bond et al., 2010). If transitive inference evolved in response to selection pressures for learning dominance relationships, highly social species, including rhesus monkeys, should solve TI tasks using mechanisms that can be applied to learning dominance information; they should rely on inference instead of the associative values of the stimuli. There have been no studies modeling TI data in monkeys, therefore the theoretical contributions of inference and associative values to TI performance in monkeys is not known. Only one study explicitly measured the associative value of TI stimuli to empirically test the predictions of the computational models, and it was found that in pigeons, associative values do not predict TI performance (Lazareva & Wasserman, 2012). This type of measurement is needed, but has not been conducted in monkeys.

In the next chapter I will assess the validity of a novel method for cognitive testing in animals housed in a semi-naturalistic social group by determining whether these subjects perform comparably to laboratory animals on cognitive tasks, including TI. This lays the groundwork for future studies that will explore the content and structure of naturally acquired dominance knowledge; questions unanswerable in traditional laboratory housed subjects. In Chapter 3 I will address the cognitive mechanisms underlying non-social transitive inference performance in monkeys to determine if the methods used to solve non-social transitive inference tasks could also be recruited to learn real world dominance relationships. This will advance our understanding of the evolution and cognitive bases of TI and of cognition in general. Finally, in chapter 4 I will summarize my findings with respect to the social intelligence hypothesis and the evolution of cognition, and lay out plans for future studies that will empirically test the predictions from my findings.

Chapter 2. Comparison of automated cognitive test performance by monkeys in laboratory and large semi-natural social groups

Animals live complex 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 a species natural environment. 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 animals' experiences, and facilitate studies of processes such as learning that may be difficult or impossible to explore through observation of wild animals. However, laboratory environments also limit subjects' exposure to natural cognitive demands, and may deprive them 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 broader and more developed range of cognitive abilities. It would be advantageous to develop study settings that provide semi-natural living conditions while preserving much of the 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 (2007), and Fagot (2009) have created cognitive testing systems that use RFID chips to indentify individual monkeys as they complete computerized motor, perceptual, and cognitive tasks. Studies in these environments have shown that monkeys living in small groups perform trials at a steady rate and perform well on motor and cognitive tasks (motor: Andrews & Rosenblum, 1994; visual search: Barbet & Fagot, 2011; Bonte et al., 2011; Fagot & Bonte, 2010; working memory: Fagot & De Lillo, 2011; complex matching: Fagot & Paleressompoulle, 2009), indicating the feasibility of this method of testing.

However, no studies have compared learning and performance between monkeys living in complex environments and monkeys living in traditional laboratory environments. Performance between these groups of monkeys might differ for at least two reasons. First, as indicated earlier, semi-natural environments may foster superior cognitive development leading to differences in central cognitive capacities and thus performance. Second, differences in housing conditions may have peripheral effects on performance 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 & Wallen, 1999). These factors have the potential to prevent acquisition of reliable cognitive data.

To assess the viability of using monkeys housed in complex social environments as subjects for cognitive research, we compared learning and performance on 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 were used to compare visual perceptual function across the two environments. A transitive inference test was used to assess complex cognitive function. Finally, we used delayed-matching-to-sample to compare memory performance.

Subjects and apparatus

Subjects were rhesus macaque monkeys (*Macaca mulatta*) housed at the Yerkes National Primate Research Center in Atlanta 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 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 the age of approximately 2.5 years. Whenever possible, monkeys were pair-housed. 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 water. All subjects had an extensive history with computerized cognitive testing.

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). Correct responses were rewarded with nutritionally balanced fruit flavored pellets on a majority of trials and miniature chocolate candies on a minority of trials.

Figure 1. The testing equipment used in the laboratory. *Left.* Each testing system was equipped with a computer, touch screen, food dispensers, and speakers. *Right.* Monkeys were tested in their home cages. The cage door was lifted so that monkeys had direct access to the touch screen and food cups during testing.

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 the equipment in the adjacent cage. 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 (Figure 1, *right*). Test sessions were conducted daily 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 80 individuals (excluding young infants). Monkeys lived primarily outdoors in a 100 x 100 foot enclosure with access to a temperature-controlled 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 6 inches behind a poly panel, in an enclosed area the limited incursion of ambient light. The touch screens could be viewed through a 6 X 8 inch mesh window and reached through a 2 inch 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.

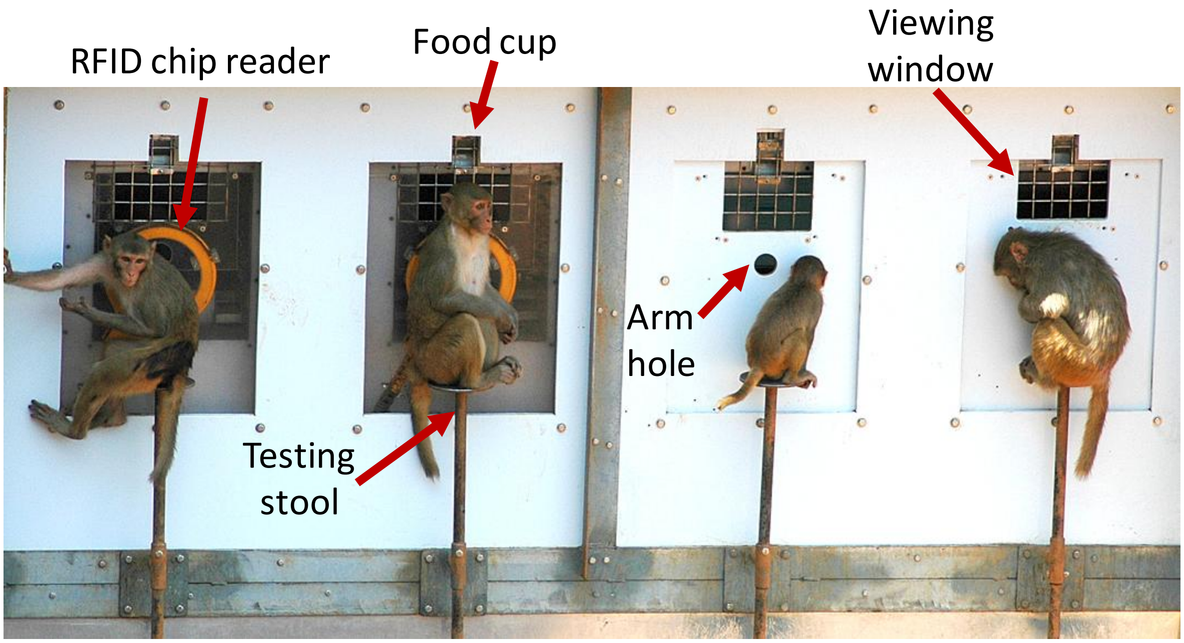

Figure 2. The four testing stations used by the Field station subjects. Monkeys sat on a stool and could view the touch screen through the window. When subjects reached their arm through the hole to touch the screen, the RFID chip reader connected to the testing computer read the subjects' RFID chip and indentified the subject. The computer selected the appropriate trial for the subject, and images appeared on the touch screen. Correct responses were reinforced with an auditory reinforcer and a food reward, which was dispensed above the viewing window.

The RFID reader at each testing station was connected to a testing 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.

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 downloaded via internet at any time to assess progress. At the start of these experiments subjects had experience with a touch screen shaping program that trained them to touch images presented on the screen and had achieved above 90% performance on a four choice 200ms delay matching-to-sample task. 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 March 2012 are reported. The sex, age, and rank of subjects 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 March 2012. During that time 39 of 80 non-infant monkeys completed the initial shaping program and were eligible to participate in the experiments. Whatever monkeys had completed training and testing in each experiment below were included in the analysis for that experiment. Thus, different experiments have different subjects and different numbers of subjects. Altogether, monkeys completed over 850,000 trials during the test period.

Juvenile (animals under 5 years of age; Table 1, *left*), adult (Table 1, *right*), male (Table 1, *top*) and female (Table 1, *bottom*) subjects used the computer testing system. However, the four adult male group members never participated.

The dominance hierarchy of the group was established through creation of a matrix based on observation of agonistic and submissive behaviors (Bernstein, 1970; De Vries & Appleby, 2000). Females were assigned to one of three equal sized groups based on their rank; High, Middle, and Low ranking individuals (10 adult females in each).

Juvenile individuals were assigned their mother's rank group for the purpose of these analyses. Animals from all three rank groupings used the testing system (Table 1).

	Juveniles			**Adults**			
	.60 *(46)*			.35 *(34)*			
	High	**Middle**	**Low**	**High**	**Middle**	**Low**	**Total**
Male	.50 *(5)*	1.0 *(5)*	0.0 *(2)*	0.0 *(2)*	---	0.0 *(2)*	.38 *(16)*
Female	.64 *(14)*	.80 *(10)*	.40 *(10)*	.30 *(10)*	.40 *(10)*	.50 *(10)*	.51 *(64)*

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.

Experiment 1: Visual psychophysics

One major concern about outdoor group testing is that it lacks the stringent environmental controls of a laboratory setting (Fagot & Bonte, 2010). For example, whereas laboratory lighting is constant, outdoor lighting is variable, which could alter the appearance of stimuli between trials. Visual psychophysical tasks provide an important comparison between the Laboratory and Field station environments because they require subjects to perceive and discriminate subtly different on-screen stimuli, a necessary prerequisite for other types of testing.

To assess monkeys' visual perception, we tested their ability to choose a target stimulus from among distracters of varying discriminability in two tasks, a size discrimination task and a brightness discrimination task. If conditions in the outdoor environment significantly alter the perception of on-screen stimuli, then the two groups should show different patterns of accuracy across different difficulty levels 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 ranked group, 5 were members of the middle ranked group, and 2 were members of the low ranked group.

General procedure

Monkeys were trained on two visual psychophysics tasks. Within each task the target stimulus remained the same across trials, and task difficulty was varied on a trial by trial basis by changing the discriminability of three identical distracters. To start a trial, monkeys touched a green start box at the bottom center of the screen (FR2). 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 different amounts along one stimulus dimension (size or brightness). Each level of difficulty consisted 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, *bottom*).

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 difficulty level.

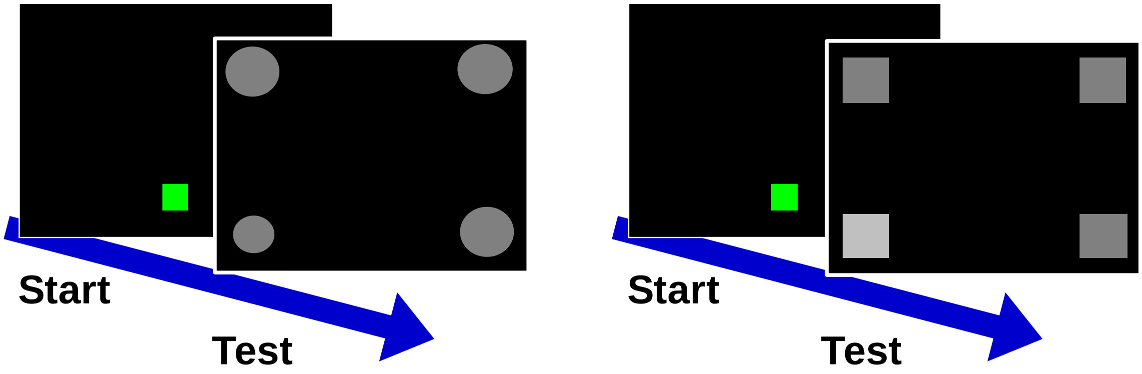

Figure 3. Trial progression during the size discrimination (*left)* and brightness discrimination (*right*). Monkeys were required to select the target stimulus from among three identical distracters that differed in either size (size discrimination task, *left*) or brightness (brightness discrimination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

Size discrimination

Monkeys were required to select a 100 pixel 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 (± 40 pixels difference in diameter from the target) to hard (± 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 x 100 pixel squares that varied by 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 (± 64

RGB value difference) to hard (± 24 RGB value difference; Figure 4, *middle*) with two distracters per level, one darker than the target and one lighter.

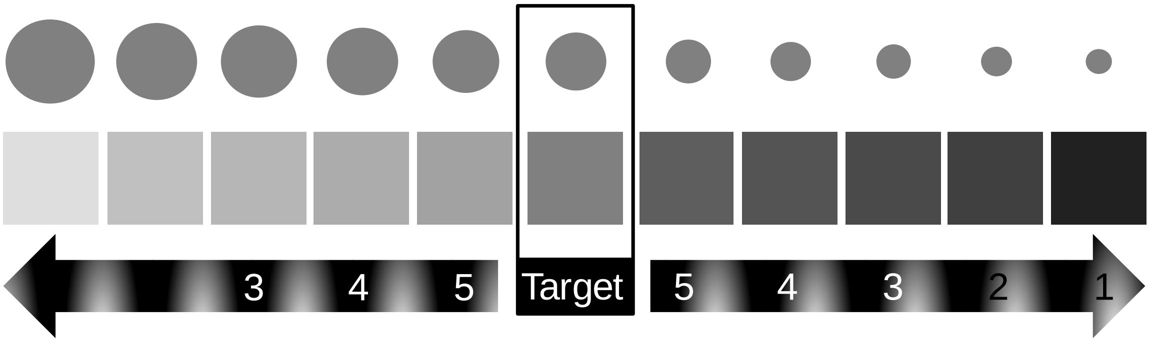

Figure 4. Stimuli used in the size discrimination (*top row)* and brightness discrimination (*middle row*). The target stimulus remained the same throughout each task (*center column*). Distracter stimuli differed from the target in two directions along five levels of difficulty (*bottom*). The distracters that were more similar to the target *(center*) were harder to differentiate (e.g. Level 5), while the distracters that are more distinct from the target (*outside*) are easier to differentiate (e.g. Level 1). Pictured brightness and sizes of the stimuli are relative.

For all experiments presented in this paper, proportion correct was transformed using an arcsine transformation (Aron & Aron, 1999) and an alpha level of .05 and two-tailed tests were used for all analyses.

Results and discussion.

There was no difference in the number of errors required before reaching criteria on the size discrimination between the Laboratory and Field station groups (Mean$_{Lab}$ + SEM = 107.17+ 6.83, Mean$_{FS}$ + SEM =156.50 + 25.72; independent samples t-test, t_{15}= -1.52, p=0.15). However subjects in the Laboratory group learned the brightness discrimination to criteria with fewer errors than the Field station group (Mean$_{Lab}$ + SEM = 335.83 + 69.40, Mean$_{FS}$ + SEM =900.18 + 166.21; independent samples t-test, t_{15}= -2.41, p=0.03).

Difficulty level affected asymptotic discrimination performance of the two groups equally. A two way repeated measures ANOVA (Difficulty Level X Group) comparing average accuracy on the two criteria 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.

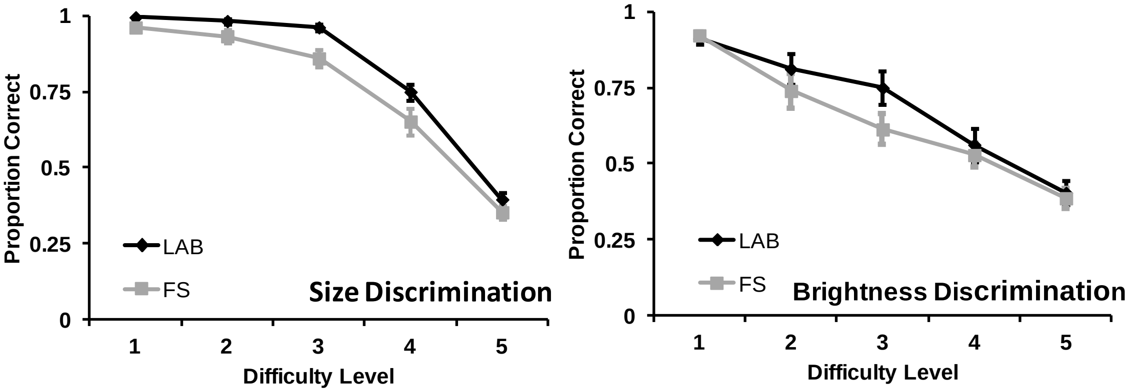

Figure 5. Mean proportion correct on the two criterion sessions across the five discrimination difficulty levels in the size discrimination (*left*) and the brightness discrimination (*right*) in the Laboratory (**black line**) and Field station (**grey line**) groups. Error bars indicate standard error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

The Laboratory subjects were 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, accounting 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 lab monkeys tended to perform numerically better than the Field station monkeys on the size discrimination, and that the main effect of group in the size discrimination approached statistical significance.

