Distribution Agreement

Signature:

______________________________ _______________

Jennifer J. Pokorny Date

Social categorization in capuchin monkeys (*Cebus apella*): In-group vs. out-group

By

Jennifer J. Pokorny
Doctor of Philosophy

Psychology

Frans B. M. de Waal, Ph.D.
Advisor

Robert R. Hampton, Ph.D.
Committee Member

William D. Hopkins, Ph.D.
Committee Member

Philippe Rochat, Ph.D.
Committee Member

Kim Wallen, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

Date

Social categorization in capuchin monkeys (*Cebus apella*): In-group vs. out-group

By

Jennifer J. Pokorny
M.A., Emory University, 2007

Advisor: Frans B. M. de Waal, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the Graduate School of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2009

Abstract

Social categorization in capuchin monkeys (*Cebus apella*): In-group vs. out-group

By Jennifer J. Pokorny

The social groups of most primates consist of individuals of different ages, sexes, ranks and relatedness. This complexity requires that group members recognize and remember the individuals within ones own group. Furthermore, it is necessary to distinguish these individuals from those outside the group as outsiders pose a threat to a group's food and mating resources. Humans typically rely on faces as one way to recognize individuals, allowing us to distinguish between familiar and unfamiliar individuals, as well as group individuals into other categories, such as friend, family, stranger, male or female, young or old. A previous study in our laboratory demonstrated that capuchin monkeys (*Cebus apella*), a New World primate species, could discriminate faces of conspecifics using an oddity task (Pokorny & de Waal, in press). In the current study, we examined capuchins' ability to categorize the faces of conspecifics as belonging to the in-group or out-group, relative to the subject, using the oddity paradigm. Follow-up tests ruled out alternative explanations for how subjects could solve the task, primarily addressing possible color cues in the images. Subjects successfully transferred to novel sets of images, suggesting that they used knowledge of familiar individuals depicted in the images to solve the task of selecting either in-group or out-group individuals. We also examined whether subjects were using an oddity concept to perform the task or whether they were using rules, such as: 1) select the in-group member, 2) select the out-group member. Findings reveal that subjects are likely doing the latter and not applying the oddity concept to the dimension of group membership. Overall, this study provides supporting evidence that nonhuman primates can recognize conspecifics in two-dimensional images and differentiate in-group from out-group members.

Social categorization in capuchin monkeys (*Cebus apella*): In-group vs. out-group

By

Jennifer J. Pokorny
M.A., Emory University, 2007

Advisor: Frans B. M. de Waal, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2009

Table of Contents

Introduction 1

- Face discrimination and recognition in nonhuman primates 1
- Knowledge of familiar and unfamiliar conspecifics 4
- Natural concepts 5
- Knowledge of social relationships 9
- Preliminary studies 11
- Current study 15
- Characteristics of the oddity paradigm 15
- Alternative explanations for the categorization performance of capuchin faces 18
- Brown capuchin monkeys (*Cebus apella*) 19

General Methods 20

- Subjects and housing 20
- Apparatus 21
- Stimuli 21
- Procedure 22
- Previous training and experience 25
- Data collection and analysis 25

Experiment 1: Group Membership Task 27

- Procedure 27
- Stimuli 28

Results Experiment 1 28

Discussion Experiment 1 34

Methods Experiment 2: Alternatives for group membership categorization –

Generalize the identity of the depicted individual 37

Procedure 38

Specific predictions 39

Results Experiment 2 39

Discussion Experiment 2 43

Methods Experiment 3: Alternatives for group membership categorization –

Color cues – grayscale 45

Procedure 45

Specific predictions 47

Results Experiment 3 47

a) Analysis of color or luminance differences 47

b) Testing with grayscale images 52

Discussion Experiment 3 66

Methods Experiment 4: Alternative for group membership categorization –

Color cues – self 67

Procedure 68

Stimuli 68

Specific predictions for before and after mirror exposure 69

Results Experiment 4 70

Discussion Experiment 4 73

Methods Experiment 5: Categorization or oddity? 75

Procedure 75

Results Experiment 5 77

Discussion Experiment 5 79

General Discussion 80

References 87

Appendix A 100

List of Tables

Table 1. Individual's correct performance on group membership transfer trials, 40 trials each condition

Table 2. Comparison of typical and juvenile probe trials by subject for In-group Odd and Out-group odd conditions

Table 3. Comparison of latencies to respond to typical (N = 40) and juvenile probe trials (N = 40) in each condition

Table 4. Comparison of the latency to respond by subject in the In-group Odd and Out-group Odd condition for typical trials (N = 40) and juvenile probe (N = 40) trials

Table 5. Independent samples t-tests for the mean and median levels of red, green, blue and the luminance from Group 1 (N = 54) and Group 2 (N = 54) used in the initial Group Membership stimulus set

Table 6. Independent samples t-test for the mean and median levels of red, green, blue and the luminance of photos from Group 1 (N = 30) and Group 2 (N = 27) used in the transfer test (Experiment 1)

Table 7. Independent samples t-test for the mean and medial levels of red, green, blue and the luminance of photos of the juveniles from Group 1 (N = 16) and Group 2 (N = 16) used in Experiment 2

Table 8. Independent samples t-test for the mean and median levels of the luminance of photos used in the initial stimulus set, converted to grayscale in Experiment 3 (Group 1: N = 54; Group 2: N = 54)

Table 9. Independent samples t-test for the mean and median levels of the luminance of photos used in the transfer set, converted to grayscale in Experiment 3 (Group 1: N = 30; Group 2: N = 27)

Table 10. Independent samples t-test for the mean and median luminance levels of new photos in grayscale used in Experiment 3, the luminance intentionally reversed for the two groups (Group 1: N = 30; Group 2: N = 18)

Table 11. McNemar tests comparing performance on typical trials and probe trials of the original stimulus set converted to grayscale

Table 12. Means of each condition by subject for typical trials and probe trials of the original stimulus set converted to grayscale

Table 13. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and grayscale probe trials

Table 14. Comparison of latencies to respond to typical (N = 40) and grayscale original probe trials (N = 40) in each condition

Table 15. Comparison of the latency to respond by subject in the In-group Odd and Out-group Odd condition for typical trials (N = 40) and grayscale original probe (N = 40) trials

Table 16. McNemar tests comparing performance on typical trials and probe trials of the transfer stimulus set converted to grayscale

Table 17. Means of each condition by subject for typical trials and probe trials of the transfer stimulus set converted to grayscale

Table 18. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and probe trials

Table 19. Comparison of response latencies to typical (N = 40) and grayscale transfer probe trials (N = 40) in each condition

Table 20. Comparison of the latency to respond by subject in the In-group Odd and Out-group Odd condition for typical trials (N = 40) and grayscale transfer probe trials (N = 40)

Table 21. 2-way chi square contingency tests comparing performance on transfer trials presented in color and probe trials of the transfer stimulus set converted to grayscale

Table 22. McNemar tests comparing performance on typical trials and probe trials of the new grayscale stimulus set intentionally reversed in luminance

Table 23. Means of each condition by subject for typical trials and probe trials of the novel grayscale stimulus set reversed in luminance

Table 24. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and probe trials

Table 25. Comparison of response latencies to typical (N = 40) and grayscale luminance reversed probe trials (N = 40) in each condition

Table 26. Specific predictions for Self Group Membership test

Table 27. McNemar tests comparing performance on typical trials presenting four pictures and probe trials presenting two pictures

Table 28. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and probe trials

Table 29. Comparison of response latencies to typical trials in which four images appeared (N = 40) and probe trials in which two images appeared (N = 40)

Table 30: Individual's correct performance on Group Membership trials when In-group Odd and Out-group Odd conditions were both presented within sessions, 13 (Wilma) or 19 sessions (Bias and Mason) per subject

List of Figures

Figure 1. Comparing group performance on Individual Recognition task for a) transfer 1 and b) transfer 2 of previous Individual Recognition study

Figure 2. Example of a single Group Membership trial

Figure 3. Individual performance on In-group and Out-group conditions when switching from one condition to the other

Figure 4. Individual's performance correct on blocked Group Membership trials, before and after transfer to new images

Figure 5. Individual's performance on Group Membership trials with original stimuli and on probe trials with novel images of juveniles in both the Out-group Odd and In-group Odd conditions

Figure 6. Individual's performance on Group Membership trials with original stimuli and on probe trials with the original stimulus set converted to grayscale in both the Out-group Odd and In-group Odd conditions

Figure 7. Individual's performance on Group Membership trials with original stimuli and on probe trials with the transfer stimulus set converted to grayscale in both the Out-group Odd and In-group Odd conditions

Figure 8. Individual's performance on Group Membership trials with original stimuli and on probe trials with the novel grayscale stimulus set reversed in luminance

Figure 9. Displays the ideal pattern that should be found if subjects categorize their own image as either a) an in-group member or b) an out-group member

Figure 10. Individual's performance on the Group Membership task when images of own portrait were presented prior to mirror exposure

Figure 11. Individual's performance on the Group Membership task when images of own portrait were presented after mirror exposure

Figure 12. Representation of Experiment 5 trial types

Figure 13. Individual performance on typical and probe trials for In-group Odd and Out-group Odd conditions in Experiment 5

Figure 14. Group performance on initial sessions where In-group Odd and Out-group Odd conditions were both presented within sessions (mixed sessions)

Both in captivity and in the wild, capuchin monkeys typically respond agonistically to outsiders (Cooper, Bernstein, Fragaszy, & de Waal, 2001; Crofoot, 2007; Fragaszy, Baer, & Adams-Curtis, 1994; Spironello, 2001), although the level of direct aggression varies. This behavior is directed specifically toward out-group individuals, mainly by and to males, and is not behavior that is normally displayed toward members of the in-group. In wild white-faced capuchins (*Cebus capucinus*), both males and high-ranking females are involved in these territorial attacks (Perry, 1996), whereas in brown capuchins (*Cebus apella*), typically only the alpha male or other adult males are involved in these interactions (Defler, 1982; Janson, 1986). These natural observations indicate that capuchins can discriminate in-group from out-group members, and that they know the identity of these individuals. What we do not know is what information they use to make this discrimination. While multiple cues are likely used, for our purposes here, we will focus on the visual modality, with particular attention to the face. In humans, faces provide a rapid means for determining the identity of individuals, as well as provide information about the age, sex, and emotional state. Nonhuman primates, similar to humans, have a repertoire of facial expressions that inform others of one's motivation or potential actions (Andrew, 1963; Darwin, 1872; van Hooff, 1962, 1967). Given the importance of these signals, it is likely that in nonhuman primates the face provides information about individual identity, as it does in humans.

Face discrimination and recognition in nonhuman primates

There is evidence that nonhuman primates are able to discriminate faces of conspecifics when presented as a two-dimensional image. One task used to evaluate this is the visual paired comparison (VPC) task, which is a task that requires no training as it takes advantage of a natural bias to look at novelty. The VPC paradigm presents an image to a subject, long enough for the subject to explore and familiarize themselves with the image. The image then disappears and two comparison images appear. One image is the same as the image initially presented and the other is novel. If subjects detect a difference in the two comparison images, they should look more at the novel image than the familiarized image. A study utilizing the VPC paradigm presented humans and rhesus monkeys (*Macaca mulatta*) with images of objects and faces of humans and rhesus monkeys. Both groups demonstrated a novelty preference for objects and conspecific faces, but not for faces of the other primate species (Pascalis & Bachevalier, 1998). Similarly, Tonkean macaques (*M. tonkeana*) and brown capuchins show this species-specific effect when presented with images of humans, conspecifics and six other nonhuman primate species (Dufour, Pascalis, & Petit, 2006). Whereas these studies indicate that individuals have the capacity to detect identity changes only of conspecifics, another study found that cotton-top tamarins (*Saguinus oedipus*), a New World monkey species, discriminated faces of both conspecifics and humans in a VPC task (Neiworth, Hassett, & Sylvester, 2007), illustrating the role that experience can play in the ability to discriminate faces.

The aforementioned studies demonstrate the ability of several primate species to discriminate the faces of conspecifics, and possibly other highly familiar species. The recognition of conspecifics from two-dimensional images has also been investigated by testing whether subjects can generalize the identity of the individual depicted across multiple viewpoints or other visual changes of the image. Some of these studies employ tasks that require an explicit response from the subject following a training period, as opposed to the VPC task that takes advantage of an untrained looking response. Rosenfeld & van Hoesen (1979) presented rhesus monkeys with two pictures of faces of two unfamiliar conspecifics in a discrimination task. One image was chosen as the correct image to select, the positive stimulus, and the other was the negative stimulus, the one to avoid. After subjects acquired the task, they were transferred to images taken from other viewpoints of the same two conspecifics. Transfer performance was gauged by comparing transfer acquisition to initial acquisition, which was shorter in the transfer phase, suggesting positive transfer to the novel viewpoints. A similar discrimination task with long-tailed macaques (*M. fascicularis*) found that upon transfer to five additional pairs of images depicting the same two individuals used in training but differing in viewpoint, subjects performed above 67% (one case 43%) in the first 100 trials (Bruce, 1982). Heywood and Cowey (1992) also employed a discrimination task and during the transfer phase added two individuals to the initial two individuals used in the training phase. Again, subjects performed well, averaging 76 errors before reaching a performance criterion of 90% correct.

A more recent study directly compared the ability of rhesus monkeys and chimpanzees (*Pan troglodytes*) to match non-identical facial images of unfamiliar conspecifics, thereby matching the identity of the depicted individual (Parr, Winslow, Hopkins, & de Waal, 2000). Species-specific differences were found, such that chimpanzees acquired the task much faster, in two to 14 exposures of 14 novel problems, than did the rhesus monkeys, who needed over 200 trials on 15 unique problems before performing significantly above chance. Rhesus monkey subjects were given another transfer phase consisting of 10 novel trials and this time their performance was similar to the chimpanzees' initial performance, reaching criterion (significantly above chance) within two to 14 exposures. While this final acquisition may have been similar, the rhesus monkeys required training on the individual recognition task first, whereas the chimpanzees performed the task immediately, suggesting that chimpanzees more readily generalized across viewpoint changes of depicted individuals than did rhesus monkeys.

Capuchin monkeys have also demonstrated recognition of both in-group and out-group conspecifics using an oddity task (Pokorny & de Waal, in press). Subjects were first presented with three identical images of one individual and one image of another individual, both either from the in-group or the out-group relative to the subject, testing the ability to discriminate conspecific faces. To determine if subjects were able to generalize the identity of the individual depicted in the image, subjects were given a recognition task that presented three different images of one individual and one image of a different individual. Subjects readily transferred to this recognition task, performing significantly above chance in the first 50 novel trials. Two additional transfer phases were carried out, presenting both novel stimuli and novel combinations of stimuli, and subjects continued to maintain above chance performance in the first 50 novel trials of each transfer. Performance did not differ between in-group and out-group conspecifics in any of the face recognition tests.

Knowledge of familiar and unfamiliar conspecifics

The above studies demonstrated that subjects were able to discriminate and recognize images of conspecifics. Only a handful of experiments have directly assessed familiarity, whether subjects know which individuals are familiar and which are unfamiliar. Those studies have presented two-dimensional images of both familiar and unfamiliar conspecifics to assess differences in the outcome variable of interest, indirectly examining the concept of familiarity. One study suggested that long tail macaques could match body parts of familiar, but not unfamiliar, conspecifics, indicating that they have knowledge of the whole body of familiar individuals (Dasser, 1987). This study was conducted before computerized technology, however, and has never been successfully replicated in the twenty years since it was conducted. In addition, one must be wary forming conclusions from this study as the reported numbers of comparison trials between familiar and unfamiliar were vastly different. For example, subjects successfully performed 44 out of 60 familiar trials, and only 7 out of 18 unfamiliar trials. If one looked at the first 18 familiar trials compared to the 18 unfamiliar trials, the conclusion may have been very different. But there is now convincing evidence that chimpanzees can match the face of familiar conspecifics to an image of the corresponding behind (anogenital region) of the same individual, suggesting whole-body knowledge of group mates (de Waal & Pokorny, 2008).

Aside from performance differences on an explicit task, other studies have demonstrated differences in physiological measures while subjects passively viewed images. Boysen and Berntson (1986, 1989) measured the heart rate of chimpanzees while they viewed images of either conspecifics or humans and found that the heart rate varied depending on the quality of the relationship the subject had with the individual depicted. Similarly, the familiarity of the individual depicted resulted in the modulation of two components of event related potentials recorded while squirrel monkeys (*Saimiri sciureus*) viewed images of conspecifics (Pineda, Sebestyen, & Nava, 1994). Whereas these studies were not directly investigating the concept of group membership, they do provide evidence that subjects connected images with real-life individuals, discriminating between familiar and unfamiliar parties.

Natural concepts

A concept is a knowledge of relations (Adams-Curtis, 1990) or knowledge that allows one to group together similar items (Zentall, Galizio, & Critchfield, 2002). Several studies have investigated natural concepts in nonhuman primates, these concepts being "natural" because they are based on stimuli present in nature and biologically relevant to the individual

(Herrnstein, Loveland, & Cable, 1976). The basis of conceptualization is the ability to categorize items, and this is what the majority of animal studies have examined – categorization ability. For instance, rhesus monkeys discriminated between images that contained rhesus monkeys and those that did not, followed by successfully discriminating between rhesus monkeys and Japanese macaques (Yoshikubo, 1985). The authors conclude that subjects demonstrated a concept of rhesus versus non-rhesus placing items into the respective categories. Similarly, squirrel monkeys also exhibited a category of "squirrel monkey," successfully transferring to new images on the first trial (Phillips, 1996). Other studies have found knowledge of categories such as humans (D'Amato & Van Sant, 1988; Schrier, Angarella, & Povar, 1984; Schrier & Brady, 1987), monkeys or animals in general (Brown & Boysen, 2000; Neiworth, Parsons, & Hassett, 2004; Roberts & Mazmanian, 1988; Schrier et al., 1984; Vonk & MacDonald, 2002), food (Bovet & Vauclair, 1998; Deputte, Pelletier, & Barbe, 2001; Inoue et al., 2008; Santos, Hauser, & Spelke, 2001), plants (Tanaka, 2001), gender/sex, (Inoue et al., 2008; Koba & Izumi, 2006) and even letters or numbers (Schrier et al., 1984; Vauclair & Fagot, 1996).

Similar tasks have also investigated knowledge of social relationships, such as mother-offspring (Dasser, 1988), social group, mated pairs and siblings (Vonk, 2002). Dasser (1988) presented a long-tailed macaque with images of familiar mother and daughter pairs in a two-choice stimulus discrimination task. Once the subject acquired the task, a transfer session of 14 unique trials was conducted and the subject identified all mother-offspring pairs. No further tests were done to rule out whether cues such as physical similarity could have been used to identify the pairs. Other subtle cues, such as facial expressions or body postures, may have been perceived by the subject but not the experimenter. Again, this possibility was not explored. It is feasible that this macaque was able to use their knowledge of the relationships between familiar group mates and connect this information with the images presented in the task. However, this study has never been replicated, so we await additional evidence.

Vonk (2002) also investigated knowledge of social relationships, this time including social group, mated pairs, siblings, in addition to mother-offspring. One major difference from the Dasser (1988) study, aside from using a different task, is that the images were of unfamiliar individuals, and even unfamiliar species. Two orangutans and one gorilla performed a delayed match-to-sample task, the correct image being the one that matched the type of social relationship demonstrated in the sample. Results revealed that subjects performed significantly above chance when examining the first several hundred trials performed.

This appears to support the notion that nonhuman primates can form categories of social relationships, but one should be cautious of this conclusion, particularly for the category of mated pairs. For mated pairs and social groups, one subject did not perform above chance. Aside from that, it is surprising that subjects could match pairs of unfamiliar species using these categories. For instance, orangutan and gorilla social structures do not include the category of "mated pair" so it is not clear why they would be tested on this concept. The definition of the category was also ambiguous, as no clear description was made as to what constituted a mated pair and in one instance, the term *mating* pair was used, which would suggest different behaviors being displayed in the image. An example was: the sample was of a mated pair of chimpanzees, the match was a mated pair of birds and the foil was an image of chimpanzee siblings (Vonk, 2002, p. 266). One should note that chimpanzees also lack "mated pair" in their social structure, so it remains vague as to what image was presented. In addition, the description of siblings was unclear and it is possible that play behavior was displayed more frequently in sibling images. If this is the case, then subjects may have been matching observed behaviors. We do know that rhesus monkeys are able to select the dominant individual after watching a brief video of conspecifics interacting with one another (Bovet & Washburn, 2003), so behavioral displays are salient cues for nonhuman primates. However, it is difficult to determine how well subjects would be able to interpret behaviors of unfamiliar species from a still photo. Our study also investigated knowledge of social relationships in capuchin monkeys, specifically the knowledge of group membership. The strength of our study is that we are examining the categories of in-group and out-group within their own species, which we know capuchins behaviorally differentiate (Cooper et al., 2001; Crofoot, 2007; Fragaszy et al., 1994; Spironello, 2001).

The reason for not immediately concluding that these studies demonstrated conceptual knowledge is because one must establish that all the members of a stimulus set in a given category are responded to in the same manner in a different situation, demonstrating an equivalence class (Lea, 1984). Some studies do address this by reversing the reward contingencies, as typically subjects are trained to respond to one of the categories (S+) and withhold a response to the other category (S-). The categories are reversed and one assesses whether the reversal applied to all members of the category (Jitsumori, 2004). Others also examine whether subjects are able to transfer to novel stimuli, as this demonstrates that subjects have knowledge of the categories and the features or functions that make up those categories, rather than a trained preference for certain oft-repeated stimuli over others. This is particularly important for natural concepts as they are often referred to as open-ended categories (Herrnstein & Loveland, 1964), meaning that there is no single feature that determines membership in a given category (Jitsumori, 2004). However, others have argued that there is an explicit manner in which to demonstrate equivalence classes, namely one must first establish conditional discriminations (A→ B and B→ C) and then test for three emergent relations: reflexivity (A→ A, B→ B, C→ C), symmetry (B→ A, C→ B) and transitivity (C→ A) (Sidman, 1994; Sidman & Tailby, 1982). Studies that investigate these relations typically use artificial stimuli, not stimuli that may be more "natural" for subjects. The lack of positive findings in nonhuman primates for all three relations may be due to the fact that researchers are testing this in such an abstract manner that is not relevant to the subject. Nonetheless, because this has not been conclusively demonstrated in nonhuman primates, specifically monkeys, we will address these studies, and ours, as one of categorization, not of conceptualization. One should also note that this is an over-simplification of concept learning with regards to nonhuman primates, and animals in general, as equivalence classes are one component of concept formation, of which there are many levels (e.g. perceptual, associative, relational, analogical), each with their own criteria of how items are grouped, essentially what holds those items together in a given category (see for reviews: Shettleworth, 1998; Tomasello & Call, 1997; Vauclair, 2002; Wasserman & Zentall, 2006; Zentall et al., 2002; Zentall, Wasserman, Lazareva, Thompson, & Rattermann, 2008).