Experiment 2: Perceptual classification

Performance on two learned visual psychophysics tasks did not differ significantly between Laboratory and Field station subjects in experiment 1. In experiment 2, we evaluated performance on a perceptual classification task in which subjects had to classify complex photographs of birds, fish, flowers, and people based on their shared perceptual features. This classification task follows logically from the discrimination task in that it also requires accurate perception of the stimuli, but with an increased demand on cognitive processing.

Perceptual classification tasks are interesting because subjects can successfully solve them using different strategies. At a narrow level, subjects may memorize information about each individual image (Schrier et al., 1984) or an irrelevant perceptual feature that correlates with category membership (D'Amato & Van Sant, 1988). At a broader level, subjects may abstract a wide set of generalized perceptual features that accurately describe the category as a whole (Schrier & Brady, 1987). One can determine whether subjects have learned a narrow or broad rule by requiring them to classify novel images. A subject that has extracted a broad set of common features will show high transfer performance to novel images that share those broad features, whereas a subject using a narrow set of features will show low transfer performance because novel images are unlikely to conform to a narrow rule. In experiment 2, we evaluated whether the different environments of the Laboratory and Field station subjects would result in different patterns of learning on a perceptual classification task by assessing acquisition and transfer of four natural categories.

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 could be goldfish or sharks, alone or in schools, in the ocean or on a plate). Images were screened manually to ensure that they did not contain representatives from any of the other three categories.

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 icon for the sample's category 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 categorization 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. For reasons unrelated to the present comparison, Laboratory and Field station monkeys were trained to different criteria on this task. For Laboratory monkeys criterion was overall accuracy above 75% for one session, for Field station monkeys it was accuracy over 80% on each category on one session.

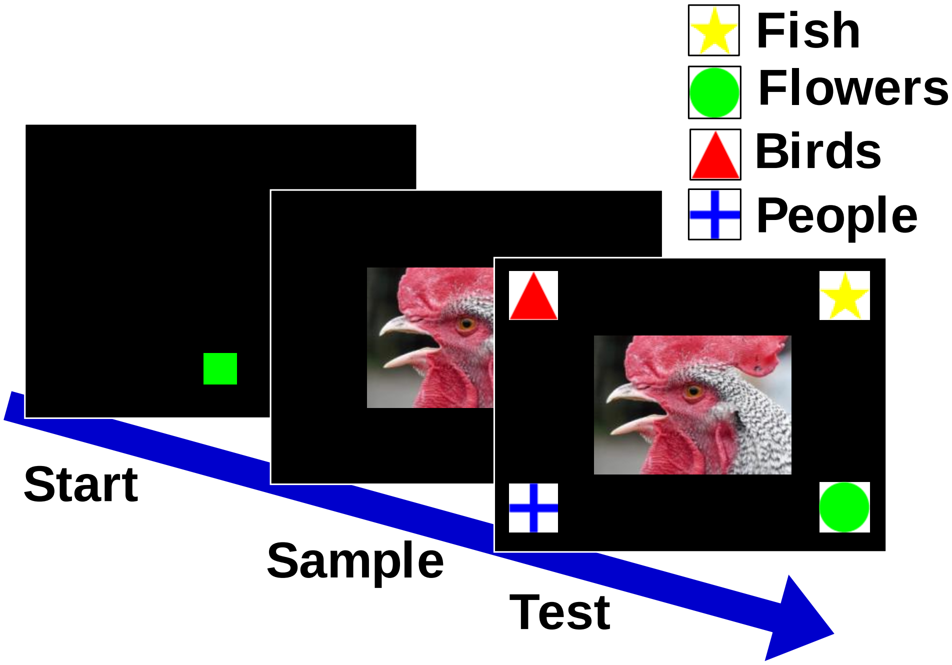

Figure 6. Example trial from the perceptual classification experiment. Subjects began a trial by pressing the green start box. A sample image from one of the four categories (bird, fish, flowers, people) appeared in the center of the screen. After the sample was touched, the four classification icons appeared in the four corners of the screen. Selection of the correct icon for the sample image category, in this case the triangle for "birds," was rewarded with an auditory and food reinforcer.

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 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$_{Lab}$+

SEM = 1837.00 + 139.97, Mean$_{FS}$ +SEM=1976.00 + 177.29 ; 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 one transfer session indicated that the Laboratory group showed no difference in performance between trained and transfer images (paired samples t-test, t_5=1.5, p= 0.30) and performance did not differ from criterion level for either training or transfer images (one sample t-tests: t*ransfer:* t_5=.15, p= 0.89; t*raining:* t_5=.73, p= 0.50; Figure 7, *left*). The Field station group showed a small but significant decrease in performance on the transfer images (paired samples t-test, t_{12}=6.17, p< 0.01); however performance on the transfer images was not different from criterion level (one sample t-test, t_{12}=.48, p=0.14). This difference between training and transfer performance was driven by *above* criterion performance on training images (one sample t-test, t_{12}=7.08, p<0.01; Figure 7, *right*).

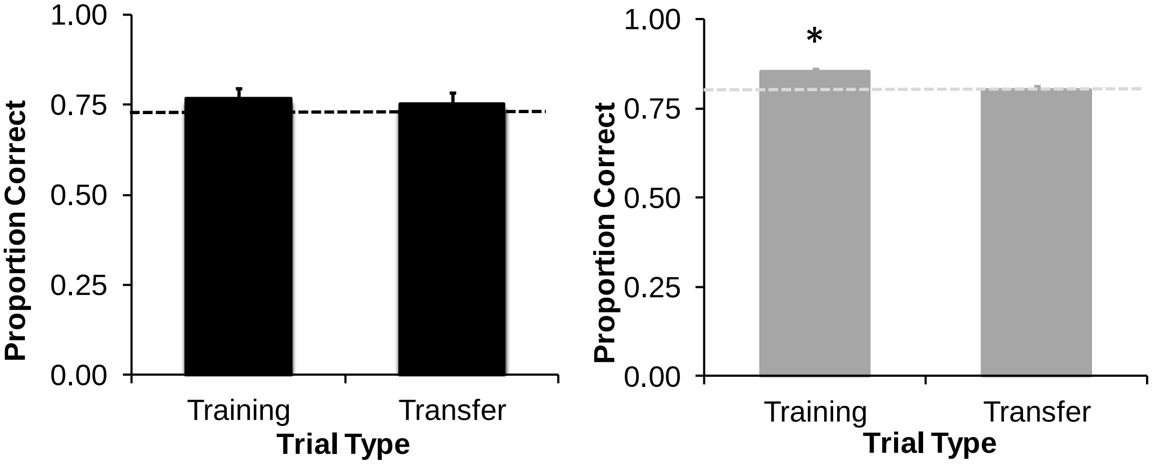

Figure 7. Mean proportion correct during the transfer session by the Laboratory (**black bars**) and Field station (**grey bars**) groups. Dotted lines indicate criterion level. Performance did not differ from criterion level on transfer images in either group, but the Field station group performed significantly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

The two groups learned to classify images at the same rate, indicating that the testing environment had no effect on task acquisition. Additionally, both groups performed at criterion level on 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. Overall, testing environment did not have a substantial effect on perceptual classification performance in monkeys.

Experiment 3: Transitive inference

In experiments 1 and 2 Laboratory and Field station monkeys performed similarly on three perceptual tasks, indicating that there are no major differences in lighting or visibility that would 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 in which the basis for correct responses is not perceptually available. Transitive inference is a cognitive task that requires subjects to infer non-perceptual relations between items based on their relations to a shared item. 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 & 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 can infer relations between these stimuli (ex: BD, CF). If subjects correctly select the higher ranked item (ex: B in BD, or C in CF), it suggests that they have a representation of the implied order of the stimuli.

Many laboratory housed species solve TI tasks (corvids: Bond, et al., 2010; fish: Grosenick, et al., 2007; crows: Lazareva et al., 2004; lemurs: MacLean et al., 2007; squirrel monkeys: McGonigle & Chalmers, 1977; rats: Roberts & Phelps, 1994; pigeons:Von Fersen, et al., 1991). Additionally, most studies report a symbolic distance effect, 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. This effect is often taken as indication of an ordered cognitive representation of the TI stimuli.

The only study to examine TI performance in a free ranging group of animals 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 it is unclear if these geese perform at the same rate or using the same cognitive mechanisms as geese housed in a laboratory setting. 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 mechanism to solve the task.

Subjects

Laboratory subjects were 6 pair-housed 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.

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 items 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 items (FR2). Selection of the correct item always resulted in an auditory reinforcer that was coupled with a food reward on 80% of trials, whereas selection of the incorrect item 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 & 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 semi 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. 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 item in the list (A or G), as these had either been always or never reinforced, respectively (Figure 8, *right*).

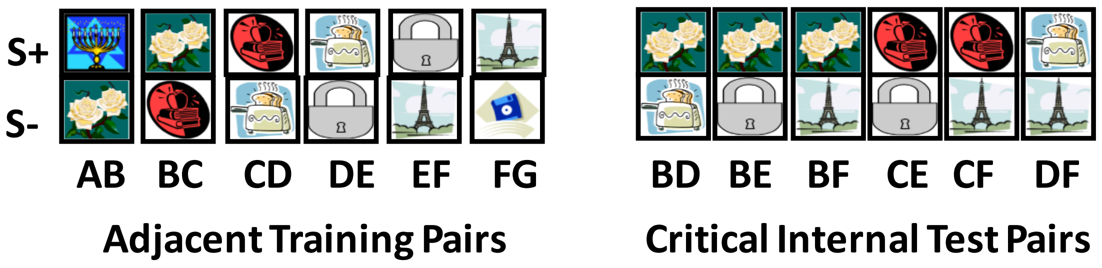

Figure 8. Adjacent training pairs (*left*) and critical non-adjacent internal test pairs (*right)* used in the transitive inference experiment. During training, correct selection of the S+ in a given pair resulted in an auditory reinforcer paired with a food reinforcer on 80% of trials. On trials containing test pairs all choices resulted in an auditory reinforcer only.

Results and discussion

Housing environment had no effect on adjacent training pair learning, as 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_{Lab} + SEM = 785.33 + 115.63 , Mean_{FS}+ SEM = 933.67 +114.04; independent samples t-test: t_{16}= 0.82, p=.47).

There was no significant difference in internal test trial performance between the Laboratory and Field station groups (independent samples t-test: t_{16}= -1.46, p=0.16), and both groups selected the higher ranked item significantly above chance on internal test trials (Mean_{Lab} + SEM = 70.83 + 4.93%; one sample t-test, t_5= 3.93, p=0.01; Mean_{FS} + SEM = 60.71 + 4.26%; one sample t-test, t_{11}= 2.50, p=0.03).

Subjects in both groups displayed the symbolic distance effect for test trial accuracy, showing increased accuracy as the symbolic distance between the items increased, and there was no significant difference in this effect between the two groups (RMANOVA Group X Symbolic Distance: Group F_1=.004, p=.95; Symbolic Distance F_4=39.23, p=.00; Interaction $F_{4,\,64}$=1.40, p=.25; Figure 9, l*eft*). Neither group showed the symbolic distance effect 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=.09, p= 0.77; Symbolic Distance F_4=1.15, p= 0.34; Interaction $F_{4,\,64}$= 2.02, p= 0.10; Figure 9, *right*).

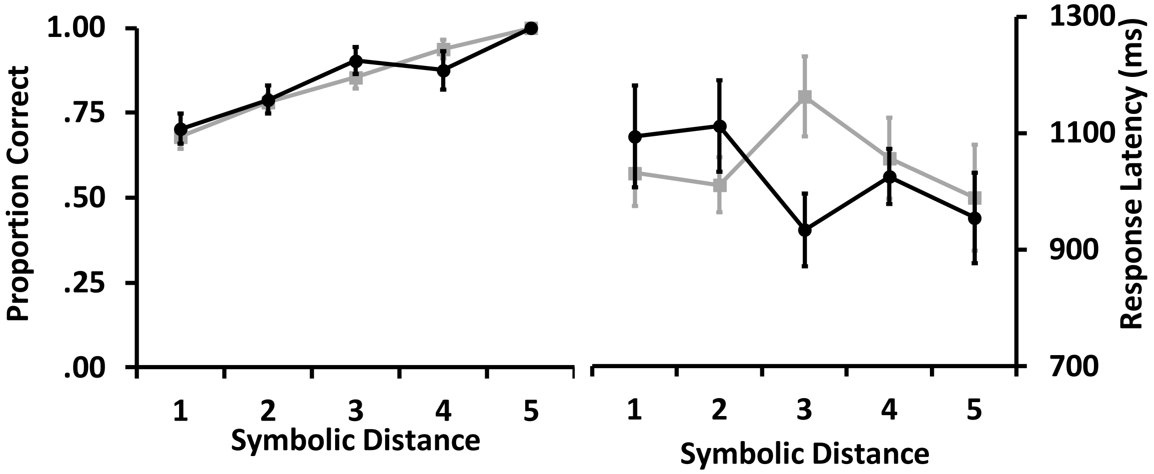

Figure 9. *Left.* Accuracy increased as the symbolic distance between the two items in the test pair increased for both the Laboratory (**black lines**) and Field station (**grey lines**) subjects. *Right.* Averages of median response latency on correct trials did not vary systematically with symbolic distance for subjects in either group. Error bars indicate standard error.

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, classification, and transitive inference tasks. However, all of these tasks require 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.

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 & Christie, 1977; Washburn & Astur, 1998). Because the Field station monkeys live in a more complex environment than Laboratory monkeys, there may be more interference from the environment, making it unfit for studies of memory. The effect of interference is strongest when the to-be-remembered images are from a small, frequently repeating image set (Basile & Hampton, in prep). 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 6 pair-housed 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.

Field station subjects were 2 male and 3 female subjects aged 2-4 years (Mean = 3.0 years). Two subjects were members of the high-ranking group, 2 were members of the middle-ranking 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

Color clip art images were presented in a delayed matching to sample procedure. A set of four session-unique clip art images were used in each 100-trial test session, such that all possible choice images were familiar to the subjects within each session. 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. A touch 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, while touches to any of the three incorrect comparison images resulted in a negative auditory stimulus and a black screen time out (15 seconds). A 3-second (Field station) or 30 second (Laboratory) inter-trial interval (ITI) separated each trial. Delays of 0.2, 6, 12, 24, and 48 seconds were counterbalanced and pseudorandomized within 100-trial sessions.

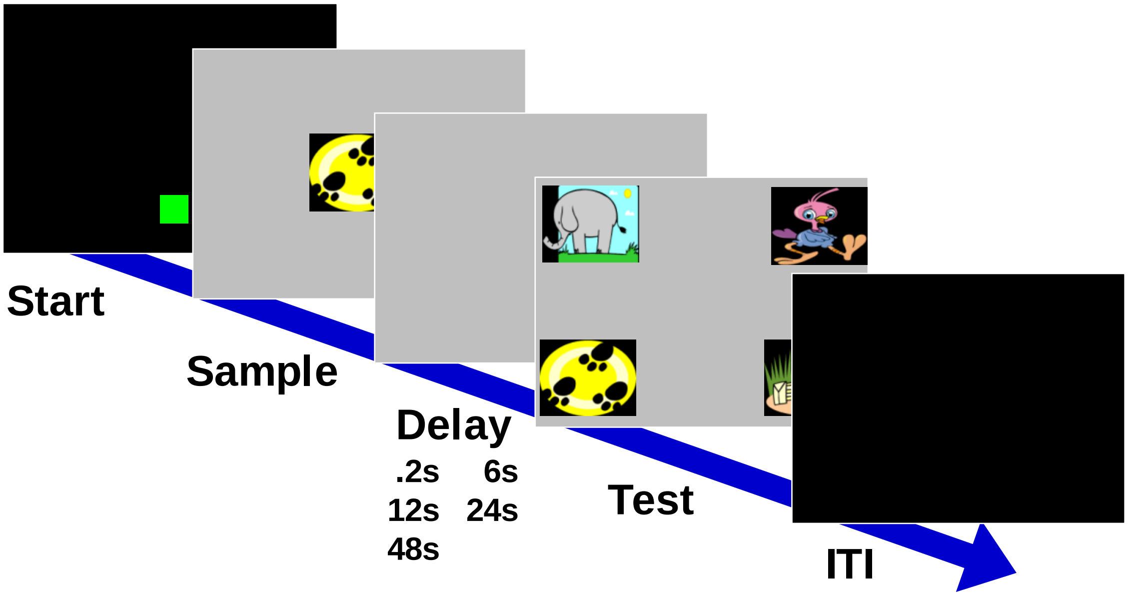

Figure 10. Example trial from the memory experiment. Subjects began a trial by pressing the green start box. One of the four possible sample images appeared in the center of the screen. After the sample was touched, a blank screen delay was presented. The four choice images then appeared in the four corners of the screen. Selection of the item identical to the sample image was rewarded with an auditory and food reinforcer.

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 walked away during the delay, the trial was aborted and the same trial was repeated when the monkey returned. The number of times a trial was aborted was recorded.

Results and discussion

Field station monkeys aborted more trials with increasing delay length (RMANOVA: F_4= 7.10, p<0.01; Figure 11), but did not abort more than 36% of trials at any delay. Field station monkey performance did not differ across delays with the inclusion of repeated trials (RMANOVA: Main effect of repeated trials: F_4= 0.01,

$p<0.94$; Interaction: $F_{4,16}= 0.44$, $p<0.78$). To better equate the task between the two groups, all further analyses excluded repeated trials.

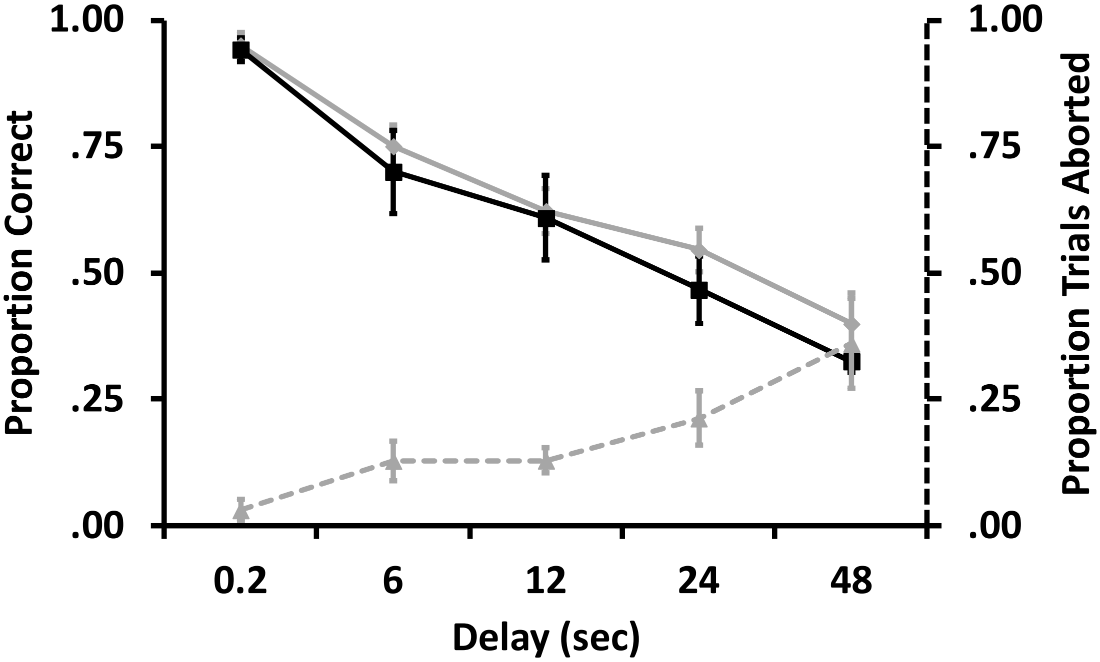

Figure 11. Accuracy on a small set matching to sample task by subjects in the Laboratory (**black line**) and Field station (**grey line**) groups across the 5 delays tested. Dashed grey line indicates proportion of trials aborted across delays by animals in the Field Station group (right axis). Error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

On the final 100 trial session, monkeys in both groups showed a significant decrease in performance with increasing delay (RM ANOVA: $F_4= 40.08$, $p<0.001$; Figure 11), but there was no difference between the two groups ($F_1= 0.36$, $p<0.56$) or interaction between group and delay ($F_{4,36}= 0.28$, $p<0.61$). 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 controlled environment. This indicates that animals in complex environments are viable research subjects even for tasks requiring attention over delays.

General discussion

Despite dramatic differences in housing, testing environment, sex, and ages of the subjects, Field station and Laboratory monkeys learned and performed comparably on psychophysical, classification, 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.

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 & Wallen, 1999) and subjects from low, medium, and high ranking families were represented in all experiments. Although subject numbers in the present experiment were too low to compare performance based on rank, future studies will use this testing system to test for rank related differences in learning rates, cognitive abilities, and social knowledge.

Although monkeys from all age groups completed some trials on the Field station testing system, all subjects who advanced far enough to participate in the four experiments presented here were juveniles without 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 simply be due to increased 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 juveniles' increased success on the testing system (Biondi et al., 2010; Glickman & Sroges, 1966; Morand-Ferron et al., 2011).

We have shown that testing socially housed monkeys in their home group produces viable data in 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 experiment. 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. Nonetheless our findings suggest that living in complex environments provides no advantage in the cognitive processes tested here. However, any conclusions based on these comparisons are tentative given the variety of differences between our Field station and Laboratory monkeys. Future studies will test for possible effects of housing conditions on other areas of cognition. The present results show that these studies can be carried out.

Chapter 3. Mechanisms underlying transitive inference performance in rhesus macaques

Knowing that *Ben* is taller than *Emily* and that *Emily* is taller than *Dina*, we readily infer without direct comparison that *Ben* is taller than *Dina*. This is known as transitive inference, the process of inferring the relation between two items based on their shared relation with a third item. Transitive inference is a protypically cognitive process thought to emerge late in development in humans (Bryant & Trabasso, 1971; Piaget, 1960). It can be used to correctly determine relations among any items along linear ordered continua, such as height, mass, and linear social dominance (Paz-y-Miño, et al., 2004).

In typical laboratory tests of transitive inference subjects are trained on a set of overlapping two-choice conditional discriminations such that that there is an implicit order (A+B-, B+C-, C+D-, D+E-, E+F-, F+G-; A>B>C>D>E>F>G). After subjects master individual premise pairs, inference is evaluated in tests of never before seen non-adjacent pairs (e.g. *BD*). The most strict tests of transitive inference exclude the end anchor items because these items are either always (A), or never (E), reinforced in training. The internal non-adjacent pairs such as *BD* have more complex reinforcement histories because they consist of items that were both reinforced (B when presented with C; D when presented with E) and non-reinforced (B when presented with A; D when presented with C) in premise pair training. Larger image sets are preferable because they provide more of these critical non-adjacent internal pairs. A set of 5 images provides only one critical test pair (*BD*), while a set of 7 images provides 6 critical test pairs (*BD, BE, BF, CE, CF, DF*). Four year old children and a diverse group of animal species perform above chance on these critical test trials, consistent with use of transitive inference (jays:

Bond et al., 2003; children: Bryant & Trabasso, 1971; rats: Davis, 1992; chimpanzees: Gillan, 1981; crows: Lazareva et al., 2004; squirrel monkeys: McGonigle & Chalmers, 1977; Merritt & Terrace 2011; Rapp et al. 1996; mice: Van der Jeugd et al., 2009; pigeons: von Fersen, et al., 1991; geese: Weiβ et al., 2010; but see Benard & Giurfa, 2004 for an exception in honeybees).

Transitive inference requires items to be processed as a ranked set in which there are no circular relationships (i.e. A>B, B>C, C>A). Differences in spacing along a linearly organized continuum result in a Symbolic Distance Effect (SDE) such that widely separated items are easier to rank correctly than are less widely spaced items (e.g. *BF* tests are easier than *BD* tests). Humans show the SDE in tests with ranked items, such as height and number (Lovelace & Snodgrass, 1971; Moyer & Landauer, 1967; Woocher et al., 1978) and humans and non-humans also show it in inference experiments, suggesting that they rank TI trained items onto an ordered continuum (Bond, et al., 2003; D'Amato, 1991; Maclean et al., 2008; Merritt & Terrace, 2011; Woocher, et al., 1978).

Logical inference is not the only cognitive process that could give rise to the patterns of performance observed in many studies of transitive inference. Both successful performance with non-adjacent internal test pairs and the SDE may be accounted for by other processes. For example, choice behavior may be controlled by the associative values of individual stimuli in some cases (Siemann, Delius, Dombrowski, et al., 1996; von Fersen & Lea, 1990; Wynne, 1998). In premise pair training with nonhuman animals, one item in a pair is reinforced with food, while the other is not. Many models have been developed to explain how inference-like patterns could emerge from the resultant variation in associative values of individual stimuli, with varying success depending on the species tested and the training procedures used (Lazareva & Wasserman, 2006; Siemann, Delius, Dombrowski, et al., 1996; Steirn et al., 1995; von Fersen, et al., 1991a; Wynne, 1998). The integral role played by reinforcement in these tasks makes it difficult to distinguish between the contributions of associative values and inference.

Whereas modeling studies show that it is possible for associative value to generate transitive inference like patterns of performance, empirical studies are required to directly test predictions based on associative value. One method used to differentiate between these two mechanisms is to manipulate the associative values of individual stimuli so that associative mechanisms and inference mechanisms would produce different patterns of choice. When the DE pair is over-trained, potentially increasing the associative value accrued to item D, pigeons (*Columbia livia)* continue to perform above chance on the *BD* test pair (Lazareva & Wasserman, 2006), whereas similarly trained crows (*Corvus cornix L.)* fall to chance (Lazareva, et al., 2004), suggesting that crows but not pigeons were relying on associative values to solve the task. Resistance to extinction measures show no systematic differences between B and D for pigeons after bias reversal, but resistance to reinforcement measurements indicated an increase in the value of D compared to B for most birds (Lazareva & Wasserman, 2012). However discrepancy models of associative learning, like Rescorla-Wagner, operate such that once performance has reached asymptote the only way to significantly increase associative value is to increase the reinforcer (Rescorla & Wagner, 1972).

Another method for dissociating the contributions of associative value and inference to TI task performance is to create a modified task that can only be solved using one of these methods. One such task is to require linking of two separate ordered lists (e.g. A>B>C; X>Y>Z) into a single larger ordered list by training only a single linking pair (C>X). Test trials consist of never before seen pairs of items, one from each of the original lists. If subjects used inference to create a single ordered representation of the two previously separate lists, then they will correctly choose the item from the higher ranked list in these probe tests. By contrast, if choice is driven by associative values, monkeys would be unable to correctly select the higher ranked item in between list pairs. This is because items occupying the same position in the two lists would have acquired similar associative values during initial training. None of the current associative models predict above chance performance on linked lists (Lazareva, 2012). Monkeys taught five item lists can link two and three of these lists together after training on single linking pairs (Treichler & Raghanti, 2010; Treichler et al., 2003; Treichler & Van Tilburg, 1996). There are no published reports of list linking by other species.

In the present series of experiments we determined the contributions of inference and associative values to performance on TI tasks in monkeys. In experiment 1 we documented performance patterns on a 7 item TI task. In experiment 2 we measured the associative values of individual stimuli to evaluate whether these values predicted performance on test trials. In experiment 3 we manipulated the associative values of multiple stimuli in a TI task. In experiment 4 we presented monkeys with a list linking task that cannot be solved by associative value.

Experiment 1: Transitive inference

In Experiment 1 we established baseline patterns of performance of rhesus monkeys in computerized 7-item transitive inference tests.

Method

*Subjects and apparatus***.** Subjects were twelve four to six-year-old male rhesus monkeys (*Macaca mulatta*) that had been raised by their biological mothers in a large social group until the age of approximately 2.5 years before moving to the laboratory. Monkeys were pair-housed whenever possible and kept on a 12:12 light:dark cycle with light onset at 7:00 am. They received a full ration of food daily and water was available ad libitum.

Procedure. Monkeys were tested in their home cages. Computerized touch-screen test systems, each consisting of a 15-inch LCD color monitor (3M, St. Paul, MN) running at a resolution of 1024 X 768 pixels, generic 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. Test sessions were conducted daily between 10 am and 5 pm, six days per week.

During testing, each pair of monkeys was separated by an opaque plastic divider with holes that allowed visual, auditory, and tactile contact, but prevented the monkeys from touching the computer screen in the adjacent cage. Computer screens were locked to the front of each monkey's cage and the door was raised, giving subjects full access to the screen during testing. After a 3-second inter-trial interval (ITI), a green box appeared at the bottom of the screen and remained until the monkey touched it (FR2) to start a trial. Two images from the training set appeared on the right and left sides of the screen (counterbalanced over trials), and remained until one was touched (FR2). Selection of the correct item always resulted in an auditory reinforcer; a food reward was also delivered on 75% of trials (85% of food rewards were nutritionally balanced banana flavored pellets; Bio-Serv, Frenchtown, NJ and a random 15% of food rewards were miniature chocolate candies). Selection of the incorrect item in the pair resulted in a negative auditory stimulus and a five second time out during which the screen was black.

Premise pair training. Stimuli consisted of seven 300 X 300 pixel color clip art items presented in overlapping adjacent pairs (A+B-, B+C-, etc.) that could be organized into an implied linear hierarchy (A>B>C>D>E>F>G).

Training proceeded one premise pair at a time, with pairs at the bottom of the implied hierarchy (FG) trained first (Treichler & Van Tilburg, 1996). Each premise pair was introduced individually in 25 trial sessions until subjects reached 80% correct in a single session. Twenty-five trials of that pair were then pseudo-randomly intermixed in a session including 25 trials of each of the previously learned pairs until subjects performed above 80% on each pair present in one session. This pattern continued as indicated in Table 1 until monkeys met criterion with all 6 training pairs in a single session (phase 11, Table 1).

	Premise pair training phase										
	1	2	3	4	5	6	7	8	9	10	11
Pairs presented	FG		FG		FG		FG		FG		FG
		EF	EF		EF		EF		EF		EF
				DE	DE		DE		DE		DE
						CD	CD		CD		CD
								BC	BC		BC
										AB	AB
Number of trials per session	25	25	50	25	75	25	100	25	125	25	150

Table 1. Training and test phases in experiment 1. The pairs presented in each phase are shown in each column, with the total number of trials per session in each phase is shown in the bottom row. Subjects were moved on to the next phase after performing over 80% correct in a single session on all pairs present in that phase.

Transitive inference test trials. One trial of each non-adjacent test pairing (e.g. AF, BD, CE, etc.) was pseudo-randomly intermixed with 25 trials of each of the 6 premise pairs to generate a session of 165 trials (15 TI test pairs and 150 premise pairs). TI test trials were non-differentially reinforced with the auditory reinforcer only, consistent with auditory reinforcement only subjects had learned to expect on 25% of premise pair training trials. Subjects received 4 sessions of these test trials.

The entire training and testing procedure was completed twice for each subject with two distinct sets of stimuli. Two distinct image sets were counterbalanced across subjects for each run of the experiment.

Data analysis. All response latency analyses in this paper used log-transformed median latencies from correct trials only (Montgomery, 1953). All accuracy data were arcsin transformed before analyses (Aron & Aron, 1999). All analyses were conducted using an alpha level of .05.