Knowledge of social relationships

Aside from knowing one's own relationship with individuals within a social group, nonhuman primates also have knowledge of third-party relationships, the relationships between others in the group. The knowledge of relationships in which the individual is not itself involved has been described by Kummer (1971), and was termed 'triadic awareness' by de Waal (1982, p. 175) and 'non-egocentric' social knowledge by Cheney and Seyfarth (1990a, p. 61). Auditory playback studies have provided support for this type of knowledge in nonhuman primates. When calls of a juvenile vervet monkey were played to a group of individuals, including the juvenile's mother, other individuals looked to the mother, often before the mother herself responded (Cheney & Seyfarth, 1980). This suggests that the others recognized who the call was from and looked to the individual most likely to respond, the mother. In another example, spectrographic analysis of the screams rhesus macaque juveniles gave during various agonistic interactions with other group members revealed that there was a difference in the screams depending on the intensity of the interaction as well as the status of the opponent (Gouzoules, Gouzoules, & Marler, 1984). When these screams were played to the mother of the juvenile, mothers responded differentially depending on the rank and kinship of the opponent. The information decoded in the screams informed the mother in what circumstance the juvenile was most in need of her assistance and she responded appropriately.

Redirected aggression, present in many species, is a behavior that is directed specifically toward the kin of individuals who were previously involved in an altercation. Specifically, the victim of an attack may retaliate against the kin of the aggressor, suggesting that they know which individuals are closely associated with the attacker (Aureli, Cozzolino, Cordischi, & Schucchi, 1992). This can be even more complex and further removed in vervets who will threaten an individual if that individual's kin and their own kin fought earlier in the day (Cheney & Seyfarth, 1986, 1990b, 1990c). This knowledge is not restricted to aggression, as it has been observed through reconciliation as well (Judge, 1991).

Soliciting aid and joining in alliances also demonstrates knowledge of relationships between other members of the group. In a group of wild white-faced capuchins (*Cebus capucinus*), individuals involved in a fight will solicit aid from others in the group. It was found that they preferentially recruit individuals with whom they have affiliated relationships and that are higher in rank than their opponent (Perry, Barrett, & Manson, 2004), as has also been demonstrated in chimpanzees (de Waal & van Hooff, 1981) and male bonnet macaques (Silk, 1999).

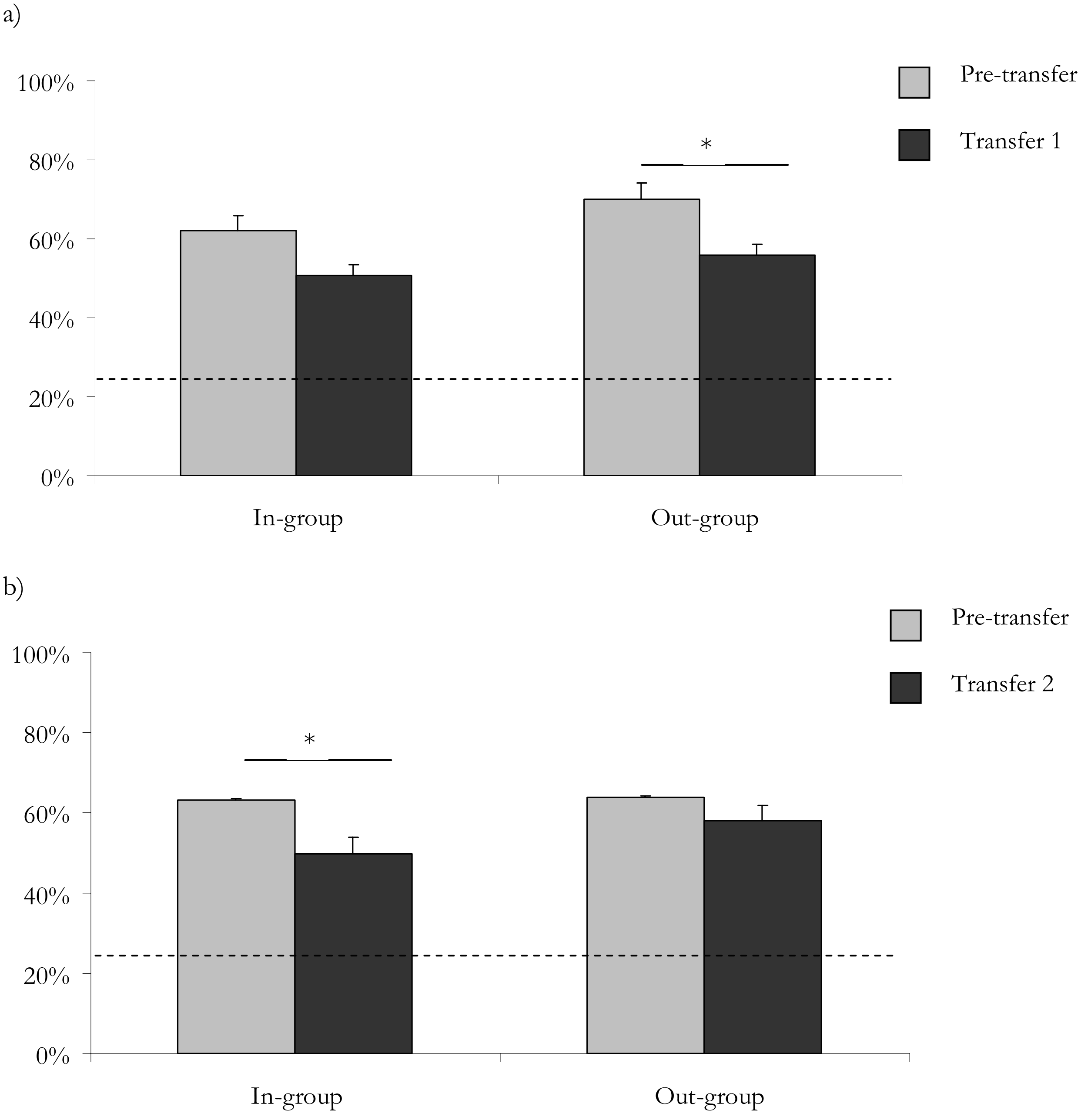
Observational studies do provide some evidence for triadic awareness but these studies are unable to rule out alternative explanations and the researcher has no control over the situation. Complementary findings from controlled experiments would make it more convincing, but very few studies have been conducted, hence our motivation for the current study. Deaner and colleagues (2005) found that rhesus subjects preferred to view images of high-ranking familiar conspecifics, so we can infer that subjects knew rank relations, but it is possible that they only knew the rank of the individual depicted in relation to themselves, not in relation to one another.

Preliminary studies

We previously conducted a study investigating conspecific face recognition in capuchin monkeys (Pokorny & de Waal, in press). The study presented an oddity task on touch-sensitive computer monitor. Four images that were the same or related and one that was different, or odd, were presented on the screen. Subjects needed to select the odd image to receive a reward. Training involved presenting three face images (portraits) that were identical and one that was from another monkey. Both individuals depicted on the screen were either from the same group (in-group) as the subject or from the neighboring social group (out-group). Once subjects were trained on this task and familiarized with conspecific portraits, they were presented with trials in which three stimuli were different portraits of the same individual, and a portrait of a different individual, again both were either members of the in-group or out-group relative to the subject. We separately analyzed performance on in-group and out-group trials. Subjects quickly attained above chance performance on this task, and as a group performed significantly above chance in the first 50 trials with both in-group and out-group faces.

A transfer test was conducted using a new set of images, assessing generalization to novel stimuli. Again, we analyzed the first 50 trials of both trial types and found that the group performed significantly above chance. Performance of the first 50 transfer trials was compared to the last 50 trials prior to transfer and while there was a decrease for out-group trials, performance remained significantly above chance (Figure 1a). Subjects continued demonstrating successful generalization when a final transfer test was presented that combined the two previous stimulus sets, so the combinations of images were novel. Transfer performance was compared to performance prior to transfer and in this case there was a decrease in the in-group trials, but performance was still above chance (Figure 1b). There were no accuracy differences for in-group or out-group faces, except in the face training stage when subjects performed better on out-group faces. This may be related to extra attention being given to unfamiliar individuals as this was the first time subjects experienced viewing images of conspecifics. This study provided convincing evidence that capuchin monkeys were able to recognize conspecific faces across viewpoint changes in two-dimensional images.

Figure 1. Comparing group performance on Individual Recognition task for a) transfer 1 and b) transfer 2 of previous Individual Recognition study. Light gray bars indicate 50 trials prior to transfer and the dark gray bars indicate the first 50 trials of transfer performance. An asterisk (*) indicates a significant difference (p < 0.05) between the comparison conditions. The horizontal dotted line designates chance (25%) performance.



We then investigated whether capuchins show specialized face processing. It is argued that humans have an area in the brain, the fusiform face area, specialized for discriminating faces (Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1996). Behavioral evidence for this specialization comes from inversion studies in which subjects are presented with a task involving upright and inverted (rotated 180º) faces. What is found is a general recognition impairment when faces, as opposed to other objects, are inverted, such that subjects' accuracy and response time decreases when inverted faces are presented compared to upright faces (for a review Valentine, 1988). The results of inversion studies in nonhuman primates have been mixed, with some finding positive evidence of the inversion effect (e.g. Overman & Doty, 1982; Parr, Dove, & Hopkins, 1998; Perrett et al., 1988; Swartz, 1983; Tomonaga, 1994) and others finding negative results (e.g. Bruce, 1982; Gothard, Erickson, & Amaral, 2004; Parr, Winslow, & Hopkins, 1999; Rosenfeld & van Hoesen, 1979).

Given our previous study demonstrating conspecific face recognition, we examined whether capuchins would demonstrate specialized face processing of conspecific faces by presenting them with inverted faces. The task was similar to the face training task in the earlier study in that three images were identical and one was different. Four different classes of stimuli were presented: capuchin faces, human faces, chimpanzee faces and automobiles. All images were converted to grayscale and the faces depicted neutral expressions and had forward facing head orientations and eye gaze. Half of the trials were presented with the stimuli upright and the other trials presented inverted stimuli. Results revealed that subjects performed better on upright than inverted capuchin and human faces, but there was no difference with chimpanzee faces and automobiles, suggesting that expertise plays a role in how faces are processed. This was the first study examining the inversion effect in capuchin monkeys and suggests that the face processing strategies of various primates, including humans, are remarkably similar. Both of the previous studies we conducted add to the body of knowledge concerning face processing in nonhuman primates and establish that capuchins are adept at recognizing faces, particularly of conspecifics.