Results and discussion

Performance with the two image sets did not differ in either the first or second run of the experiment in the total number of errors made before reaching criterion (independent samples t-tests: Run 1:M_{set1}=672.80, *SEM*=143.73, M_{set2}=749.71, *SEM*=62.85; t_{10} =-0.55, *p*=.60; Run 2: M_{set1}=529.80, *SEM*=133.01, M_{set2} =626.00, *SEM*=49.54; t_{10}=-0.77, *p*=.46) or in accuracy on internal test pairs (independent samples t-tests; Run 1: M_{set1}=69.85, *SEM*=0.39, M_{set2}=69.49, *SEM*=0.23; t_{10}=0.05, *p*=.96; Run 2: M_{set1}=80.76, *SEM*=0.63, M_{set2}=70.05, *SEM*= 0.48; t_{10}=1.18, *p*=.27). The two sets within each run of the experiment were combined for further analyses.

Monkeys made more total errors before reaching criterion in the first run of the experiment than in the second run of the experiment (M_{run1}=717.67, *SEM*=67.21, M_{run2}=585.92, *SEM*=60.55; paired samples t-test: t_{11}=4.07, *p*=.002). There was no difference between the two runs of the experiment in performance on the six adjacent training pairs during the criterion session (RMANOVA: $F_{5, 55}$=.34, *p* =.88), or performance on the critical non-adjacent internal test pairs (M_{run1}=69.64, *SEM*=0.13, M_{run2}=74.71, *SEM*=0.28; paired samples t-test: t_{11}=-0.82, *p*=.43), therefore all test data were combined across the two runs for further analysis.

Subjects performed significantly above chance on critical non-adjacent internal test pairs (*M* = 71.56, *SEM*=0.08; t_{11}=7.95, $p<.001$). These results reinforce previous findings that rhesus macaques and other species solve inference tasks (corvids: Bond et al., 2010; chimpanzee: Gillan, 1981; crows: Lazareva, et al., 2004; rhesus monkeys: Buckmaster et al., 2004; Merritt & Terrace, 2011; Rapp, et al., 1996; Treichler &

Raghanti, 2010; Treichler et al., 2007; Treichler & Van Tilburg, 1996; pigeons: Von Fersen et al., 1991; greylag geese: Weiβ, et al., 2010)

To test for the SDE all test pairs were grouped according to the number of list items intervening between members of the pair. For example, a symbolic distance of 1 included internal test pairs *AC, BD, CE, DF,* and *EG,* whereas a distance of 2 included *AD, BE, CF,* and *DG.* Monkeys showed SDEs in that response latency decreased and accuracy increased as the symbolic distance between tested items increased (Figure 1; RMANOVA Response Latency: $F_{4,44}$= 9.60, p<.01; RMANOVA Accuracy: $F_{4,44}$= 52.75, p<.01).

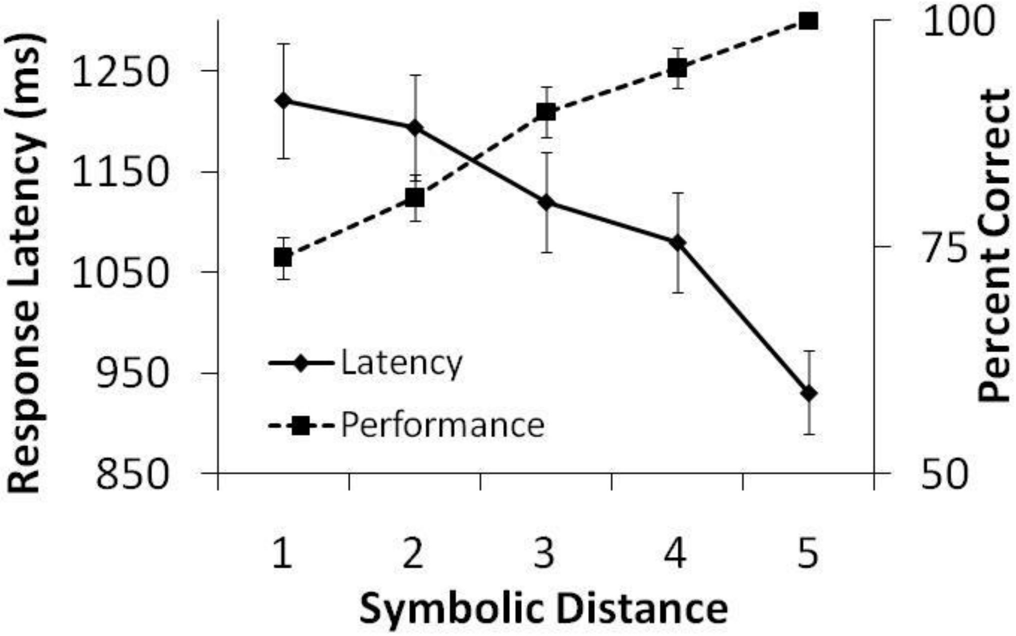

Figure 1. Average accuracy and average median response latency on all non-adjacent test pairs (end anchor and internal) in experiment 1 by symbolic distance. Error bars represent standard error of the mean.

At larger symbolic distances the proportion of contributing pairs that contain end anchor items increases (i.e. distance 5 has 1/1pairs containing items A and G whereas distance 1 has 2/5). Because end anchors are either always or never reinforced, these pairings may be easier than others and may lead to overestimates of the SDE (Vasconcelos, 2008). We examined the SDE by the first item in the pair so that only one pair contributes to performance at each distance. The SDE patterns still hold across different first items (Figure 2, Table 2), indicating that the overall effect is not driven entirely by the contributions of end items pairs to the larger symbolic distances, and suggesting that all stimuli are represented in the ordered array.

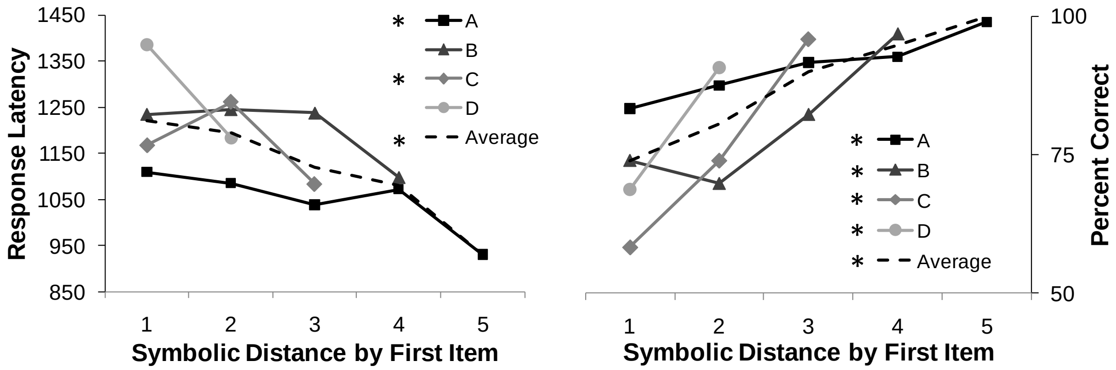

Figure 2. Symbolic distance effect for response latency (*left*) and average performance (*right*), sorted by the first item in the pair. Stars (*) indicate significant scores on a repeated measures ANOVA. The effect of significantly decreasing response latency with increasing symbolic distances holds for all first items except for *B* and *D*, and the effect of significantly increasing performance with increasing symbolic distance holds for all first items (Table 2). Because this analysis controls for the contribution of the end anchors, it indicates that the overall symbolic distance effect is not driven entirely by performance on the end anchors.

	Response latency		Percent correct	
First item	*F*	*p*	*F*	*P*
A	6.56	.03*	8.65	<.01*
B	1.48	.24	11.93	<.01*
C	3.57	.05*	46.37	<.01*
D	1.63	.23	11.88	.01*

Table 2. Statistical results of Repeated Measures ANOVA for Symbolic Distance Effect by first item as shown in Figure 2. The first item in the pairs included in each analysis is presented on the far left. There was a significant increase in performance with increasing symbolic distance for all first items, and a significant decrease in response latency with increasing symbolic distance for pairs in which A was correct and pairs in which C was correct. * indicates $p<.05$.

Monkeys performed above chance on internal test pairs in two seven item lists and showed a robust symbolic distance effect for both accuracy and latency, which suggests that they have an ordered representation of the stimuli. Although these patterns are consistent with a ranked organization and use of logical inference, the same pattern of performance could result if differences in associative values mirrored the implied rank order of the stimuli. In experiment 2, we empirically measured the associative values of individual stimuli after premise pair training to determine whether performance on standard TI tasks can be accounted for by these values.

Experiment 2: Measurement of associative values

Associative value accounts of TI performance posit that items used in TI tests gain associative values consistent with the implied TI order through association with primary reinforcers during premise pair training. Differences in associative value manifest as preference for particular stimuli on non-adjacent test trials. Studies have modeled how TI stimuli could accrue associative values that produce above chance performance on test trials (Siemann & Delius, 1998; Siemann, et al., 1996; Von Fersen, et al., 1991). However, the only study to empirically measure the values of stimuli used in transitive inference tests used resistance to reinforcement and found that these values do not predict performance on test pairs in pigeons (Lazareva & Wasserman, 2012). In order to determine the extent to which associative values produce the same performance patterns as expected from inference on standard TI tasks in monkeys, we explicitly measured the associative values of the stimuli used in experiment 1.

We presented monkeys with two identical concurrent schedules; one schedule was associated with neutral images and the other with the seven trained TI images. This allowed us to measure the relative associative value of each of the seven TI stimuli. Secondary reinforcers that have reliably signaled reinforcement elicit higher rates of responding than secondary reinforcers that sporadically signal reinforcement (Armus & Garlich, 1961; D'Amato & Lachman, 1958). Because the two schedules delivered the same reward type and amount with the same average frequency, differences in rates of responding can be attributed to differences in the associative values of the stimuli (Miller, 1976). The relative response rate to each of the seven transitive inference stimuli was used to index associative values.

If monkeys’ choices in the TI tests were determined by the associative values of individual stimuli, they should select the stimulus with the higher associative value in each test pair regardless of whether that stimulus was higher or lower in the implied hierarchy. Accordingly, larger differences in the associative values of the items in a test pair should result in greater preference for the higher valued stimulus in TI tests. We compared the extent to which choice is controlled by associative value by assessing the extent to which choice behavior on TI tests correlated with associative value.

Method

*Subjects and apparatus***.** Subjects and apparatus were the same as in experiment 1.

Stimuli and procedure. Monkeys were trained on two concurrent random interval schedules (RI 30 second) until they reached a steady rate of responding to both schedules. The two schedules were represented by 300 x 300 pixel images presented simultaneously on the left and right sides of the screen. Images for the two schedules were drawn from a pool of 20 familiar clip art images and changed in synchrony every 30 seconds, independent of the state of either schedule. The schedules operated independent of one another, and the first touch after each random interval was reinforced with one food pellet and an auditory reinforcer. All other touches were recorded, but did not result in reinforcement.

The images for which associative values were to be measured were presented as probe trials pseudorandomly intermixed with the 20 familiar images used during concurrent RI training. A probe trial began when the images at both schedule locations changed; the right location changed to different one of the 20 familiar clip art images, whereas the left location changed to be one of the to-be-measured probe stimuli. Probe images appeared on the left side of the screen to control for side bias. During the 30 seconds that the probe stimulus was on the screen, neither schedule terminated and the monkeys were not rewarded for any touches. When the 30 second probe period was over, the images each changed to one of the 20 standard images, and a shortened RI 10 second schedule was initiated on both keys. Consecutive probe trials were separated by at least 90 second. To increase the likelihood that subjects were attending to the task when the probes appeared, probes were only presented after subjects had touched each key at least once since the presentation of the last probe.

Object discrimination reversal. To evaluate the efficacy of our methods for measuring associative value, we first measured the accrual of associative value during training on a series of object discriminations and reversals. Monkeys learned 4 two-image object discriminations concurrently. One of the images in each pair was randomly designated the S+ and was reinforced, whereas the other was the S-. Monkeys received three 8 trial sessions per day (two trials of each of the 4 object discriminations) until they chose correctly at least 23 of the 24 trials in a single day. The contingencies of each discrimination were then reversed so that choice of the item that had been rewarded was now non rewarded (S-) and choice of the item that had been non-rewarded was now rewarded (S+). Monkeys were trained to the same criterion on this reversal, followed by a final reversal in which the reward contingencies were the same as in the original training. Immediately after each eight trial object discrimination training session one 25 minute concurrent RI30 second measurement session was conducted in which all of the stimuli from the object discrimination pairs were presented as probes.

Transitive inference. Following completion of the object discrimination training and measurement, the associative values of the transitive inference stimuli used in the first iteration of experiment 1 were measured using the same methods as described above. Monkeys received one transitive inference test session with reinforcement contingencies as described in experiment 1, followed by one measurement session in which the probe stimuli were the seven images from the TI set. Three more TI test sessions were conducted, one per day over the next three days. Accuracy for each of the 15 TI test pairs was calculated by averaging performance over these four test sessions.

Data analysis. The associative value of each probe stimulus was assessed by calculating the proportion of total touches allocated to the probe stimulus during the 30 second it appeared on screen during the one measurement session using the formula:

$$\text{Value Index } = \frac{\text{Touches to the Probe Stimulus}}{\text{Touches to the Probe Stimulus} + \text{Touches to the NonProbe Stimulus}}$$

Associative value scores therefore range from 0 (aversion to the probe stimulus) to 1 (exclusive preference for the probe stimulus).

For each of the 4 object discrimination pairs and each of the 15 possible non-adjacent TI test pairs we calculated an associative value difference score using the formula:

$$\begin{aligned}&\text{Associative Value Difference Score}\\&\qquad\qquad = \text{Value Index of Correct Item} - \text{Value Index of Incorrect Item}\end{aligned}$$

Positive difference scores indicate that the correct item (the S+ in the object discrimination task and the higher ranked item in the TI task) had a larger value index than the incorrect item. Negative difference scores indicate the incorrect item had a larger value index than the correct item. If monkeys rely on associative values of individual stimuli to make their choices, then the valence and magnitude of these differences should predict choice in object discrimination and TI test trials.

We used a Generalized Linear Mixed Model with subject as a random factor to compare average daily associative value difference scores with average daily choice on the object discrimination task over the initial training, first reversal, and second reversal. A Generalized Linear Mixed Model with subject as a random factor compared associative value difference scores with choice on the 15 transitive inference test pairs. One subject's data were excluded from the transitive inference measurement analysis due to insufficient touching (this monkey touched both control and probe stimuli less than 0.1 times per second).

Results and discussion

Measured associative value differences correlated with object discrimination performance. There was a significant relationship between associative value difference scores and percent choice of the S+ across testing days (Figure 3; Generalized Linear Mixed Model, daily associative value difference X daily discrimination accuracy: $F_{1,170}$= 7.90, p =.01). This positive relationship demonstrates that the method used to measure associative values can detect differences relevant to choice behavior in tasks in which associative values are expected to control choice.

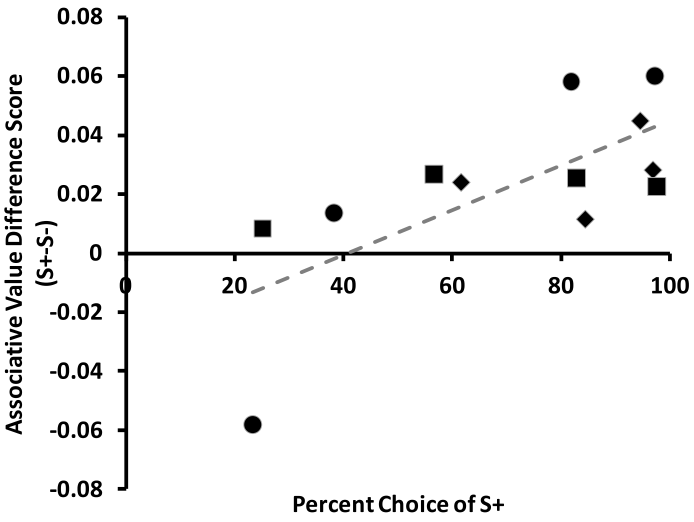

Figure 3. Average percent choice of the S+ during object discrimination acquisition and reversal plotted by associative value difference score. Each point represents the average of all data from one day of measurements (3 measurement sessions, 24 object discrimination trials). The first two days of measurement and the last two days of measurement are depicted for initial training (diamonds), first reversal (squares) and second reversal (circles) phases. For each phase the left-most point on the x axis corresponds to the first measurement day, and points progress rightward through training to the final criterial point. Trend line is indicated with a dashed line. There was a significant relationship between choice and pair difference score in the object discrimination task.

If TI test pair performance was also controlled by associative values, then associative value difference scores should be correlated with TI accuracy. By contrast if performance was controlled by the implied order of the TI items, accuracy should be unrelated to value differences and instead should vary with symbolic distance between the items in the test pairs (D'Amato & Colombo, 1990). In this experiment symbolic distance was not correlated with associative value difference scores (Spearman Rank Order correlation: r_{13}=.18, *p*=.53). Therefore any effect of symbolic distance in these results cannot be explained by associative value differences alone.

There was no relationship between associative value difference scores and accuracy across the fifteen test pairs from the TI image set (Figure 4, *left;* Generalized

Linear Mixed Model, associative value difference x pair accuracy: $F_{1,12.68}$= .68, *p* =.42). In contrast, the correlation between accuracy and symbolic distance of the test pairs was nearly significant (Spearman Rank Order correlation: r_{13}=.51, *p*=.05, Figure 4, *right).* These results suggest that performance on test trials in the TI task was not controlled by associative value as was the case in the object discrimination task, but instead were driven by the implied order of the stimuli.

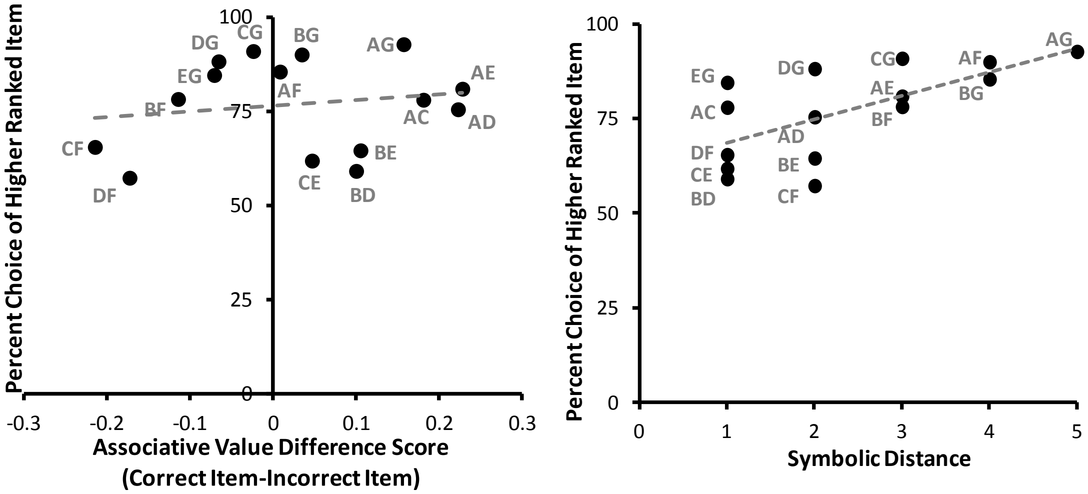

Figure 4. Average percent choice of higher ranked item for all non-adjacent TI test pair in the transitive inference stimulus sets plotted against associative value difference score (*left*) and the symbolic distance between the items in the pair (*right*). Each point represents one transitive inference test pair. Trend lines are indicated with a dashed line. There was no significant relationship between accuracy and associative value difference score, but the correlation between accuracy and symbolic distance was nearly significant.

Associative value models predict that animals will select the item with the higher associative value on TI test trials. In contrast, our results support those recently reported in pigeons (Lazareva & Wasserman, 2012) that choices in the TI tests were not correlated with associative value. In fact, in TI tests, monkeys often chose the stimulus with the lower associative value (all those points in Figure 4, *left* where associative value difference is negative and accuracy is above .50). This indicates that when associative value and implied order conflict, choice behavior tends to be controlled by the implied order. In experiment 3, we will further evaluate the relative influence of associative values and inference by explicitly manipulating the magnitude of the reinforcement associated with different images in a TI task.

Experiment 3: Manipulation of associative values

In standard transitive inference training paradigms, items high in the implied hierarchy may accrue larger associative values than items lower in the hierarchy. To the extent this is the case, associative value is congruent with the implied order of the hierarchy, making it difficult to determine whether inference or associative value control choice behavior. In experiment 2, we found that associative value and implied order were often incongruent: in six of the TI test pairs the associative value of the item lower in the implied order had accrued a larger associative value than item at the top of the implied order. Despite this incongruency, monkeys often chose images consistent with the implied order. However, the small number of these pair types and the variability in the associative value differences make it difficult to draw firm conclusions about the influence of the two mechanisms. In experiment 3, the associative values of individual stimuli were directly manipulated to more clearly dissociate the contributions of inference and associative values in situations in which the two mechanisms conflict.

Other investigators have overtrained selected premise pairs in an effort to increase the associative value of a lower ranked item. For example, overtraining the DF pair could increase the value of D relative to other stimuli because it is reinforced on many additional trials. Results from this manipulation in pigeons and crows are mixed; pigeons continued to select B over D in transitivity tests, suggesting use of inference, whereas crows chose D and B equally often, suggesting that the associative value manipulation influenced their choices (Lazareva, et al., 2004; Lazareva & Wasserman, 2006). The fact that in neither case was D chosen over B in transitivity tests may suggest either that choice is not controlled by associative value or that overtraining has only a modest effect on associative value. Because all premise pairs are trained to a high accuracy criterion using the same reinforcer, all items in the TI set may already have values close to maximum supported by the reinforcer (Rescorla & Wagner, 1972). Additional reinforced trials administered in overtraining may have only a small effect.

To produce large systematic differences in associative value among stimuli in an implied hierarchy, reward magnitude was manipulated. Selection of some images during training was rewarded with a single food pellet, whereas selection of other images was reinforced with two pellets. When this manipulation produces associative values that are Congruent with the implied order, monkeys should perform above chance whether or not their behavior is controlled by associative value. When the resultant associative values are Incongruent with the implied order, accuracy should decrease to the extent that choice is controlled by associative values.

Method

Subjects and apparatus. Subjects were the same 12 rhesus monkeys used in experiments 1 and 2.

Procedure.

Premise pair training. Stimuli were two new sets of seven 300 x 300 pixel photographs. Premise pair training was conducted as in experiment 1, with the following exceptions. In the Congruent condition, correct choices of items A, B, or C were rewarded with 2 pellets, whereas correct choices of items D, E, or F were rewarded with one pellet (G was never correct and therefore never rewarded). In the Incongruent condition, correct choice of images A, B, or C were rewarded with 1 pellet, whereas correct choices of items D, E, or F were rewarded with 2 pellets rewards. Incorrect choices in both conditions resulted in no food reward, a negative auditory stimulus, and a five second time out during which the screen was black. All monkeys received both conditions with order of the conditions and image sets counterbalanced across subjects.

Transitive inference test trials. Test trials were presented as in Experiment 1. Correct choices were defined as those consistent with the implied order regardless of how the stimuli had been rewarded during training.

We assessed the influence of associative value in two ways. First, we compared the number of errors required to reach criterion in premise pair training in the Congruent and Incongruent conditions. Second, we examined performance on critical internal non-adjacent test pairs BD, BE, BF, CE, and CF in the Congruent and Incongruent conditions. For these pairs, one image in the pair had been reinforced with a single pellet and the other image had been reinforced with two pellets. Because the implied order (A>B>C>D>E>F>G) did not differ between the two conditions, choices based on the implied order would result in above chance performance in both the Congruent and Incongruent conditions. In contrast, if choice behavior is controlled by associative value monkeys should select the item that was reinforced with two food rewards during training over the item that was reinforced with only one, even when these choices conflict with the implied order. In the Congruent condition choices based on associative value would still result in above chance performance, as the higher ranked items also had higher associative values, but in the Incongruent condition choices based on associative value would result in below chance performance, because the lower ranked items had higher associative values.

Results and discussion

Training order did not affect learning rates or accuracy within either condition (errors to criterion, Congruent: t_{10}=.12, p=.90; Incongruent: t_{10}=-1.20, p=.26; internal test pair performance, Congruent: t_{10}=1.40, p=.19; Incongruent*:* t_{10}=-1.58, p=.14). Therefore data were collapsed with respect to order of testing.

*Premise pair training***.** In both the Congruent and Incongruent conditions, reinforcement of premise pairs FG, EF, and DE is congruent with implied order and should not be difficult to learn. In the Incongruent condition, reinforcement of pairs CD, BC, and AB is in conflict with the implied order. To the extent that choices are controlled by associative value, acquisition of these pairs, but not others, should be retarded relative to the congruent condition. There were significant main effects of Condition (Congruent or Incongruent) and Premise Pair (FG, BC, etc.), and a significant interaction in errors to criterion (2x6 RMANOVA: Condition: $F_{1,11}$=11.39, p=.01; Premise Pair= $F_{5,55}$= 9.99, p<.01; Condition X Premise Pair: $F_{5,55}$=8.77, p<.01). Monkeys made significantly more errors reaching criterion on pairs CD, BC, and AB in the Incongruent condition than in the Congruent condition (Figure 5; paired samples t-tests: CD: t_{11}= -5.15, p<.00; BC: t_{11}= -2.35, p=.04; AB: t_{11}= -2.82, p=.02). Monkeys learned pairs FG, EF, and DE in the two conditions with similar numbers of errors (Figure 5; paired samples t-tests: FG: t_{11}= 0.22

p=.83; EF: t_{11}= 0.20, *p*=.85; DE: t_{11}= -1.44, *p*=.18). Despite differences in the rate of learning, by the end of training monkeys performed above 85% correct on all six test pairs in both conditions.

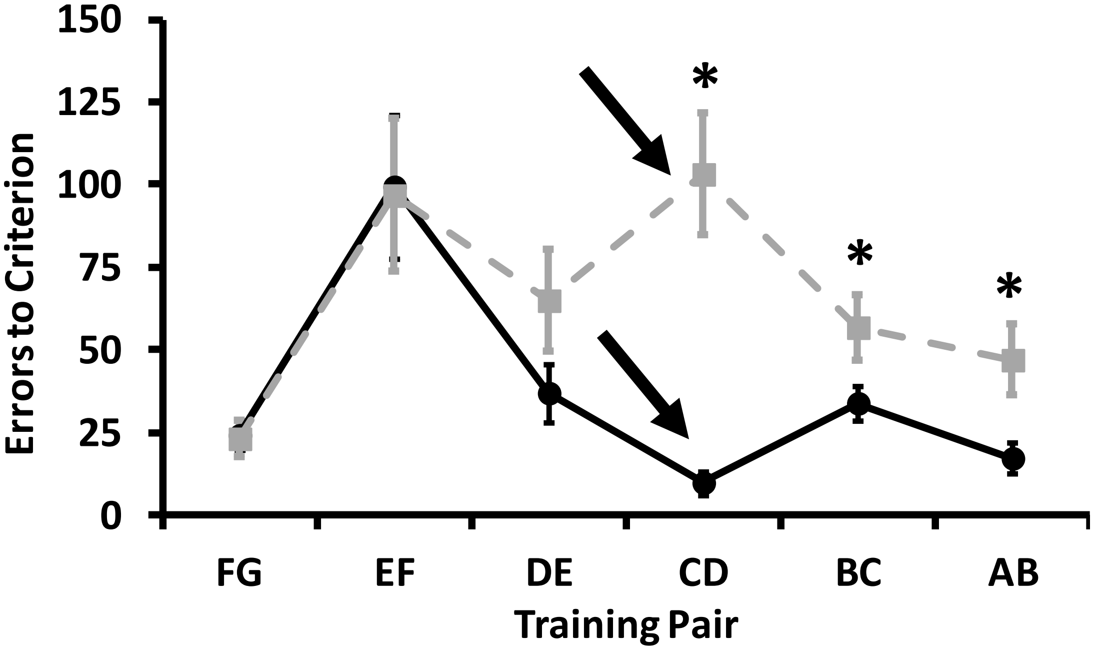

Figure 5. Total errors to criterion for the 6 premise pairs in the Congruent (**solid black**) and Incongruent (**dashed grey**) conditions. Premise pairs were learned in the order they are shown from left to right (i.e. pair FG was learned first and AB last). Arrows indicate change in number of reinforcers during training. Reinforcement on pairs *CD, BC*, and *AB* in the Incongruent condition is in conflict with the implied order of the stimuli. ***** indicates a significant difference between conditions (paired-samples t-tests, *p*<.05). Error bars are standard errors.

Transitive inference testing. On TI test trials where the two items tested had been trained with different reinforcement values (BD, BE, BF, CE, CF), monkeys performed above chance in the Congruent condition, but below chance in the Incongruent condition (Congruent: *M*= 89.58, *SEM*=5.79%; one sample t-test: t_{11}= 6.50, p<.00; Incongruent; *M*= 37.08, *SEM*=4.62%; one sample t-test: t_{11}=-2.79, *p*=.02). This indicates that when differences in associative value are sufficiently large, the influence of associative value on choice can overwhelm control by implied order. Because premise pairs in both conditions were trained to the same criterion, deficits in internal test pair performance in the Incongruent condition cannot be explained the differences in learning.

Monkeys correctly chose the higher ranked item in both the Congruent and Incongruent conditions if items in the tested pair were associated with the same reward (i.e. pairs AC and DF; Congruent: *M*= 89.58, *SEM*=4.57, one sample t-test: $t_{11}=7.01$, $p<.00$; Incongruent: *M* = 76.04, *SEM*=3.59, one sample t-test: $t_{11}=5.42$, $p<.00$). This result suggests that monkeys had latent knowledge of the implied order of the stimuli, but the two fold reinforcement difference between test items in pairs BD, BE, BF, CE, and CF led associative values to control choice on these trials.

This is the first study we know of in which associative value was manipulated using different reward magnitudes. The results show that when associative value differences are large they do influence both premise pair learning and performance in TI tests. These results differ from previous studies that found little or no effect of efforts to manipulate associative value in crows and pigeons via overtraining (Lazareva, et al., 2004; Lazareva & Wasserman, 2006). Differences in reward magnitude may have considerably stronger effects on associative values than does overtraining.

We found that choice in TI tasks can be controlled by differences in associative value when associative value and inference are incongruent with one another. But even in the Incongruent condition, monkeys selected images consistent with the implied order if the target and the distracter had been reinforced with the same number of pellets (pairs AC and DF). This suggests that monkeys may have encoded both implied order and associative value during premise pair training and associative value may mask knowledge of implied order when differences in associative value are sufficiently large. To further evaluate whether monkeys infer order in TI tests, we presented a list linking task that can only be solved using implied order. Choices based on associative value would produce a very different pattern of performance in this test from choices based on inference, yet differences in associative value are not expected to be strong enough to mask implied order as may have been the case in experiment 3.