Current study

The current study investigated whether a New World primate, *Cebus apella*, was able to use faces to categorize conspecifics based on group membership (in-group vs. out-group). Three adult subjects were tested using an oddity task, which presented four images simultaneously, three of which were the same or related, and one was different, or odd, this being the correct choice. The first experiment required subjects to select the odd image based on group membership using facial images of subjects' own social group and images of individuals in an out-group. After initial acquisition of the task, transfer tests were conducted to demonstrate that subjects could generalize to novel images. Further follow up tests were done to rule out alternative explanations for how subjects may have been able to perform well on the task, using information other than knowledge of in-group and out-group members.

Characteristics of the Oddity Paradigm

Before discussing the follow up tests that were conducted, first we will address why the oddity paradigm was used for this project. Several studies have used a match-to-sample (MTS) paradigm, not oddity. The basic layout of an identical MTS task is to present a sample image, then to present two alternative images: a match and a foil (or non-match). The subject is rewarded for selecting the image that is identical to the sample, the match.

For example, a sample image of a red circle would be presented, then two alternatives, a blue square and a red circle. The subject would select the red circle as the correct response. In order to determine that subjects have learned the concept of matching and are not just displaying traditional stimulus-response learning, new stimuli must be introduced to establish generalized matching-to-sample. If the subjects transfer to the new stimuli and perform with high accuracy, it is then concluded that they have acquired the matching concept.

An alternative to identical MTS is relational MTS, which can be used to determine relationships and concept identity. The layout is similar to identical MTS, except that the sample and the match are related in some fashion, but are not visually identical. For instance, the sample could be a red Papermate pen and the two choices would be a pencil and a blue Bic pen. The correct response would be the blue pen, indicating an understanding of the concept pens as opposed to pencils. This has been used to look at knowledge of social relationships, such as in the case of Dasser (1988) and Vonk (2002) who both looked at the relationship of mother-offspring.

The basic format of the oddity paradigm is different from the MTS paradigm in that it presents several items at once, of which some are identical or related, and one that is different, or odd; the objective being to select the odd object. The task requires the subject to determine the relationship between the stimuli in order to conclude which stimulus is "odd." Therefore, there is no constant quality (e.g. color or shape) of the stimuli that would allow a subject to correctly identify the odd stimulus; it depends on the dimension that applies to the majority of stimuli presented from which the choice stimulus deviates. For example, the objects presented could be a blue triangle, a red circle, a yellow triangle, and a green triangle. One would determine that the characteristic of similarity is that three are triangles, and would select the red circle as the correct response. In later trials, the blue triangle may be odd if it were presented with three different colored squares. This relational aspect of the oddity task allows one to present stimuli that are not identical visually, but instead have a more abstract dimension in common.

This task is most often used to demonstrate use of relative class concepts (Thomas, 1996). In relative class concepts, the features that determine an object's class membership are not inherent in the objects themselves. Membership can only be determined by comparing the object with other objects. For instance, which object is larger, which set of objects is less in number, etc. These concepts are relative because in the example of size, an object that is 5x5 cm may be selected as the large object when partnered with objects that are 3x3 cm and smaller. However, if the other objects were 7x7 cm and larger, the original "large" object no longer fits in the category of large, and would not be selected. Therefore, there is nothing inherent in the object itself that would indicate class membership. Instead it is the relationship between the objects that determines the current membership. In this study, the quality that grouped three of the individuals together was that they lived in the same social group, a group that was separate from the individual who was "odd."

This task has not specifically been used to explore knowledge of social relationships but has been used in other cognitive investigations with varying species (e.g. Adams-Curtis, 1990; Bernstein, 1961; Berryman, Cumming, Cohen, & Johnson, 1965; Buckley, Booth, Rolls, & Gaffan, 2001; Moon & Harlow, 1955; Noble & Thomas, 1970; Noble & Thomas, 1985; Roitberg & Franz, 2004; Strong, Drash, & Hedges, 1968; Strong & Hedges, 1966; Thomas & Boyd, 1973; Thomas & Frost, 1983; Wolin & Massopust, 1972; Wright & Delius, 1994; Wright & Delius, 2005). We selected the oddity task to investigate social relationships because the task itself requires the subject to compare the relationship of the items presented to determine the odd item. Thus we sought to use a different behavioral paradigm which could provide a new way to extract how individuals understand social relationships, in this case, group membership. We also examined whether subjects were using the oddity concept to correctly select the odd individual, suggesting that they were using knowledge of the social relationships between individuals presented on the screen. Alternatively monkeys might use rules, such as select the in-group individual, or select the out-group individual, which may point to subjects using knowledge of their *own* relationship with the depicted individuals as opposed to using knowledge of the relationship among the individuals depicted. We found that subjects were likely doing the latter and not necessarily applying the oddity concept to the dimension of group membership.

Alternative explanations for the categorization performance of capuchin faces

As was mentioned previously, after subjects acquired the current task, a transfer test was conducted to determine whether subjects would generalize to new images. While the images used in transfer tests were novel, they depicted the same individuals that had been presented for approximately 14 months. At least six different images of each individual had been used in the prior tasks, so it is possible that the "novel" images were not much different than the images previously presented. This would allow subjects to perform well on the task using the identity of the individuals depicted during training and acquisition. Fortunately, there were individuals in each social group, specifically the juveniles, whose images had not been used in any previous task. Therefore, we presented trials consisting of pictures taken of the juveniles to determine whether successful performance was based 1) on the generalization of prior images or 2) on the knowledge of individuals.

An additional explanation is that color cues might be present in the images that allow subjects to categorize the images, regardless of identity. This could be the case if monkeys in one group on average have a different color than monkeys in another group. To control for this, we a) measured luminance and color distribution of images of the two groups to determine if systematic differences could be detected, and b) presented images in grayscale, using three different sets of stimuli, described in the methods. We hypothesized that if categorization was based on color differences, subjects should perform at chance levels when categorizing grayscale images. Alternatively, if categorization was based on the identity of the individual depicted, then presenting the images in grayscale should have little or no effect on their performance and subjects should perform significantly above chance. In addition, we presented subjects with their own image under the assumption that if categorization was based on color differences, they should categorize their own image as an in-group member.

Brown capuchin monkeys (Cebus apella)

Capuchins are a good subject for this project for several reasons. First, they are a New World primate species and the majority of studies reviewed above have been conducted with Old World species and apes. In fact, aside from our own work (Pokorny & de Waal, in press)only one other study has investigated the face processing abilities of brown capuchins (Dufour et al., 2006). Capuchins are also highly intelligent and are often compared to chimpanzees on various cognitive tasks (Visalberghi, 1997). The similarities with respect to tool- use, food sharing and vertebrate predation likely developed independently in the two species, as they are geographically and phylogenetically quite distant (Rose, 1997; Visalberghi & McGrew, 1997). In the wild, capuchins live in groups of 14-17 individuals, similar to how they are housed at the Yerkes National Primate Center., and regularly encounter neighboring groups of conspecifics (Defler, 1982; Spironello, 2001).

These encounters typically elicit behavior that is specific to interactions with out-group members and is not demonstrated when encountering in-group members, indicating that they can distinguish between in-group and out-group members. Lastly, in capuchins, males leave the natal group and join other groups (Fragaszy, Visalberghi, & Fedigan, 2004). In a member of the same genus, *Cebus capucinus*, transferring males are observed to join groups in which there are familiar males (Jack & Fedigan, 2004). This suggests that capuchins are able to individually recognize one another and know who is familiar even among out-group members.

General Methods

Subjects and Housing

Subjects were three adult brown capuchin monkeys from two separately housed groups of capuchin monkeys. Subjects were "Bias" a female aged 19 from Group 1, "Wilma" a female aged 9 from Group 2, and "Mason" a male aged 8 from Group 1. Ages of subjects are given for when training began. These three subjects all participated in the earlier study on face recognition (Pokorny & de Waal, in press). Training began in September 2006 at which time the groups consisted of 15 individuals each in Group 1 and Group 2. Training and testing lasted until February 2009, except for Mason who stopped testing in February 2008 due to obtaining alpha status. During this time (February 2008), three individuals from Group 2 were transferred to a new facility for social reasons, bringing the final total of individuals to 12.

The groups were housed in the same facility, with differently sized indoor/outdoor areas (25 m^2 and 31 m^2 respectively) that the monkeys have free access to unless inclement weather requires closing off the outdoor area. Monkey chow and water were available *ad* *libitum*, with supplemental food trays containing fresh fruits, vegetables, bread and protein solution provided in the late afternoon, after testing sessions were completed. Tests were conducted once a day between the hours of 12:00 and 17:00 hours, approximately 5 days per week.

Apparatus

Tests were conducted in a mobile test chamber that was positioned in front of the indoor animal enclosure. Subjects were removed from their group by allowing them to come up to the test chamber and enter while the rest of the group was moved to the back of the enclosure. The test chamber measured 155 x 64 x 58 cm and was divided into 2 separate sections using an opaque partition, which blocked visual contact between subjects, allowing for the simultaneous testing of two subjects. The front of the test chamber was made of clear Lexan® with multiple holes cut out so that an individual could stick a hand out to perform the task.

A touch-sensitive monitor (43 cm diameter Elo Entuitive Touchmonitor) connected to a desktop PC was placed upon a mobile cart positioned directly in front of the test chamber. The cart also held an automatic pellet dispenser that contained either Kix cereal or Bio-Serv Precision Pellets which were provided as a reward. Visual Basic 6.0 was used to control the display presentation, reward delivery, playback of appropriate sound response, and data collection.

Stimuli

Stimuli used as baseline were three-dimensional clip art images sized 300x300 dpi. The facial images used in training and testing were digital photographs taken of all individuals in both

Group 1 and Group 2. These served as both the in-group and out-group stimuli for each subject. In-group refers to individuals who live in the same social group as the subject, while out-group individuals are those from the other social group. The two groups do not have visual access to one another, though four of the older adult females (including one subject, Bias) were housed together over 14 years prior. The terms in-group and out-group are used to denote the current living situation of subjects, whether they had at one point in the past been familiar with one another.

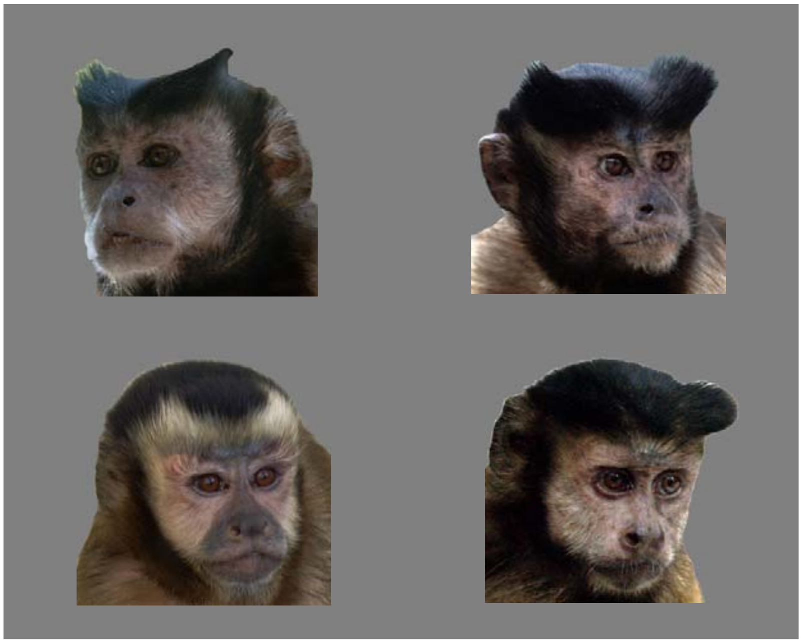
All photos were taken using a Konica Minolta Maxxum 7D digital camera and edited using Adobe Photoshop 6.0. Photos were cropped to only include the head, face and neck. Background information was removed by filling in the remaining area around the face with a solid gray color. Brightness and contrast were adjusted to control for differences in lighting conditions. Images presented in grayscale were converted using Adobe Photoshop. All images were sized to 8.4 cm^2 with a resolution of 300 pixels per inch.

Procedure

The experiments were conducted using an oddity paradigm. Trials began with a colored square located at the center of the screen. Upon contact, the square disappeared and four images appeared simultaneously on the screen in either a diamond or square layout. Three of the images were identical or related, and one was different, or odd, this being the correct choice. The location of the odd stimulus was randomly chosen at the beginning of each trial. If subjects contacted the correct image, all images disappeared, a high tone was played, and a reward delivered automatically via the pellet dispenser. For incorrect selections, all images disappeared, a low tone was played, no food reward was delivered, and an additional penalty of four seconds was added to the inter-trial interval. Subjects were given 30 seconds to make their selection or the trial ended and was recorded as aborted. Aborted trials were not included in the data analysis. A correction procedure was employed for regular trials such that when an incorrect selection was made, the trial was repeated four times or until the subject selected the correct response, which ever occurred first. Only the first presentation was included in data analysis.

Typical trials consisted of presenting subjects with either a) three different in-group individuals and one out-group individual (Out-group Odd) or b) three different out-group individuals and one in-group individual (In-group Odd, Figure 2). The two trial types were presented in blocked sessions such that the Out-group Odd condition was presented for several consecutive days before switching to the In-group Odd condition for an equivalent number of days. We initially presented subjects with the two conditions within one session, but performance was less than expected (see Appendix A), so we used blocked sessions throughout the rest of training and testing.

Figure 2. Example of a single Group Membership trial. All individuals are adult females. Correct response is the individual depicted in the top right position, who is from Group 2 while other individuals are from Group 1.



Later tests used the same paradigm, but no correction procedure was employed on test trials. Test trials were presented as probes, only once or twice, and the subject was rewarded for any response, correct or incorrect. This was so we could assess performance without contamination of reinforcement history. Sessions consisted of 20 probe trials, 35 original/typical trials (original stimulus set, reinforced for only correct responses, correction procedure employed) and 20 clip art trials that assessed the motivation and attention of the subject. As subjects typically performed above 80% on clip art trials, if performance on those trials fell below 60% on a given session, all the data for that session would be removed. This never occurred during the current study and therefore, no data were removed.

Previous training and experience

All subjects were trained to perform the oddity task on the touch-sensitive monitor using clip art images and then images of conspecifics. The specific training procedure can be found in Pokorny & de Waal (in press). Briefly, six subjects began training on the oddity task with clip art images. One subject was dropped during this stage as he obtained alpha status and refused to continue testing. Once subjects were successfully performing oddity with clip art images, we presented them with images of conspecific faces. Three images were identical and one was of a different individual. After reaching criterion (60% correct on two consecutive sessions), subjects began the Individual Discrimination task.

The Individual Recognition task presented trials in which three stimuli were portraits of the same individual but from different views, and the odd stimulus was a portrait of a different individual. Again, all individuals presented within a trial were either both from the in-group or both from the out-group in relation to the subject. Subjects were never presented with an image of themselves. Two transfer phases were conducted to assess generalization using 1) a novel set of stimuli, and 2) novel combinations of both sets of stimuli. Following completion of the Individual Recognition task, subjects began the current study. One subject was removed from the facility while performing the Individual Recognition task so was unable to participate in the current study. Another female subject never attained criterion on the Individual Recognition task after performing the task for 28 months and therefore was not included in the current study.

Data collection and analysis

All data collection was recorded via the computer controlling stimulus presentation. Information that was recorded on each trial included: subject, experimenter, date, type of test, trial condition (In-group Odd/ Out-group Odd), trial number, the names of the image files presented at each location, the location of the odd stimulus, the latency to start the trial as well as the latency to make a response (in milliseconds), the image and location that was selected by the subject, and whether the trial was correct, incorrect or aborted (no response from subject). Data were analyzed using SPSS 16.0.

The primary dependent variable of interest was the response (correct/incorrect) and the independent variable was the type of condition, In-group or Out-group. The latency, the time from starting a trial to selecting a response, was also used in analyses as a dependent variable. Trials in which latencies were less than 500ms or longer than 10,000ms were removed from analyses, as less than 500ms does not provide enough time for subjects to move to a target and latencies longer than 10,000ms usually indicate that the subject was distracted. This criterion resulted in the removal of a small number of trials, usually one trial per subject in each experiment. As the latencies were typically skewed even with those outlying trials removed, we followed a set procedure to determine whether to transform the data, starting with conducting a Kolmogorov-Smirnov (K-S) test ($\alpha = 0.05$) to determine if the distribution deviated from a normal distribution. If significant, latencies were transformed via a natural log transformation and again tested for deviation from a normal distribution with a K-S test. If the result was still significant, an inverse transformation was applied to the original latencies and again tested with the K-S test. In the results, medians (Mdn) and interquartile ranges (IQR) are reported untransformed, but the results of any other analyses are from the transformed data, if transformations were necessary. McNemar tests were conducted to compare performance on a task prior to transfer and after transfer as well as to assess differences between In-group and Out-group trial types. These analyses compared the last 40 trials prior to transfer (“pre-transfer”) to the first 40 trials of transfer

(“transfer”) for each subject. For probe sessions, analyses compared the 40 probe trials to 40 original trials performed within the same session, analyzing In-group and Out-group separately. To evaluate performance above chance level (25%), Heterogeneity G-tests were conducted. Heterogeneity G-tests compare performance with random chance, similar to a chi-square, but the G-test takes into account individual contributions.

Experiment 1: Group Membership task

Procedure

Group membership trials consisted of a) three different in-group individuals and one out-group individual (Out-group Odd), and b) three out-group individuals and one in-group individual (In-group Odd). Within one session, subjects were presented with 25 In-group Odd trials, 25 Out-group Odd trials and 25 3-dimensional clip art trials. Performance was less than expected, so sessions were changed to only present one condition at a time for several sessions before switching to the other condition. For example, the In-group Odd condition was performed for several days, after which the Out-group Odd condition was performed for an equivalent number of days. Blocked sessions consisted of 40 group membership trials (In-group Odd or Out-group Odd) and 35 clip art trials used as baseline.

To assess generalization, subjects were transferred to a new set of stimuli. After subjects attained a performance criterion of 60% correct (compared to 25% chance level) for at least two consecutive sessions on the given condition type, subjects were transferred to a new stimulus set. The transfer was performed in the same manner as the training sessions, such that there were 40 group membership trials and 35 clip art trials. Analyses compared the last 40 trials of the task using the original set of images with the first 40 trials of transfer using the new stimulus set.

Stimuli

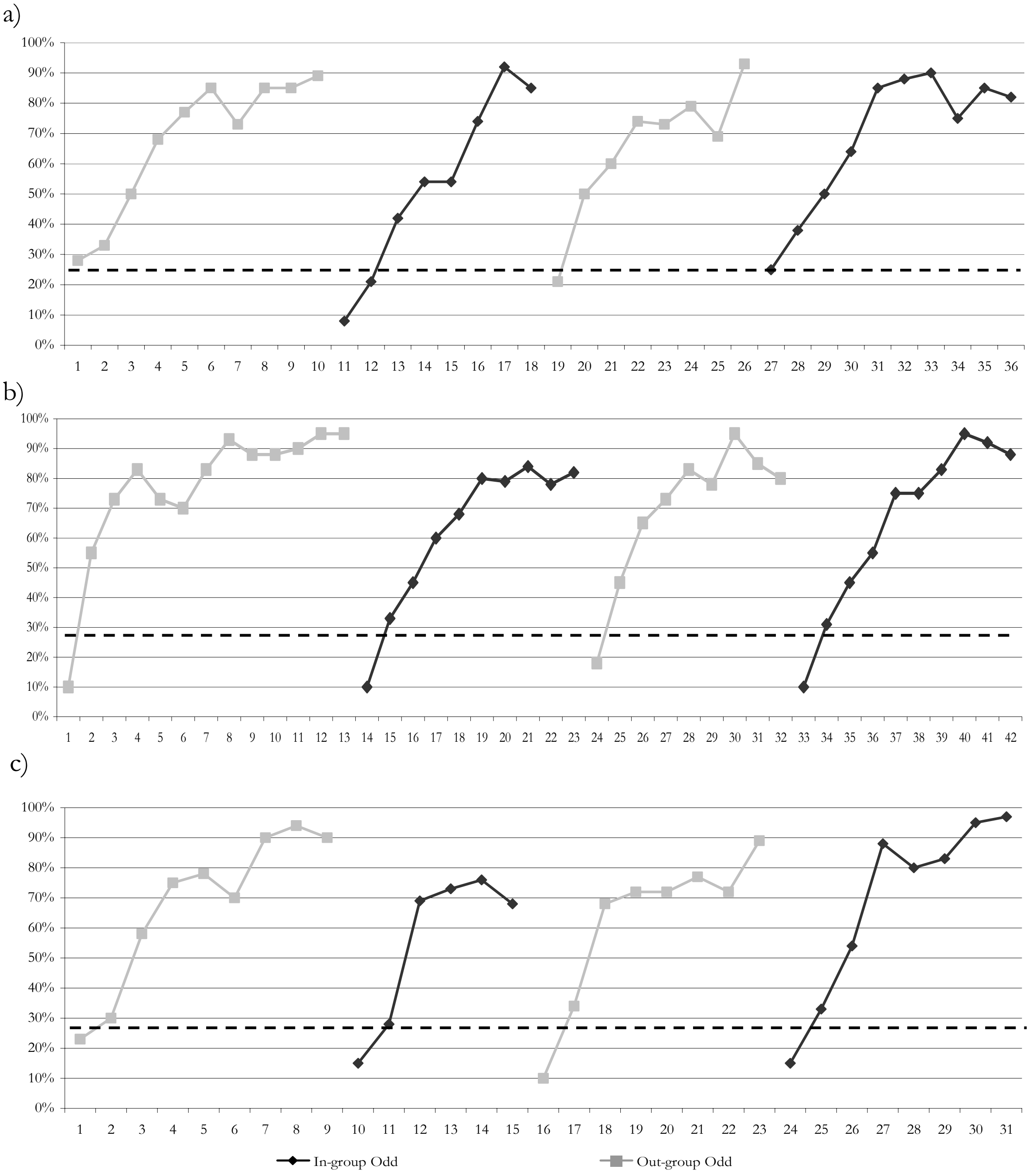
The initial stimulus set consisted of six different views of nine individuals from Group 1 and six views of eight individuals and three views from two individuals from Group 2 for a total of 108 pictures. The transfer stimulus set contained three views of nine individuals from Group 1 and ten individuals from Group 2. The initial stimulus set is also the set that was used for typical trials in each of the following experiments. In February 2008 when three individuals were removed from Group 2, the images of those individuals were also removed from the stimulus set.