Experiment 4: List linking

To further evaluate whether choices on TI tasks are controlled by the implied order of stimuli, monkeys were presented with a list linking task that cannot be solved using the associative values of the stimuli. Because the two original lists were trained separately and performance did not differ between them, images occupying the same relative location in the lists (e.g. B and I, D and K) should have accrued similar associative values during training. Consistent with this, on un-linked independently trained TI lists, monkeys select the item with the higher relative position in its originally trained list (Merritt & Terrace, 2011; Treichler, et al., 2003). We taught monkeys that the lowest item (G) in one of the lists learned in Experiment 1 (A>B>C>D>E>F>G) ranked above the highest item (H) in the other list (H>I>J>K>L>M>N). If choices in TI tests are controlled by associative values, training on the linking pair GH will not result in systematic selection of images from the higher ranked list over images in the lower ranked list but instead would result in performance similar to that seen in unlinked tests. Presentation of Same Location pairs would result in chance performance. In tests with pairs in which the image from the higher ranked list occupied a relatively lower position in the training list (Lower Location pairs, e.g. F and I) choice by associative value would result in below chance accuracy with respect to implied order. Additionally, because images in the two lists have similar associative values, choice by associative value does not predict an SDE spanning the two linked lists. In contrast, if choices on TI tasks are controlled by inferred order, then monkeys should link the two independently learned lists into a single 14 item list (A>B>C>D>E>F>G>H>I>J>K>L>M>N) and correctly select any item from the higher list over any item from the lower list. Additionally, because they have linked the two lists into one large list, monkeys should show a SDE that spans the entire 14 item list.

Method

*Subjects and apparatus***.** Subjects were the same twelve monkeys used in experiments 1, 2 and 3, and general testing procedures and apparatus remained the same.

*Stimuli and procedure***.** Stimuli consisted of the two sets of seven color clip art images that were used in the two repetitions of experiment 1. Training and testing reward contingencies were the same as in experiment 1.

Re-familiarization. In order to ensure that subjects remembered the premise pairs learned in experiment 1, they were presented with re-familiarization sessions consisting of 25 trials of each of the 6 previously trained adjacent premise pairs from one of the two lists (AB, BC, CD, DE, EF, FG). Once they reached >80% on all six premise pairs in one session, they were presented with sessions containing the 6 premise pairs from the second list (HI, IJ, JK, KL, LM, MN) until they reached this same criterion. Finally, they were presented with sessions in which all 12 of the premise pairs from the two lists were intermixed. During this re-familiarization phase none of the pairs spanned the two lists, thus monkeys were familiarized with test sessions containing 12 test pairs intermixed, but could not link the two previously learned lists at this stage.

Linking. List linking training sessions included 25 trials of the linking pair in which the lowest item (G) from the higher ranked list was rewarded when paired with the highest item (H) from the bottom ranked list until subjects performed above 80%. For half of the subjects the higher ranked list was the first one learned in experiment 1, for the other half of the subjects it was the second one learned in experiment 1. Next, subjects received training sessions in which all 13 training pairs were intermixed (the 12 premise pairs from the two previously learned lists and the one linking pair) until they were performing above 80% on all 13 pairs in a session.

TI testing. Test sessions consisted of all possible non-adjacent test pairings pseudo-randomly intermixed with the 13 training pairs in a session containing 403 trials. The 13 premise pairs and linking pair made up 325 of these trials (25 of each trial type), within list test pairs (non-adjacent pairs of stimuli from within the same list, e.g. AC, BD, HK) made up 30 of these trials, and between list test pairs (never before seen non-adjacent test pairs spanning the two lists, e.g. AN, BK) made up 48 of the trials. Monkeys received four test sessions.

Results and discussion

Re-familiarization and training. The number of errors made before reaching criterion in the re-familiarization phase did not differ between the two previously trained lists (M_{List1}= 361.75, *SEM*=369.97; M_{List2}= 359.75, *SEM*=75.95; paired samples t-test

$t_{11}=0.04$, $p=.97$). Monkeys made an average of 94.33 (*SEM*=9.74) errors on the linking pair (GH) before reaching criterion.

In order for subjects relying on associative values to perform above chance on both within and between list test trials, there must be systematic changes in the values of the items in the previously trained lists such that all items in the lower list accrue a value below all items in the higher list, but items within each list maintain their relative values. This change would occur after linking when all pairs were intermixed, and would likely result in performance changes on some of the premise pairs. However, when the 13 premise pairs (6 from the higher list, 6 from the lower list, and 1 linking pair) were intermixed, subjects performed at pre-linking rates on all higher and lower list pairs except for the pairs containing the linking items (FG, GH, and HI) on their first intermixed sessions, and maintained this above chance performance on all other pairs over all intermixed training sessions (Figure 6). Performance on pairs containing the linking items (pairs FG, GH, and HI) did decrease compared to performance during re-familiarization (paired samples t-tests: FG*:* $t_{11}= 9.48$, $p<.01$*;* GH*:* $t_{11}=4.47$, $p<.01$; HI*:* $t_{11}=6.23$, $p<.01$), but this pattern of decreased performance after addition of a new adjacent pair is typical in sequentially presented transitive inference training (Treichler & Van Tilburg, 1996).

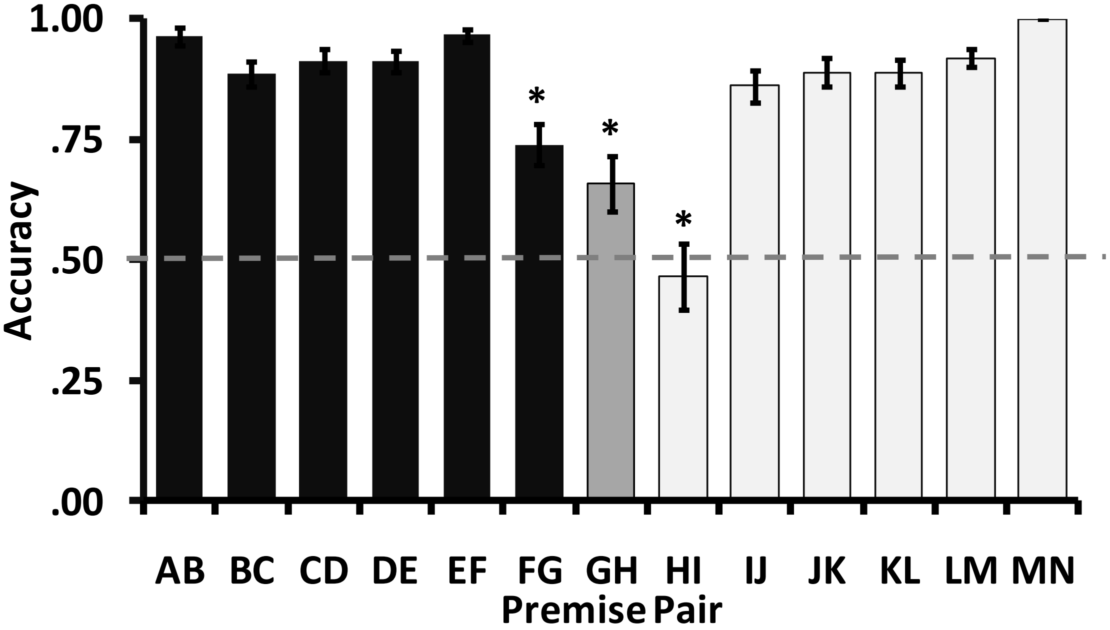

Figure 6. Performance on the 13 premise pairs in the first intermixed training session of experiment 2. Performance on all pairs except pair HI remained above chance, and only pairs FG, GH, and HI showed significant decrements in performance from the last session of re-familiarization (significant difference on a paired samples t-test indicated by *****). Error bars indicate standard errors.

Test. After linking, subjects maintained test trial performance within the higher and lower lists, performing significantly above chance on the internal within list test pairs that were previously tested in experiment 1 (Higher List: M = 70.31, SEM=3.95; t_{11}=5.17, p<.01 *;* Lower List: M= 66.78, SEM=4.61; t_{11}=3.56, p<.02). They also showed the SDE for accuracy within both lists, although there was no difference in response latency across the symbolic distances (Higher List: accuracy: $F_{4, 44}$=15.69, p<.01, latency: $F_{4, 44}$=0.81, p=.53*;* Lower List: accuracy: $F_{4, 44}$=6.07, p<.01, latency: $F_{4, 44}$=0.32, p=.86).

Accuracy and response latency on between list test pairs did not differ as a function of whether the list learned first or second in experiment 1 took the higher or lower position when the lists were linked (independent samples t-tests: accuracy t_{10}=0.18,

p=.86; response latency (t_{10}= 1.75, p=.11).Therefore data were collapsed for further analyses.

Monkeys performed above chance on critical non-adjacent internal between list test trials (M= 66.54, SEM=3.68; t_{11}=7.15, p <.001). In contrast with associative value accounts which predict below chance performance on Lower pairs and chance performance on Same pairs, monkeys performed above chance regardless of the relative locations of the test items in their originally trained lists (Lower: M = 59.43, SEM=2.74; t_{11}=3.43, p=.01*;* Same: M = 77.28, SEM=2.50; t_{11}= 9.82, p<.01*)*, suggesting that their choices were driven by the inferred order of the stimuli.

To determine if subjects integrated the two separate lists into one large 14 item list after linking training, we examined the SDE for novel between list test pairs. Accuracy data were consistent with the SDE across all 12 between list symbolic distance (Figure 7; RMANOVA: $F_{11,121}$= 38.27, p<.01). This pattern of results suggests that monkeys formed a linear representation of one large 14 item list. However monkeys did not show a significant SDE for response latency (Figure 7: RMANOVA: $F_{11,121}$=.79, p=.65). The latency data on between list pairs had high variability, as indicated by the large standard errors (Figure 7). The novelty of the task and large memory load of 14 items may have led to this large variability and may account for the lack of a systematic pattern in response latency.

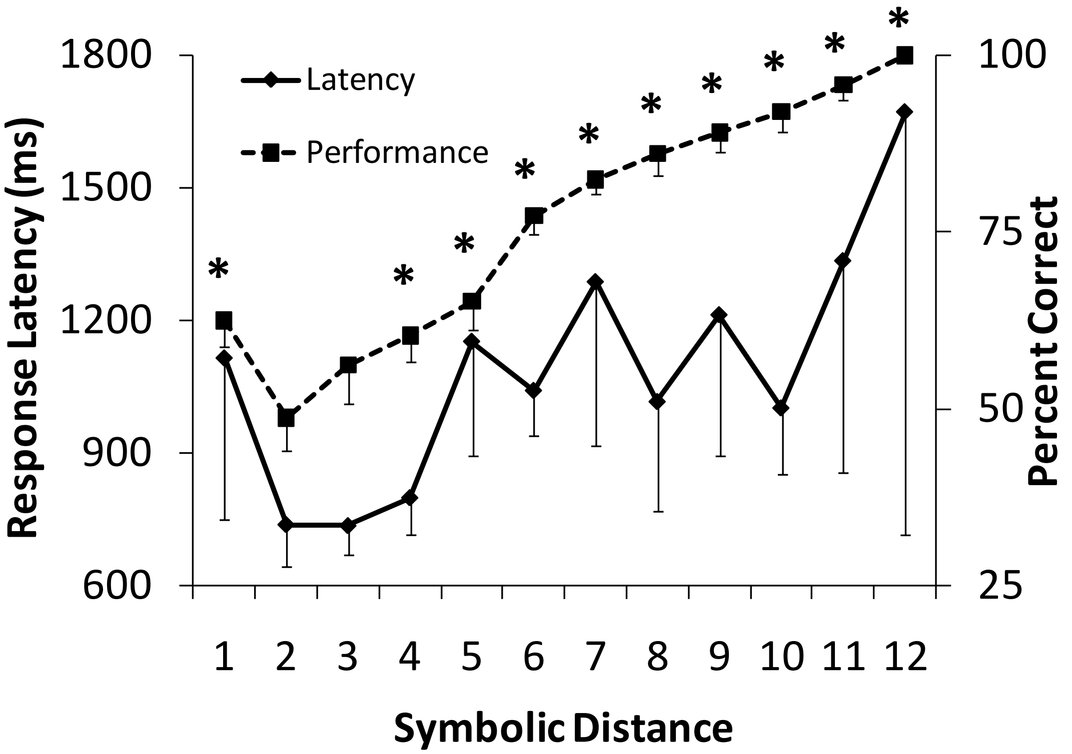

Figure 7. Average performance and response latency on all *between* list test pairs (end anchor and internal) in experiment 2 by symbolic distance. Accuracy follows the pattern expected if monkeys used transitive inference, latency did not differ systematically across the symbolic distances. Error bars represent standard error of the mean. * indicates above chance performance according to one sample t-test.

Together, between list test pair accuracy and the symbolic distance pattern suggest that monkeys linked the two separately learned seven item lists into one fourteen item list. This linking was done after exposure to only one linking pair (GH) and on the monkeys' first experience with this type of task. These results support and expand upon previous findings that experienced monkeys link five item lists, and can select the higher ranked item on between list pairs regardless of their relative rankings in their initially trained lists (Treichler & Raghanti, 2010; Treichler et al., 2007; Treichler & Van Tilburg, 1999; Treichler & VanTilburg, 1996). List linking results like those shown here cannot be explained by any of the current associative models of performance (Lazareva, 2012), and suggest that monkeys performance on TI tasks can be driven by the implied order of stimuli.

General discussion

Together the findings from these four experiments suggest that while monkeys' choices on TI tasks can be driven by associative values, choice on standard TI tasks is driven by the inferred order of stimuli. Monkeys selected the item associated with a larger reinforcer when the inferred order and associative values were incongruent in experiment 3, indicating that associative values can drive TI performance. However, the results of experiment 2 indicated that monkeys' correct choices on standard TI test trials were not related to the measured associative values of individual stimuli. Above chance performance on the equally reinforced pairs in the Incongruent condition in experiment 3 suggested latent knowledge of the inferred order not expressed when the reinforcement differences between stimuli were large. Monkeys seamless linking of two seven item lists into one 14 item list in experiment 4 cannot be explained by any of the current associative models.

Associative value models have successfully predicted performance on transitive inference tasks by pigeons (Siemann, et al., 1996; Von Fersen, et al., 1991; Wynne, 1995, 1998). Although these models have not yet been applied to monkey data, our empirical measurements of associative values in experiment 2 suggest that monkeys' performance on TI tasks cannot be well accounted for by these models. Pigeons require more extensive training to reach criterion level on the premise pairs than was required by the monkeys in the present study (Monkeys: Experiment 1, average 718 errors to learn 6 premise pairs; Pigeons, average 4512 trials, Lazareva & Wasserman, 2006), and while this may simply be indicative of differences in memory, motivation, or attention, it may also indicate that these species rely on different cognitive mechanisms. Even across closely related species, learning rates and performance patterns on transitive inference tasks vary so dramatically as to suggest choice is controlled by different cognitive mechanisms (Bond, et al., 2010; Bond, et al., 2003; Lazareva, et al., 2004; Maclean, et al., 2008). In corvids, species that live in complex social environments or rely heavily on caching for food show performance patterns that are more consistent with use of inference than species whose natural environments do not contain these complex cognitive demands (Bond, et al., 2010). It is therefore important to take into account a species' naturally history when discussing the mechanisms underlying task performance, and is invalid to assume that all species choices are controlled by the same mechanism.

Recent studies have examined use of transitive inference for learning dominance relationships in animals. Inference may allow animals to rapidly, safely, and flexibly learn dominance hierarchies (Bond, et al., 2010; Grosenick, et al., 2007; Paz-y-Miño, et al., 2004) as it could be used to construct a dominance hierarchy after observing interactions between a small number of pairs of animals, avoiding many costly aggressive encounters (Cheney & Seyfarth, 1990). Fish, chickens, and jays all correctly identify the dominant subject from observing adjacent pairs of dominance interactions even though the stimuli (individual animals) are not reinforced with food. This suggests that these animals can use inference to learn relationships between stimuli in a social context, and may therefore use it over associative values in non-social contexts, such as in standard laboratory TI tasks.

In three of the four TI experiments, monkeys' choices were primarily controlled by the inferred order of stimuli. This indicates that monkeys are capable of extracting an implied order from overlapping pairs of stimuli. However, the below chance performance on the unequally reinforced pairs in the Incongruent condition in experiment 3 shows that under certain conditions choices can also be controlled by associative values. This highlights the dual nature of the mechanisms underlying TI performance; both associative values and inference are likely employed in concert to solve this task. The extent of the contribution of each may depend on species, task parameters, or subject expertise (Bond, et al., 2010; Lazareva, et al., 2004; Lazareva & Wasserman, 2006; Maclean, et al., 2008).