Results Experiment 1

This task presented two conditions: 1) In-group Odd, three individuals were from the out-group and one individual was from the in-group relative to the subject, and 2) Out-group Odd, three individuals were from the in-group and one was from the out-group relative to the subject. Initially, these two conditions were both randomly presented, along with clip art trials, within a session. This approach was not very successful (Appendix A), so we separated the conditions and only presented one within a session.

Given the less than expected performance, group membership conditions were presented in blocks across sessions. For instance, In-group Odd and clip art trials were presented for several days, and then Out-group Odd and clip art trials were presented for an equivalent number of days. All subjects were presented with the In-group Odd condition first. Performance increased dramatically within the first two sessions (Bias: M = 51.9%, SEM = 5.6%; Mason: M = 65.0%, SEM = 5.4%; Wilma: M = 65.8%, SEM = 5.5%) and was higher than performance on mixed sessions (Table 30 in Appendix A). Subjects continued to perform the In-group Odd condition for several sessions before switching to the Out-group Odd condition. Upon switching to the other condition, we found that subjects performed poorly for a few sessions before returning to previous performance levels. This phenomenon was consistent across subjects and occurred every time subjects switched from one condition to the other (Figure 3).

Figure 3. Individual performance on In-group and Out-group conditions when switching from one condition to the other. Out-group Odd is lighter squares (1st condition represented), and In-group Odd is dark diamonds (2nd condition represented). Horizontal dotted lines indicate chance performance (25%). Panels refer to each subject: a) Bias, b) Mason, c) Wilma.



To assess generalization of the in-group / out-group discrimination, a new set of images was prepared and presented. Subjects were required to obtain a performance criterion of 60% correct for two consecutive sessions on the given condition before transferring to the new set of stimuli. One session consisting of 40 transfer trials was completed for each condition. As a group, subjects performed above 25% chance upon transfer under both conditions (In-group Odd: Gh = NS, Gp = 81.24, *df* = 1, $p < 0.001$; Out-group Odd: Gh = 7.05, *df* = 2, $p < 0.05$, Gp = 59.14, *df* = 1, $p < 0.001$). Individual performance can be found in Table 1. The significant heterogeneity in the Out-group Odd condition is due to Bias's poorer performance compared to Mason's and Wilma's, though it was significantly above chance (z = 1.67, $p = 0.048$, binomial test).

Table 1: Individual's correct performance on group membership transfer trials, 40 trials each condition

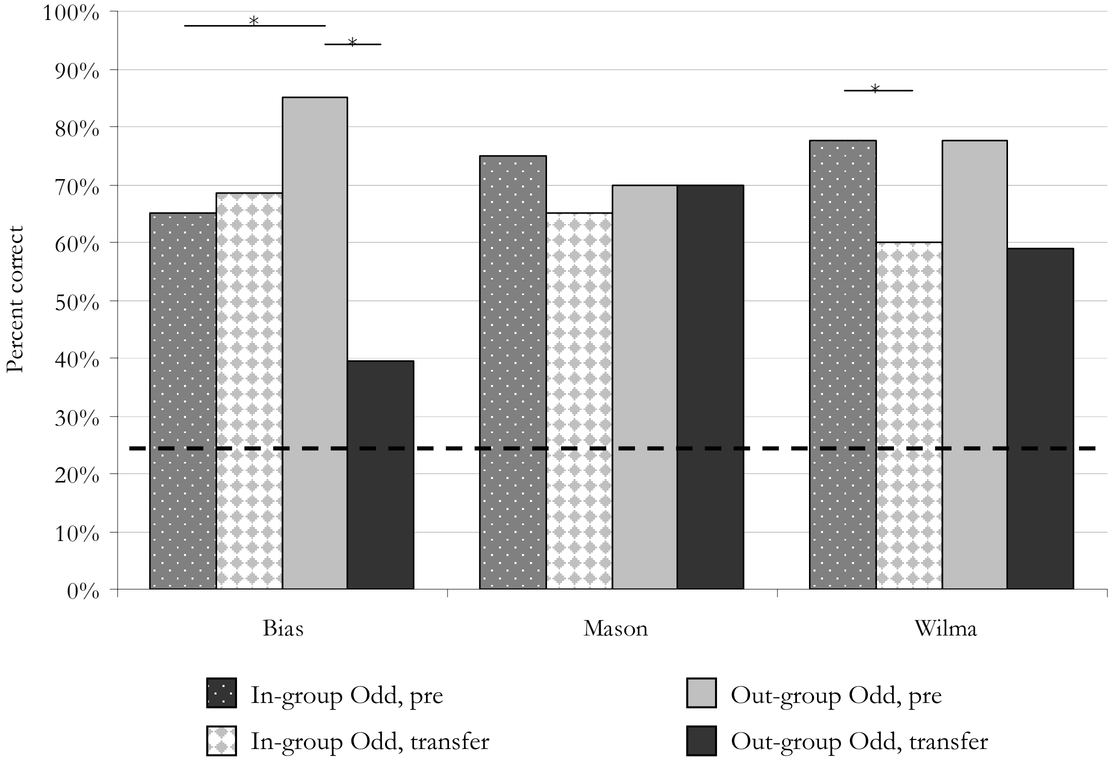
	In-group Odd	Out-group Odd
Subject	*Mean*	*Mean*
Bias	68.4%	39.4%
Mason	65.0%	75.0%
Wilma	60.0%	58.9%

Transfer performance was also assessed by comparing the last 40 trials prior to transfer with the first 40 trials of transfer for each condition. This assessed whether transfer performance, while significantly above chance, was at the same level as pre-transfer performance. Analyses revealed that Mason did not significantly differ on transfer trials from prior to transfer on both conditions (In-group Odd: McNemar test, $N = 40$, $p = 0.454$;

Out-group Odd: McNemar test, $N = 40$, $p = 1.000$). Bias performed worse after transfer on Out-group Odd trials (McNemar test, $N = 33$, $p < 0.001$; pre-transfer: $M = 85.0\%$; transfer $M = 39.4\%$) but there was no difference between pre-transfer and transfer performance on In-group Odd trials (McNemar test, $N = 38$, $p = 1.000$). The opposite was found for Wilma, who performed worse in the In-group Odd condition transfer trials (McNemar test, $N = 40$, $p = 0.049$; pre-transfer: $M = 82.5\%$; transfer $M = 60.0\%$) but no different in the Out-group Odd condition (McNemar test, $N = 39$, $p = 0.143$).

We conducted individual McNemar tests to determine whether there was a difference in performance between the two conditions, In-group Odd and Out-group Odd, suggesting an ease to select one over the other. Results showed that no subject performed differently between conditions prior to transfer (Bias: McNemar test, $N = 40$, $p = 0.057$; Mason: McNemar test, $N = 40$, $p = 0.791$; Wilma: McNemar test, $N = 40$, $p = 0.791$). This was also assessed for the transfer tests (Bias: McNemar test, $N = 33$, $p = 0.049$; Mason: McNemar test, $N = 40$, $p = 0.815$; Wilma: McNemar test, $N = 39$, $p = 1.000$), and in this case, Bias did show a difference between conditions, performing better on the In-group Odd condition (In-group Odd: $M = 68.4\%$; Out-group Odd: $M = 39.4\%$), though her performance was significantly above chance in the Out-group Odd transfer condition ($z = 1.67$, $p = 0.048$). Results are summarized in Figure 4.

*Figure 4. Individual's performance correct on blocked Group Membership trials, before and after transfer to new images. Pre = 40 trials immediately before transfer session. Transfer = first 40 trials when new stimuli were presented to the subject. Individual subject is on the X-axis and the percentage correct on the Y-axis. Dotted line indicates the chance level (25%). Subjects all performed significantly above chance. A * indicates whether the performance on that condition was significantly (p < 0.05) different to the comparison condition according to a McNemar test.*



The latency to make a response was analyzed to determine if there was a difference depending on the condition type, with faster response times pointing to faster cognitive processing. Prior to transfer, Bias and Wilma responded significantly faster (Bias: *t*(38) = -3.529, *p* = 0.001, two-tailed; Wilma: *t*(38) = -3.361; *p* = 0.002, two-tailed) in the In-group Odd condition (Bias: *Mdn* = 2444.00, *IQR* = 1101.00; Wilma: *Mdn* = 1592.50, *IQR* = 380.25) than in the Out-group Odd condition (Bias: *Mdn* =3175.00, *IQR* = 1372.00; Wilma: *Mdn* =

1863.00, *IQR* = 591.00), while Mason did not show a significant difference (In-group Odd: *Mdn* = 1537.00, *IQR* = 550.25; Out-group Odd: *Mdn* = 1672.50, *IQR* = 958.25; $t(39) = -1.304$, $p = 0.200$). In the transfer session, Wilma was faster on In-group Odd trials (In-group Odd: *Mdn* = 1483.00, *IQR* = 451.00; Out-group Odd: *Mdn* = 1893.00, *IQR* = 701.00; $t(38) = -4.014$, $p < 0.001$) consistent with her performance prior to transfer. Neither Bias ($t(38) = 0.155$, $p = 0.878$) nor Mason ($t(39) = 1.993$, $p = 0.053$) responded differently to In-group (Bias: *Mdn* = 3605.00, *IQR* = 1932.50; Mason: *Mdn* = 1932.50, *IQR* = 1024.50) or Out-group (Bias: *Mdn* = 3665.00, *IQR* = 2244.00; Mason: *Mdn* = 1722.00, *IQR* = 753.50) trials in the transfer session.

Discussion Experiment 1

The group membership oddity task presented three different individuals from one group and one individual from the other group. This task required subjects to discriminate the faces of similar looking conspecifics who differed in their relationship to the subject and each other. The initial presentation of this task had conditions mixed within a session and did indicate that subjects could select individuals based on group membership, as performance was significantly above chance over all sessions. However, performance was relatively low compared to tasks that subjects have performed in the past (Pokorny & de Waal, in press) and was not showing improvement (Appendix A). Therefore, we separated conditions and presented them in blocks across sessions. Once sessions were separated by condition, performance improved dramatically within the first two sessions, providing convincing evidence that subjects could select the in-group and out-group individuals depicted in the images. As for why subjects could not successfully perform the group membership oddity task with mixed sessions is unclear. If they were applying the oddity concept, it seems that they should be able to perform the task. It seems unlikely that they did not have an understanding of in-group and out-group as performance rose quickly for all subjects once the two conditions were separated. One possibility is that they were not using the oddity concept and instead were applying a rule, such as "select the in-group member, avoid the out-group member." This possibility will be address later in Experiment 5.

The finding that subjects required a few sessions before performing above chance when switching between condition types is similar to what is found in reversal learning tasks. These tasks typically take the form of a stimulus discrimination task (A+, B-) and after subjects reach a given criteria, the contingencies are changed such that the stimulus that was previously positive is now negative and vice versa (A-, B+). Rumbaugh (1970) developed a transfer index, which is a way to measure reversal learning and compare the value with other species. Essentially it measures how long it takes an individual to learn to respond correctly to the reversal. If subjects learn quickly, then they will have a higher transfer index than individuals who take longer to learn the reversal and perseverate on the previously learned association. It appears that there is an increase in the transfer index that corresponds phylogenetically when looking across primate species, increasing as ones moves from prosimians to the great apes, with capuchins falling in the middle (Rumbaugh & Pate, 1984). The mediational paradigm (Rumbaugh, 1971) allows one to assess what strategy, associative or rule based, is utilized by subjects when presented with reversal learning tasks. Associative learning is based on stimulus-response associations, namely excitatory and inhibitory associations, whereas rule-based is more cognitively advanced in that one can learn from the first reversal and immediately solve the task and all following problems using the same strategy of win-stay/lose-shift. Again, it appears that there are qualitative differences in learning strategies across the primates, such that apes use more rule based strategies while monkeys employ associative strategies (Rumbaugh, 1971).