Chapter 4. Discussion

Laboratory and socially housed monkeys learned and performed similarly on visual psychophysical, perceptual classification, transitive inference, and delayed matching to sample tasks. Broad similarities in learning and performance occurred despite differences in age, sex, experience, and housing environment in the two groups. This suggests that monkeys rely on similar cognitive mechanisms to solve these tasks despite demographic differences. Additionally these similarities provide strong evidence that animals living in complex environments can be productive research subjects in a broad range of cognitive domains.

Increased exposure to cognitive demands in the Field station environment did not facilitate cognitive task performance as predicted. Performance by monkeys in the Laboratory and Field Station groups did not differ even on the TI task, which may engage cognitive mechanisms refined by experience in complex social environments. However, 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 have been long enough for critical environment-dependent cognitive development to occur. Future studies with older Field station subjects are needed to draw conclusions about the long term cognitive benefits of complex housing environments.

Contrary to the predicted benefit of housing environment on TI performance, there were three major demographic differences between the two groups that predicted a decrement in TI performance by Field station monkeys compared to Laboratory monkeys. First, our Laboratory subjects had 5 years of experience with cognitive testing whereas Field station monkeys had only an average of 9 months. For both the Laboratory and Field station subjects the data presented in chapter 2 are from their first experience with TI tasks, however differences in experience with other cognitive tasks could have resulted in differences in the cognitive mechanisms driving choices. Second, all of our Laboratory monkeys are male, whereas the majority of the Field station monkeys are female. Female and male rhesus monkeys have different life histories that may have exerted different selection pressures on cognitive development. Additionally they have different social roles which may result in differences in attention or motivation. Within Field station subjects, males performed significantly better than females on TI test trials (independent samples t-test: $t_{10}=2.39$, $p=.038$), but there were no significant differences between males and females on the three other tasks (visual psychophysics, RMANOVA sex X discrimination difficulty interaction: $F_{4,36}=2.35$, $p=.07$; brightness $F_{4,36}=1.01$, $p=.42$; classification transfer performance, independent samples t-test: $t_{11}=0.54$, $p=.60$; memory: RMANOVA sex x delay interaction: $F_{1,3}=0.094$, $p=.80$). Due to the small number of male subjects (n=2) this result is only suggestive that sex differences within a group may lead to differences in TI performance. Future studies with increased sample size will address this comparison further.

Finally, Field station monkeys averaged 5 years younger than Laboratory subjects. Our study is the first to test TI performance in young monkeys (M =2.8 years), and whereas no effects of age within or between the two groups were found, subjects differed in many demographic factors which prevent conclusions about the effects of age on monkey TI performance. Human children as young as 4 years old can solve TI tasks when stimuli can be ordered along a physical dimension, such as size, length, or spatial arrangement (Bryant & Trabasso, 1971; Goswami, 1995; Pears & Bryant, 1990; Russell et al., 1996), but perform at chance when stimuli have no inherent physical order (Russell, et al., 1996). Adults, in contrast, perform well regardless of the orderability of stimuli, suggesting that the mechanisms used and/or expertise at solving TI tasks in humans may vary with age (Lazareva & Wasserman, 2010; Moses, Ostreicher, et al., 2010; Russell, et al., 1996).

In both monkeys and humans, age may be a determining factor in the contributions of associative value and inferred order to TI task performance. Findings from our detailed TI studies in Laboratory subjects show that there is flexibility in the mechanisms that drive choices on TI tasks even within individual subjects. Although Laboratory monkeys primarily solved TI tasks based on the implied order of the stimuli, they did rely on associative values to make choices when the value differences were large. In contrast to younger adult monkeys (4-6 years), aged monkeys (M= 22 years) display TI response patterns that indicate choices driven by associative values (Rapp, et al., 1996). In human adults, awareness of the hierarchical relations between items in a TI task is often positively correlated with test trial performance (Lazareva & Wasserman, 2010; Moses et al., 2006; Smith & Squire, 2005). Older adults (M= 68 years) are significantly less aware of the order of stimuli in TI tasks than younger adults (M= 23 years), and subsequently perform less well on test trials (Moses, Ostreicher, et al., 2010). Choices by highly aware participants appear to be primarily driven by an explicit representation of the relations between TI items, while correct choices by unaware participants appear to be driven by implicit knowledge of which item was correct (Siemann & Delius, 1996; Smith & Squire, 2005). This dissociation is further supported by findings from neuroimaging studies that unaware participants show different brain activity patterns during TI test trials than aware participants (Greene et al., 2006; Moses, Brown, et al., 2010).

Experience, sex, and age differences between Laboratory and Field station monkeys suggest that there should be differences in the relative contributions of implied order and associative value to TI test choices between the groups. To fully understand how animals solve TI tasks, the mechanisms that underlie use of associative values and implied order need to be elucidated. While the mechanisms of associative learning are well understood (Domjan, 2004; Rescorla & Wagner, 1972), the mechanism underlying relational knowledge of TI stimuli are unclear. Online inferences of the type generally referred to in TI examples (if Ben is taller than Emily and Emily is taller than Dina, then Ben is taller than Dina), would not produce the performance patterns seen in TI tasks (Vasconcelos, 2008). When presented with B and D on a TI test trial, an individual could actively infer *if B>C and C>D, then B>D,* but this online inference would lead to *longer* response latencies and *decreased* accuracy with increasing symbolic distance between items, as more inferences need to be made for more disparate items. However the prevalence of the symbolic distance effect in TI tasks, which shows *shorter* response latencies and *increased* accuracy with increasing symbolic distance, suggests that a representation of the relations between items is formed during training then later referenced to solve test trials.

The hypothesis that the a mental representation of the ordered list is created during training is supported by limited evidence that animals trained on TI premise pairs before hippocampal system disruption perform well on post-lesion TI test trials (Van der

Jeugd, et al., 2009), whereas animals trained after lesion perform poorly (Buckmaster et al., 2004; Dusek & Eichenbaum, 1997). The hippocampus may be necessary for forming a representation of TI stimuli during training, but not for accessing a pre-existing representation at test (Van der Jeugd, et al., 2009). In future studies we will evaluate the role of the primate hippocampus in TI by comparing test pair performance on two TI sets trained before and after hippocampal removal. This will inform our understanding of both the neural basis of TI and of the cognitive mechanisms underlying TI performance.

The implication of the hippocampus in TI performance in nonhuman animals and humans (Buckmaster, et al., 2004; Dusek & Eichenbaum, 1997; Fortin et al., 2002; Greene, et al., 2006; Heckers et al., 2004; Moses, Brown, et al., 2010; Nagode & Pardom, 2002; Van der Jeugd, et al., 2009; Zalesak & Heckers, 2009) may suggest that the mental representation created in TI tasks may be spatially organized (Moses, Brown, et al., 2010). In humans, non-spatial ordered information is often spontaneously represented as a spatially organized "mental line" (Holmes & Lourenco, 2011; Prado et al., 2008; Previtali et al., 2010; Schwarz & Keus, 2004; Shaki & Fischer, 2008). In a TI task this type of representation could result in a mental line with item A on the far left, item G on the far right, and items B, C, D, E, and F located linearly between (Brunamonti et al., 2011; Chen et al., 1997; D'Amato & Colombo, 1990; Roberts & Phelps, 1994). At test this mental line would be referenced and the leftmost item in the pair would be correctly selected, producing above chance performance on TI test trials. Items further apart on this mental line would be easier to distinguish, resulting in the symbolic distance effect.

In humans, the limited evidence available suggests that TI tasks produce patterns consistent with a spatial representation of the relations between stimuli (Brunamonti, et al., 2011; Moses, Ostreicher, et al., 2010; Previtali, et al., 2010). Few studies have explicitly tested whether TI stimuli can be represented spatially in animals, and none have explored this in monkeys. Rats learn TI tasks faster when the stimuli are trained in a physical linear order (Roberts & Phelps, 1994), and crows, like human children, only perform above chance on a TI task when stimuli are followed by physically orderable post choice feedback (Lazareva, et al., 2004). We are currently testing whether monkeys and humans represent TI stimuli spatially by determining the extent to which knowledge of a pre-trained spatial order of stimuli facilitates learning TI relationships between those stimuli.

The use of implied order in our TI tasks indicates that monkeys possess the cognitive capacity to infer natural dominance relationships. However, to infer relations between third party individuals in a natural social group, monkeys must also be able to learn dominance relations through observation. Monkeys can select the dominant individual in observed social interactions (Bovet & Washburn, 2003; Paxton et al., 2010) and may alter their behavior based on information gained through “eavesdropping” on third party social relationships (Bachman & Kummer, 1980; Crockford et al., 2007; but see le Roux & Bergman, 2012 for counter evidence). Therefore, like chickens, fish, and pinyon jays, monkeys may possess the cognitive abilities necessary to both learn pair-wise dominance relationships through observation and combine those relations into an ordered hierarchy (Grosenick, et al., 2007; Hogue, et al., 1996; Paz-y-Miño, et al., 2004).

In future studies we will use the Field station testing techniques described here to determine the extent to which monkeys use TI to learn dominance hierarchies. We will manipulate the social experience of Laboratory monkeys to directly test for social TI and will determine whether the mental representation of the dominance hierarchy acquired through normal social experience in the Field station monkeys can support TI. We have established that Field station monkeys are viable research subjects on complex cognitive tasks, and that they can perform TI in non-social tasks. Using videos of individuals from the Field station dominance hierarchy, we will test whether monkeys know third party relationships, which can only be learned through observation (Cheney & Seyfarth, 1990; Grosenick, et al., 2007), and whether judgments of real dominance relations show the symbolic distance effect often seen in TI tasks. These studies will provide a direct test of the extent to which monkeys possess third party social knowledge and will be the first test of the hypothesis that monkeys learn dominance hierarchies using TI.

Together, the experiments presented here provide the first explicit evidence that monkeys solve TI tasks using the implied order of the stimuli, and confirm the viability of socially housed rhesus monkeys as subjects for cognitive testing. The combination of these findings allows for future studies that will explore whether monkeys naturally learn social dominance hierarchies through TI. Our future work will provide a direct measurement and comparison of mechanisms underlying social and nonsocial TI, will characterize the role played by TI in social dominance hierarchy learning, and will inform our understanding of the influence of social selection pressures on the evolution of cognition.

References

Andrews, M. W., & Rosenblum, L. A. (1994). Automated recording of individual performance and hand preference during joystick task acquisition in group-living bonnet macaques (Macaca radiata) *J. Comp. Psychol., 108*(4), 358-362.

Appleby, M. C. (1983). The probability of linearity in hierarchies. *Anim. Behav., 31*, 600-608.

Armus, H. L., & Garlich, M. M. (1961). Secondary reinforcement strength as a function of schedule of primary reinforcement. *Journal of Comparative and Physiological Psychology, 54*, 56-58.

Aron, A., & Aron, E. (1999). *Statistics for psychology*. Upper Saddle River, NJ: Prentice Hall.

Bachman, C., & Kummer, H. (1980). Male assessment of female choice in Hamadryas baboons. *Behavioral Ecology and Sociobiology, 6*(4), 315-321.

Barbet, I., & Fagot, J. (2011). Processing of contour closure by baboons (Papio papio). *Journal of Experimental Psychology-Animal Behavior Processes, 37*(4), 407-419.

Barchas, P. R., & Mendoza, S. P. (1984). Emergent hierarchical relationships in rhesus macaques: An application of Chase's model. In P. R. Barchas (Ed.), *Social hierarchies* (pp. 81-95). Westport: Greenwood Press.

Basile, B. M., & Hampton, R. R. (in prep). Images from small, but not large, sets are cognitively effortful for monkeys to keep in working memory.

Benard, J., & Giurfa, M. (2004). A test of transitive inferences in free-flying honeybees: Unsuccessful performance due to memory constraints. *Learning & Memory, 11*(3), 328-336.

Bernstein, I. S. (1970). Primate status hierarchies. In L. A. Rosenblum (Ed.), *Primate Behaviour*. New York: Academic Press.

Biondi, L. M., Bo, M. S., & Vassallo, A. I. (2010). Inter-individual and age differences in exploration, neophobia and problem-solving ability in a Neotropical raptor (Milvago chimango). *Anim. Cogn., 13*(5), 701-710.

Bond, A., Wei, C. A., & Kamil, A. C. (2010). Cognitive representation in transitive inference: A comparison of four corvid species. *Behavioral Processes*.

Bond, A. B., Kamil, A. C., & Balda, R. P. (2003). Social complexity and transitive inference in corvids. *Animal Behaviour, 65*, 479-487.

Bonte, E., Flemming, T., & Fagot, J. (2011). Executive control of perceptual features and abstract relations by baboons (Papio papio). *Behavioural Brain Research, 222*(1), 176-182.

Bovet, D., & Washburn, D. A. (2003). Rhesus macaques (Macaca mulatta) categorize unknown conspecifics according to their dominance relations. *J Comp Psychol, 117*(4), 400-405.

Brunamonti, E., Genovesio, A., Carbe, K., & Ferraina, S. (2011). Gaze modulated non-propositional reasoning: Further evidence for spatial representation of reasoning premises. *Neuroscience, 173*, 110-115.

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature, 232*(5311), 456-458.

Buckmaster, C. A., Eichenbaum, H., Amaral, D. G., Suzuki, W. A., & Rapp, P. R. (2004). Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. *Journal of Neuroscience, 24*(44), 9811-9825.

Byrne, R. W., & Whiten, A. (1997). Machiavellian intelligence. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 1-17). Cambridge: Cambridge University Press.

Chapais, B. (1983). Dominance, relatedness and the structure of female relationships in rhesus monkeys. In R. A. Hinde (Ed.), *Primate social relationships: An integrated approach* (pp. 208-217). Sunderland, Massachusetts: Sinauer Associates.

Chapais, B. (1988). Rank maintenance in female Japanese macaques: Experimental evidence for social dependency. *Behaviour*, *104*, 41-59.

Chase, I. D. (1974). Models of hierarchy formation in animal societies. *Behavioral Science, 19*(6), 374-382.

Chase, I. D. (1980). Social process and hierarchy formation in small groups: A comparative perspective. *American Sociological Review, 45*(6), 905-924.

Chase, I. D. (1982). Dynamics of hierarchy formation: The sequential development of dominance relationships. *Behaviour, 80*, 218-240.

Chase, I. D. (1986). Explanations of hierarchy structure. *Animal Behaviour, 34*, 1265-1266.

Chen, S., Swartz, K. B., & Terrace, H. S. (1997). Knowledge of the ordinal position of list items in rhesus monkeys. *Psychological Science, 8*(2), 80-86.

Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world*. Chicago: University of Chicago Press.

Crockford, C., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2007). Baboons eavesdrop to deduce mating opportunities. *Animal Behaviour*, *73*, 885-890.

D'Amato, M., & Lachman, R. (1958). Secondary reinforcement as affected by reward schedule and the testing situation. *journal of Comparative Physiological Psychology, 51*, 737-741.

D'Amato, M. R. (1991). Comparative cognition: Processing of serial order and serial pattern. In L. Dachowski & C. F. Flaherty (Eds.), *Current topics in animal learning: Brain, emotion and cognition* (pp. 165-185). Hillsdale, New Jersey: L Erlbaum.

D'Amato, M. R., & Colombo, M. (1990). The symbolic distance effect in monkeys *(Cebus apella*). *Animal Learning & Behavior, 18*(2), 133-140.

D'Amato, M. R., & Van Sant, P. (1988). The person concept in monkeys (Cebus apella). *J. Exper. Psychol.: Anim. Behav. Processes, 14*(1), 43-55.

Davis, H. (1992a). Logical transitivity in animals. In W. K. F. Honig, J.G. (Ed.), *Cognitive aspects of stimulus control* (pp. 405-429). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Davis, H. (1992b). Transitive inferences in rats (*Rattus norvegicus). Journal of Comparative Psychology, 106*, 342-349.