The mediational paradigm allows one to assess whether subjects learned to approach A+ or if they learned to avoid B-, as these result in different behaviors with novel stimuli and associations in the reversal stage (e.g. B+C-, A-C+). Capuchins have been tested on the mediational paradigm and the results point to more associative than rule based learning (Beran et al., 2008; De Lillo & Visalberghi, 1994). We essentially only presented A and B to our subjects so it is impossible for us to distinguish which associative strategy was being used. For example, if subjects are performing the In-group Odd condition, they may be learning to approach the in-group or to avoid the out-group. When the condition switches to Out-group Odd, subjects must either overcome the inhibition to select the out-group individuals or overcome approaching the in-group members – both appear as an inability to switch immediately to the new condition type which we observed. Beran and colleagues (2008) explicitly tested capuchins on the mediation paradigm and found that subjects mainly learned to avoid the B stimulus during training, so in the example presented, our subjects may have been learning to avoid out-group individuals and needed to overcome avoidance before performance increased on Out-group Odd trials. As was stated, we only tested condition A and B and therefore cannot be sure that our subjects were learning to avoid the S-, but this is a possible explanation for what could have been occurring given our observations of their performance when switching conditions.

A transfer session was introduced to assess whether subjects' could generalize their behavior to a new set of stimuli. Subjects did perform significantly above chance on transfer trials in both conditions, indicating positive transfer. However, one can also assess transfer by measuring whether transfer performance was the same as prior performance and in this case, only Mason performed similarly during transfer as he did prior to transfer in both conditions. Bias did maintain pre-transfer performance in the In-group Odd condition and Wilma did in the Out-group Odd condition, but they did not in the other respective condition.

As for whether subjects responded faster when selecting the in-group or out-group member, again there were some discrepancies among subjects. Prior to transfer, Bias and Wilma selected the in-group individual faster than they selected the out-group individual while Mason responded the same to both. In the transfer sessions, again Wilma selected in-group members faster but there was no difference for Mason or Bias. Therefore, little can be concluded from any latency analyses given the variability among subjects.

We conclude that overall subjects did generalize to new stimuli, though in some instances there was a drop in performance, but not below chance levels. This indicates that performance was not specific to the stimulus set that was used in training. It also suggests that they recognized the individuals depicted in the images, at least the familiar individuals, and used that knowledge and the relationship with one another, or with the subject, to complete the task.

Methods Experiment 2: Alternatives for group membership categorization – Generalize the identity of the depicted individual

While the portrait itself used in the previous transfer task was novel, the individuals depicted in the portraits were not. It is possible that positive performance was due to generalizing the specific identity of the individuals depicted. While subjects did not know the out-group members, they had been exposed to pictures of those individuals for many months, enough time for those individuals to no longer be "unfamiliar" and for the subject to select them based on identity. If this were the case, then subjects should not generalize performance if presented with novel conspecifics as no association has been created through prior training. Alternatively, if subjects used the identity of the individuals depicted in the portraits, regardless of prior experience with the image, subjects should be able to successfully transfer to portraits of novel individuals never before depicted in a two-dimensional image. Therefore, we presented images of juveniles from both social groups that had never before been presented in an image. As the identity of the individuals was completely novel, positive performance could not be explained by any prior experience in the task and could only be explained by subjects identifying the in-group individuals and categorizing them appropriately.

Procedure

Probe trials were presented as subjects performed the group membership task, first performing the In-group Odd condition and then the Out-group Odd condition.
Probe trials included only portraits of juveniles. The juvenile stimulus set consisted of four photos of eight juveniles, four from each group, for a total of 32 images. These juveniles had never before been shown as digital images to subjects, so were completely novel. Subjects first performed several sessions with the original portrait set before being presented with two test sessions that included the juvenile probe trials. Twenty probe trials were presented in each test session, resulting in 40 probe trials. This procedure was conducted for both the In-group Odd condition and the Out-group Odd condition. Performance on these 40 trials was compared to 1) 25% chance level using a heterogeneity G-test, and 2) performance on the typical trials performed in the same session. This allowed us to

determine if performance on these early trials was significantly above chance and if performance significantly differed from how subjects performed on the typical trials.

Specific predictions

1 – Previous association: If performance on the previous transfer had been based on the fact that the portraits depict the same individuals used for training and thus that subjects categorized "novel" images based on learned associations, then subjects would perform in the present test at chance levels. This would be because the present test used portraits of juveniles, which had never before been depicted in a two-dimensional image.

2 – In-group/out-group distinction: Alternatively, if subjects were categorizing based on the identity of the individual depicted and the group in which it lived, then subjects would perform significantly above chance when categorizing images of juveniles. The conclusion would be that group membership, not past associations, determined the choice.

Results Experiment 2

Subjects maintained high performance on typical trials (trials using the original stimulus set), performing significantly above chance (In-group Odd: Gh=NS, Gp=159.88, df=1, $p < 0.001$; Out-group Odd: Gh=NS, Gp=180.63, df=1, $p < 0.001$), indicating that inclusion of the probe trials did not interfere with their normal performance. As for performance on the probe trials, subjects again performed significantly above chance in both conditions (In-group Odd: Gh=NS, Gp=46.34, df=1, $p < 0.001$; Out-group Odd: Gh=8.63, df=2, $p = 0.013$, Gp=68.76, df=1, $p < 0.001$). The significant heterogeneity in the Out-group Odd condition was due to Mason's performance ($M = 42.5\%$) being considerably lower than Bias'

(*M* = 67.5%) and Wilma's (*M* = 72.5%), even though when tested individually, Mason did perform significantly above chance (z = 2.257, *p* = 0.012, binomial test).

Individual generalization was assessed by comparing performance on the typical trials with the probe trials performed within the same session. Results revealed that while all subjects performed significantly above chance on probe trials, there was considerable variability in their performance compared to typical trials (Table 2).

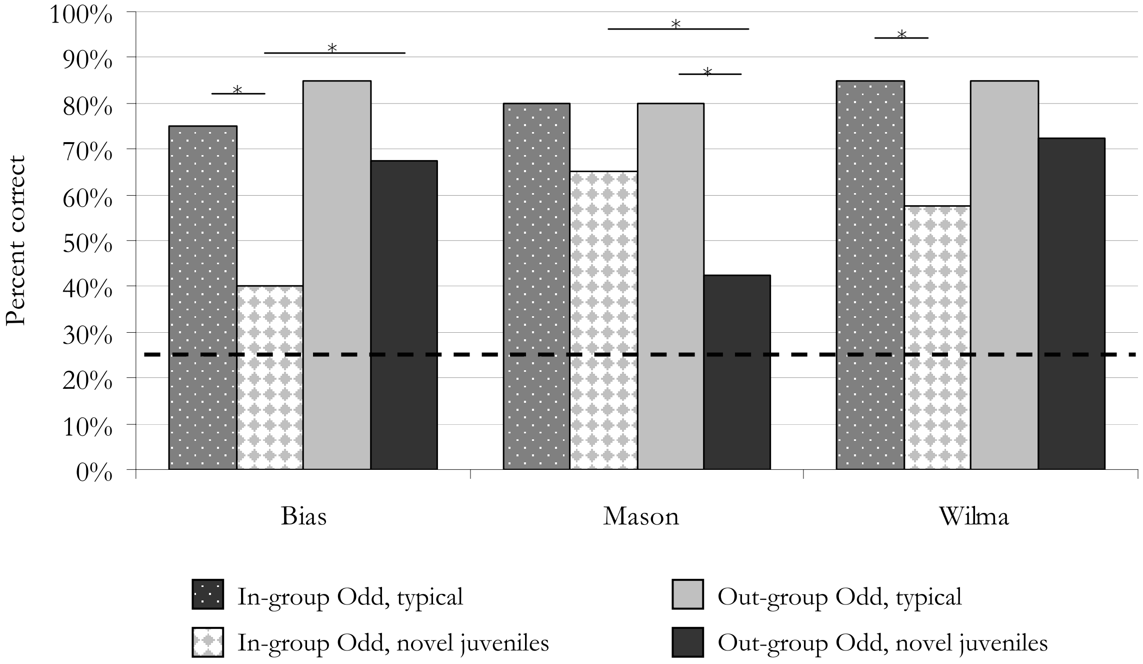
Table 2. Comparison of typical and juvenile probe trials by subject for In-group Odd and Out-group odd conditions.

		Typical	Probe	McNemar Test	
Subject	Condition	*Mean*	*Mean*	*N*	*p*
Bias	In-group Odd	75.0%	40.0%	40	0.007*
	Out-group Odd	85.0%	67.5%	40	0.118
Mason	In-group Odd	80.0%	65.0%	40	0.180
	Out-group Odd	80.0%	42.5%	40	0.001*
Wilma	In-group Odd	85.0%	57.5%	40	0.013*
	Out-group Odd	85.0%	72.5%	40	0.302

We also examined whether there were performance differences between the two condition types (In-group Odd and Out-group Odd) for typical trials or probe trials. For the typical trials, analyses found no difference between conditions for typical trials with the original stimulus set (Bias: McNemar test, *N* = 40, *p* = 0.388; Mason: McNemar test, *N* = 40, *p* = 1.000; Wilma: McNemar test, *N* = 40, *p* = 1.000). For probe trials, Bias performed better on Out-group Odd (*M* = 67.5%) than In-group Odd trial types (*M* = 40.0%;

McNemar test, $N = 40$, $p = 0.027$), while Mason showed the opposite result (McNemar test, $N = 40$, $p = 0.049$). Wilma (McNemar test, $N = 40$, $p = 0.286$) showed no difference. Results are summarized in Figure 5.

*Figure 5. Individual's performance on Group Membership trials with original stimuli and on probe trials with novel images of juveniles in both the Out-group Odd and In-group Odd conditions. Novel stimuli = 40 probe trials, Original stimuli = 40 trials during same sessions in which probe trials were presented. Subject is on the X-axis and the percentage correct is along the Y-axis. The dotted line indicates chance level (25%). All subjects performed significantly above chance in all trial types. A * indicates whether performance on the condition was significantly ($p < 0.05$) different than the comparison condition according to a McNemar test.*



Latencies were also examined using a paired samples *t*-test to determine whether there were response time differences between the typical and juvenile probe trials in each condition within the first 40 trials (Table 3). Bias performed faster on typical trials in the In-group Odd condition, while Mason and Wilma responded faster on typical trials in both conditions. We also assessed response differences when selecting the in-group or out-group member on typical trials and probe trials (Table 4). Results revealed that Mason and Wilma were faster to select the in-group member on typical trials, but no one responded differently to selecting the in-group or out-group on probe trials.

Table 3. Comparison of latencies to respond to typical (N = 40) and juvenile probe trials (N = 40) in each condition. Medians and IQRs are reported untransformed while t values may be from transformed data (see Methods).