De Vries, H., & Appleby, M. C. (2000). Finding an appropriate order for a hierarchy: a comparison of the I&SI and the BBS methods. *Anim. Behav., 59*, 239-245.

de Waal, F. B. M., & Luttrell, L. M. (1985). The formal hierarchy of rhesus monkeys: An investigation of the bared-teeth display. *Am. J. Primatol., 9*, 73-85.

Deaner, R. O., Khera, A. V., & Platt, M. L. (2005). Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Current Biology, 15*(6), 543-548.

Dittus, W. P. J. (1980). The social regulation of primate populations: A synthesis. In D. G. Lindburg (Ed.), *The macaques: Studies in ecology, behavior, and evolution* (pp. 263-286). New York: Van Nostrand Reinhold Company.

Domjan, M. (2004). *The Essentials of Conditioning and Learning*. Belmont, CA: Wadsworth Publishing.

Drea, C. M. (1998). Social context affects how rhesus monkeys explore their environment. *Am. J. Primatol., 44*(3), 205-214.

Drea, C. M., & Wallen, K. (1999). Low-status monkeys “play dumb” when learning in mixed social groups. *Proc. Nat. Acad. Sci. US, 96*, 12965.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences of the United States of America, 94*(13), 7109-7114.

Ehardt, C. A., & Bernstein, I. S. (1986). Matrilineal overthrows in rhesus monkey groups. *Int. J. Primatol., 7*(2), 157-181.

Emery, N. J., Clayton, N. S., & Frith, C. D. (2007). Introduction. Social intelligence: from brain to culture. *Philosophical Transactions of the Royal Society B-Biological Sciences, 362*(1480), 485-488.

Fagot, J., & Bonte, E. (2010). 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). *Behavior Research Methods, 42*(2), 507-516.

Fagot, J., & De Lillo, C. (2011). A comparative study of working memory: Immediate serial spatial recall in baboons (Papio papio) and humans. *Neuropsychologia, 49*(14), 3870-3880.

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods, 41*(2), 396-404.

Fairbanks, L. A., & McGuire, M. T. (1984). Determinants of fecundity and reproductive success in captive vervet monkeys. *American Journal of Primatology, 7*(1), 27-38.

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience, 5*(5), 458-462.

Gillan, D. J. (1981). Reasoning in the chimpanzee: II. Transitive inference. *Journal of Experimental Psychology-Animal Behavior Processes, 7*(2), 150-164.

Glickman, S. E., & Sroges, R. W. (1966). Curiosity in zoo animals. *Behaviour, 26*, 151-&.

Goswami, U. (1995). Transitive relational mappings in 3 year olds and 4 year olds: The analogy of Goldilocks and the 3 bears *Child Development, 66*(3), 877-892.

Gouzoules, H. (1980). A description of genealogical rank changes in a troop of Japanese monkeys (Macaca fuscata). *Primates, 21*(2), 262-267.

Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2006). An fMRI analysis of the human hippocampus: Inference, context, and task awareness. *Journal of Cognitive Neuroscience, 18*(7), 1156-1173.

Grosenick, L., Clement, T. S., & Fernald, R. D. (2007). Fish can infer social rank by observation alone. *Nature, 445*, 429-432.

Hasset, J. M., Martin-Malivel, J., Lange, H., Fischer, A., & Wallen, K. (2007). *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 Comparative Cognition, Melbourne, FL.

Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., & Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus, 14*, 153-162.

Hogue, M.-E., Beaugrand, J. P., & Lague, P. C. (1996). Coherent use of information by hens observing their former dominant defeating or being defeated by a stranger. *Behavioural Processes, 38*, 241-252.

Holekamp, K. E., Sakai, S. T., & Lundrigan, B. L. (2007). Social intelligence in the spotted hyena (*Crocuta crocuta*). *Philosophical Transactions of the Royal Society B-Biological Sciences, 362*(1480), 523-538.

Holmes, K. J., & Lourenco, S. F. (2011). Common spatial organization of number and emotional expression: A mental magnitude line. *Brain and Cognition, 77*(2), 315-323.

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology*. Cambrigde: Cambridge University Press.

Jackson, W. M., & Winnegrad, R. L. (1988). Linearity in dominance hierarchies : A second look at the Individual Attributes Model. *Animal Behaviour, 36*, 1237-1240.

Janus, M. (1992). Interplay between various aspects in social relationships of young rhesus monkeys: Dominance, agonistic help, and affiliation. *American Journal of Primatology, 26*(4), 291-308.

Kamil, A. C. (2004). Sociality and the evolution of intelligence. *Trends Cogn Sci, 8*(5), 195-197.

Lazareva, O. F. (2012). Transitive inference in non-human animals. In T. R. Zentall & E. A. Wasserman (Eds.), *The Oxford Handbook of Comparative Cognition* (2nd ed.). Oxford: Oxford University Press.

Lazareva, O. F., Smirnova, A. A., Bagozkaja, M. S., Zorina, Z. A., Rayevsky, V. V., & Wasserman, E. A. (2004). Transitive responding in hooded crows requires linearly ordered stimuli. *Journal of the Experimental Analysis of Behavior, 82*(1), 1-19.

Lazareva, O. F., & Wasserman, E. A. (2006). Effect of stimulus orderability and reinforcement history on transitive responding in pigeons. *Behavioural Processes, 72*(2), 161-172.

Lazareva, O. F., & Wasserman, E. A. (2010). Nonverbal transitive inference: Effects of task and awareness on human performance. *Behavioral Processes, 83*, 99-112.

Lazareva, O. F., & Wasserman, E. A. (2012). Transitive inference in pigeons: Measuring the associative value of stimuli B and D. *Behavioural Processes, 89*(3), 244-255.

le Roux, A., & Bergman, T. J. (2012). Indirect rival assessment in a social primate, *Theropithecus gelada*. *Animal Behaviour, 83*, 249-255.

Logie, R. H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology Section A, 38*(2), 229-247.

Lovelace, E. A., & Snodgrass, R. D. (1971). Decion times for alphabetic order of letter pairs. *Journal of Experimental Psychology 88*(2), 258-264.

MacLean, E. L., Merritt, D. J., & Brannon, E. M. (2007). Transitive inference in two species of prosimian primates. *American Journal of Primatology, 69*, 102-102.

Maclean, E. L., Merritt, D. J., & Brannon, E. M. (2008). Social complexity predicts transitive reasoning in prosimian primates. *Animal Behaviour, 76*, 479-486.

Marsh, A. A., Blair, K. S., Jones, M. M., Soliman, N., & Blair, R. J. R. (2009). Dominance and submission: The Ventrolateral prefrontal cortex and responses to status cues. *Journal of Cognitive Neuroscience, 21*(4), 713-724.

McGonigle, B. O., & Chalmers, M. (1977). Are Monkeys Logical. *Nature, 267*(5613), 694-696.

Merritt, D. J., & Terrace, H. S. (2011). Mechanisms of inferential order judgments in humans (Homo sapiens) and rhesus monkeys (Macaca mulatta). *J. Comp. Psychol., 125*(2), 227-238.

Miller, H. J., Jr. (1976). Matching based hedonic scaling in the pigeon. *Journal of Experimental Analysis of Behavior, 26*, 335-347.

Montgomery, K. C. (1953). Concerning the use of analysis of variance on latency data. [Note]. *American Journal of Psychology, 66*(1), 131-135.

Morand-Ferron, J., Cole, E. F., Rawles, J. E. C., & Quinn, J. L. (2011). Who are the innovators? A field experiment with 2 passerine species. *Behavioral Ecology, 22*(6), 1241-1248.

Moses, S. N., Brown, T. M., Ryan, J. D., & McIntosh, A. R. (2010). Neural system interactions underlying human transitive inference. *Hippocampus, 20*(8), 894-901.

Moses, S. N., Ostreicher, M. L., & Ryan, J. D. (2010). Relational framework improves transitive inference across age groups. *Psychological Research-Psychologische Forschung, 74*(2), 207-218.

Moses, S. N., Villate, C., & Ryan, J. D. (2006). An investigation of learning strategy supporting transitive inference performance in humans compared to other species. *Neuropsychologia, 44*, 1370-1387.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519-1520.

Nagode, J. C., & Pardom, J. V. (2002). Human hippocampal activation during transitive inference. *Learning and Memory, 13*(7), 939-944.

Paxton, R., Basile, B. M., Adachi, I., Suzuki, W. A., Wilson, M., & Hampton, R. R. (2010). Rhesus monkeys (*Macaca mulatta*) rapidly learn to select dominant individuals in videos of artificial social interactions between unfamiliar conspecifics. *Journal of Comparative Psychology, 124*(4), 395-401.

Paz-y-Miño, G., Bond, A., Kamil, A. C., & Balda, R. P. (2004). Pinyon jays use transitive inference to predict social dominance. *Nature 430*, 778-781.

Pears, R., & Bryant, P. (1990). Transitive inferences by young children about spatial position. *British Journal of Psychology, 81*, 497-510.

Phillips, W. A., & Christie, D. F. M. (1977). Interference with visualization. *The Quarterly Journal of Experimental Psychology, 29*(4), 637-650.

Piaget, J. (1960). *Logic and psychology*. New York: Basic Books, Inc.

Prado, J., Van der Henst, J. B., & Noveck, I. A. (2008). Spatial associations in relational reasoning: Evidence for a SNARC-like effect. *Quarterly Journal of Experimental Psychology, 61*(8), 1143-1150.

Previtali, P., de Hevia, M. D., & Girelli, L. (2010). Placing order in space: the SNARC effect in serial learning. *Experimental Brain Research*(201), 599–605.

Rapp, P. R., Kansky, M. T., & Eichenbaum, H. (1996). Learning and memory for hierarchical relationships in the monkey: Effects of aging. *Behavioral Neuroscience, 110*(5), 887-897.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Oavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. B. W. F. Prokasy (Ed.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appelton-Century-Crofts.

Roberts, W. A., & Phelps, M. T. (1994). Transitive inference in rats: A test of the spatial coding hypothesis. *Psychological Science, 5*(6), 368-374.

Rommeck, I., Capitanio, J. P., Strand, S. C., & McCowan, B. (2011). Early social experience affects behavioral and physiological responsiveness to stressful conditions in infant rhesus macaques (Macaca mulatta). *American Journal of Primatology, 73*(7), 692-701.

Russell, J., McCormack, T., Robinson, J., & Lillis, G. (1996). Logical (versus associative) performance on transitive reasoning tasks by children: Implications for the status of animals' performance. *Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology, 49*(3), 231-244.

Schrier, A. M., Angarella, R., & Povar, M. L. (1984). Studies of concept formation by stumptailed monkeys: Concepts humans, monkeys, and letter A. *J. Experimental Psychol: Anim. Behav. Processes, 10*(4), 564-584.

Schrier, A. M., & Brady, P. M. (1987). Categorization of natural stimuli by monkeys (Macaca mulatta): Effects of stimulus set size and modification of exemplars. *J. Exp. Psychol.: Anim. Behav. Processes, 13*(2), 136-143.

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics, 66*(4), 651-664.

Shaki, S., & Fischer, M. H. (2008). Reading space into numbers - a cross-linguistic comparison of the SNARC effect. *Cognition, 108*(2), 590-599.

Shettleworth, S. J. (2009). *Cognition, Evolution, and Behavior, Second Edition*. New York Oxford University Press.

Siemann, M., & Delius, J. D. (1996). Influences of task concreteness upon transitive responding in humans. *Psychological Research-Psychologische Forschung, 59*(2), 81-93.

Siemann, M., & Delius, J. D. (1998). Algebraic learning and neural network models for transitive and non-transitive responding. *European Journal of Cognitive Psychology, 10*(3), 307-334.

Siemann, M., Delius, J. D., Dombrowski, D., & Daniel, S. (1996). Value transfer in discriminative conditioning with pigeons. *Psychological Record, 46*(4), 707-728.

Siemann, M., Delius, J. D., & Wright, A. A. (1996). Transitive responding in pigeons: Influences of stimulus frequency and reinforcement history. *Behavioural Processes, 37*(2-3), 185-195.

Silk, J. B., Alberts, S. C., & Altmann, J. (2003). Social bonds of female baboons enhance infant survival. *Science, 302,* 1231-1234.

Smith, C., & Squire, L. (2005). The importance of awareness in transitive inference. *Journal of Cognitive Neuroscience*, 239-239.

Steirn, J. N., Weaver, J. E., & Zentall, T. R. (1995). Transitive inference in pigeons: Simplified procedures and a test of value transfer theory. *Animal Learning & Behavior, 23*(1), 76-82.

Thierry, B. (2007). The Macaques: A double-layered social organization. In C. J. Campbell, A. Fuentes, K. C. Mackinnon, M. Panger & S. K. Bearder (Eds.), *Primates in Perspective* (pp. 224-239). New York: Oxford University Press.

Treichler, F. R., & Raghanti, M. A. (2010). Serial list combination by monkeys (*Macaca mulatta*): test cues and linking. *Animal Cognition, 13,* 121-131.

Treichler, F. R., Raghanti, M. A., & Van Tilburg, D. N. (2003). Linking of serially ordered lists by macaque monkeys (*Macaca mulatta*): List position influences. *Journal of Experimental Psychology-Animal Behavior Processes, 29*(3), 211-221.

Treichler, F. R., Raghanti, M. A., & Van Tilburg, D. N. (2007). Serial list linking by macaque monkeys (*Macaca mulatta)*: List property limitations. *Journal of Comparative Psychology, 121*(3), 250-259.

Treichler, F. R., & Van Tilburg, D. (1996). Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List linking. *J. Exper. Psychol.: Anim. Behav. Proc., 22*(3), 105-117.

Treichler, F. R., & Van Tilburg, D. (1999). Training requirements and retention characteristics of serial list organization by macaque monkeys. *Animal Cognition, 2,* 235-244.

Treichler, F. R., & VanTilburg, D. (1996). Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List linking. *Journal of Experimental Psychology-Animal Behavior Processes, 22*(1), 105-117.

Van der Jeugd, A., Goddyn, H., Laeremans, A., Arckens, L., D'Hooge, R., & Verguts, T. (2009). Hippocampal involvement in the acquisition of relational associations, but not in the expression of a transitive inference task in mice. *Behavioral Neuroscience, 123*(1), 109-114.

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes, 78,* 313-334.

von Fersen, L., & Lea, S. E. G. (1990). Category discrimination by pigeons using five polymorphous features. *J. Exper. Anal. Behav., 54*, 69-84.

Von Fersen, L., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology-Animal Behavior Processes, 17*(3), 334-341.

Washburn, D. A., & Astur, R. S. (1998). Nonverbal working memory of humans and monkeys: Rehearsal in the sketchpad? *Memory & Cognition, 26*(2), 277-286.

Weiβ, B. M., Kehmeier, S., Schloegl, C., & (2010). Transitive inference in free-living greylag geese, *Anser anser*. *Animal Behaviour, 79,* 1277-1283.

Woocher, F. D., Glass, A. L., & Holyoak, K. J. (1978). Positional discriminability in linear orderings. *Memory and Cognition*, *6*(2), 165-173.

Wynne, C. D. L. (1995). Reinforcement accounts for transitive inference performance. *Animal Learning & Behavior*, *23*(2), 207-217.

Wynne, C. D. L. (1997). Pigeons transitive inferences: Tests of simple accounts of a complex performance. *Behavioral Processes*, *39*, 95-112.

Wynne, C. D. L. (1998). A minimal model of transitive inference. In C. D. L. Wynne & J. E. R. Staddon (Eds.), *Models of action: Mechanisms for adaptive behavior* (pp. 269-307). Mahwah, NJ: Erlbaum.

Zalesak, M., & Heckers, S. (2009). The role of the hippocampus in transitive inference. *Psychiatry Research-Neuroimaging*, *172*(1), 24-30.

Zentall, T. R., & Sherburne, L. M. (1998). The transfer of value in simultaneous discriminations: Implications for cognitive and social processes. *Brain and Values*, 323-336.