		Typical		Probe			
Subject	Condition	*Mdn*	*IQR*	*Mdn*	*IQR*	*t*	*p (two-tailed)*
Bias	In-group Odd	2514.00	1573.00	3455.00	2294.00	-2.588	0.014*
	Out-group Odd	2354.00	1732.00	2729.00	7691.00	-1.555	0.128
Mason	In-group Odd	1407.00	736.75	2333.00	1915.25	-5.355	< 0.001*
	Out-group Odd	1623.00	911.00	2298.00	1287.00	-3.010	0.005*
Wilma	In-group Odd	1322.00	448.50	1797.50	944.50	-4.657	< 0.001*
	Out-group Odd	1522.50	582.00	1872.50	786.00	-2.065	0.046*

Table 4. Comparison of the latency to respond by subject in the In-group Odd and Out-group Odd condition for typical trials (N = 40) and juvenile probe (N = 40) trials.

		In-group Odd		Out-group Odd			
Subject	Condition	*Mdn*	*IQR*	*Mdn*	*IQR*	*t*	*p (two-tailed)*
Bias	Typical	2514.00	1573.00	2354.00	1732.00	0.694	0.492
	Probe	3455.00	2294.00	2729.00	1692.50	1.539	0.132
Mason	Typical	1407.00	736.75	1623.00	911.00	-2.607	0.013*
	Probe	2333.00	1915.25	2298.00	1287.00	0.501	0.619
Wilma	Typical	1322.00	448.50	1522.50	582.00	-3.172	0.003*
	Probe	1797.50	944.50	1872.50	786.00	-0.335	0.739

Discussion Experiment 2

This experiment took advantage of the fact that not all individuals present in our two groups had been depicted in previous tasks. It was possible that the positive transfer performance in Experiment 1 was due to the fact that all individuals depicted in the transfer set were used previously in the training set. Therefore, transfer may have been based on previously learned associations of the specific individuals. By presenting novel individuals, we eliminated the possibility of any learned associations and investigated whether subjects could generalize to novel individuals. In general, subjects did positively generalize to images of juveniles as performance was significantly above chance in all conditions. This is what would be expected if group membership is a natural, open-ended category, in which one responds in the same manner to all items in a group, and continues to do so for newly encountered items as well.

As for whether performance was the same on typical trials as on probe trials, the results were not consistent across subjects. Bias and Wilma performed poorer in the In-group Odd condition, which may suggest that the novelty of the images interfered as the In-group Odd condition was performed first. On the other hand, Mason also performed the In-group Odd condition first and he didn't demonstrate any performance difference between typical and probe trials. He did, however, perform worse in the Out-group Odd condition. Again, these findings are not consistent across subjects and cannot be explained by testing condition order or by what set of individuals they were to be selecting (which would be the case if Mason and Bias had the same pattern and Wilma had the direct opposite). It did appear that introducing novel stimuli affected subjects' response times, as all subjects took longer to respond when images of juveniles were presented, except for Bias in the Out-group Odd condition. This pattern is not surprising as subjects were quite familiar with the typical stimulus set so when novel images were presented, more time was required to scan the portraits before making a response.

In the previous experiment, Mason responded faster in the transfer session when selecting out-group members, possibly because of the number of females represented. Here we did not see any difference in the latency to select the in-group or out-group among the juvenile probes. Juvenile capuchins lack the secondary sexual characteristics that are seen in the adults, and so these portraits may not have affected his attention level as did the adult images. In fact, the only latency difference between responding to in-group or out-group members was on typical trials, not on probe trials. Mason and Wilma both responded faster when selecting the in-group. Wilma's response was consistent with the previous experiment, but Mason's was not.

In general, as performance was significantly above chance for subjects in both probe conditions, we do conclude that subjects recognized the individuals depicted in the images and were able to categorize them as an in-group or out-group member. Since subjects had no prior association with the juvenile pictures on this task, subjects needed to apply their knowledge of individual members in their group to perform this task and did so successfully.

Methods Experiment 3: Alternatives for group member categorization – Color cues - grayscale

The above experiment ruled out one possible explanation for how subjects may be able to transfer to novel stimuli, namely by generalizing the identity of individuals presented before. An alternative explanation is that color cues were present in the portraits that provide information as to how the images should be categorized. To eliminate possible color cues, subjects were presented with portraits converted to grayscale.

Procedure

Probe trials consisted of portraits of the adults and subadults, but the images were converted into grayscale. Again, subjects first performed the In-group Odd condition, followed by two test sessions, and then performed the Out-group Odd condition followed by two test sessions. Three test sessions were performed, using three different stimulus sets. One stimulus set was the original stimulus set used for typical trials in all experiments but now in grayscale. The second stimulus set was the set of images that were used in the transfer test. The luminance of the images was controlled between Group 1 and Group 2 for both of these sets and the stimuli were presented in grayscale. To control luminance differences, we selected only the image by using the Select function in Adobe Photoshop 6.0. We then selected “image” and “invert” so that only the image was selected, not the gray background. The mean luminance value was obtained from the Histogram feature. This value was to be kept within a range of 50 to 110, the majority being around 75. If the image deviated from this, we manipulated the brightness and contrast of the image until the luminance fell within the predefined range. The third set of images was a novel set of images converted to grayscale, consisting of 48 images (three views of ten individuals from Group 1 and six individuals from Group 2). The luminance of the images in this set was modified such that the average luminance of Group 1, which includes several blond (light-colored) individuals, was now significantly darker in luminance than the images of Group 2. This was to determine whether luminance differences influenced their selections. For example, on typical trials if they were selecting individuals from Group 1 based on a lighter color, then in the probe trials they would have to select the individual that was darker. One subject, Mason, obtained alpha status before being tested on the final grayscale stimulus set and refused to test. His results are only reported for the first two grayscale sets.

These test sessions consisted of 35 typical trials (color, original set), 20 probe trials (grayscale), and 20 clip art images. Performance on the first 40 probe trials was compared to 1) chance level, and 2) the performance on the typical trials performed in the same session. This allowed us to determine if performance on these early trials was significantly above chance and if performance differed from how subjects performed on the typical trials. Performance on the second grayscale stimulus set was also compared to subjects’ prior performance when they viewed the images in color to see if removing the color was detrimental to their performance.

Specific predictions

If prior transfer performance (Experiment 1) was based on systematic color differences between members of the two groups of capuchins, then subjects should perform at 25% chance levels when presented with grayscale images. Alternatively, if subjects were categorizing based on the identity and group membership of the individual depicted, then subjects should perform significantly above chance even with grayscale stimuli and with grayscale stimuli that reversed the natural luminance differences between the two groups.

Results Experiment 3

a) Analysis of color or luminance differences

We measured the mean and median levels of red, green, blue and luminance in each of the previous colored stimulus sets and conducted independent samples t-tests to assess differences between the two groups for each stimulus set. To measure the color channel levels, we selected only the face of the image and recorded the mean and median of each color channel provided by the Histogram feature in Adobe Photoshop. Results can be found in Tables 5 – 7. There were significant differences in the red and blue levels in the original set (Table 5) and in the red, green, blue and luminance levels in the novel juvenile set (Table 7). Luminance differences were not found in the original set or in the transfer set. The transfer set was relatively well controlled for color, having only a significant difference in the median level of red between Group 1 and Group 2 (Table 6).

Table 5. Independent samples t-tests for the mean and median levels of red, green, blue and the luminance from Group 1 (N = 54) and Group 2 (N = 54) used in the initial Group Membership stimulus set.

	Group 1		Group 2			
Measure	*Mean*	*SEM*	*Mean*	*SEM*	*t*	*p (two-tailed)*
Mean red	91.81	1.578	82.63	1.640	4.036	<0.001 *
Median red	89.37	1.848	77.11	1.996	4.507	<0.001 *
Mean green	75.57	1.261	75.34	1.676	0.112	0.911
Median green	70.91	1.431	69.70	2.035	0.484	0.630
Mean blue	60.58	1.348	69.80	1.800	-4.103	<0.001 *
Median blue	53.93	1.507	62.59	2.205	-3.245	0.002 *
Mean luminance	78.79	1.305	76.84	1.652	0.930	0.354
Median luminance	74.52	1.485	71.19	2.020	1.330	0.187

Table 6. Independent samples t-test for the mean and median levels of red, green, blue and the luminance of photos from Group 1 (N = 30) and Group 2 (N = 27) used in the transfer test (Experiment 1).

	Group 1		Group 2			
Measure	*Mean*	*SEM*	*Mean*	*SEM*	*t*	*p (two-tailed)*
Mean red	85.39	2.373	79.72	2.095	1.776	0.081
Median red	83.80	2.870	73.82	3.215	2.325	0.024 *
Mean green	73.91	1.791	70.50	2.114	1.237	0.222
Median green	71.40	2.316	65.00	2.958	1.721	0.091
Mean blue	66.69	1.623	65.07	2.267	0.588	0.559
Median blue	63.37	2.060	58.37	3.104	1.366	0.177
Mean luminance	76.52	1.910	72.64	2.099	1.373	0.175
Median luminance	74.43	2.415	67.07	2.983	1.934	0.058

Table 7. Independent samples t-test for the mean and medial levels of red, green, blue and the luminance of photos of the juveniles from Group 1 (N = 16) and Group 2 (N = 16) used in Experiment 2.

	Group 1		Group 2			
Measure	*Mean*	*SEM*	*Mean*	*SEM*	*t*	*p (two-tailed)*
Mean red	103.26	2.090	85.33	2.005	6.190	<0.001 *
Median red	102.56	2.614	83.69	2.501	5.217	<0.001 *
Mean green	89.44	1.987	76.28	1.655	5.091	<0.001 *
Median green	86.00	2.273	73.69	2.093	3.985	<0.001 *
Mean blue	78.55	2.247	72.00	1.908	2.221	0.034 *
Median blue	74.38	2.196	68.63	2.340	1.792	0.083
Mean luminance	92.34	1.998	78.50	1.756	5.202	<0.001 *
Median luminance	89.63	2.311	75.81	2.223	4.307	<0.001 *

Independent samples t-tests were also conducted on the mean and median luminance levels of the grayscale sets of images (Tables 8-10). While we attempted to control luminance differences in the first two grayscale sets (original/typical stimulus set and the transfer stimulus set both converted to grayscale), there was a difference in the median luminance levels of the transfer set ($t(55) = 2.072$; $p = 0.043$, two-tailed), such that Group 1 ($M = 74.30$, $SEM = 2.193$) was lighter than Group 2 ($M = 67.22$, $SEM = 2.651$). As for the third grayscale stimulus set, our intention was to significantly darken Group 1 and lighten Group 2, essentially reversing the observed luminance differences. Analyses confirmed that there was a significant difference between the two groups in both the mean ($t(46) = -17.37$, $p < 0.001$) and median ($t(46) = -12.07$, $p < 0.001$) luminance levels.

Table 8. Independent samples t-test for the mean and median levels of the luminance of photos used in the initial stimulus set, converted to grayscale in Experiment 3 (Group 1: N = 54; Group 2: N = 54).

	Group 1		Group 2			
Measure	*Mean*	*SEM*	*Mean*	*SEM*	*t*	*p (two-tailed)*
Mean luminance	79.57	1.255	77.64	1.586	0.954	0.342
Median luminance	74.96	1.368	71.46	1.817	1.539	0.127

Table 9. Independent samples t-test for the mean and median levels of the luminance of photos used in the transfer set, converted to grayscale in Experiment 3 (Group 1: N = 30; Group 2: N = 27).

	Group 1		Group 2			
Measure	*Mean*	*SEM*	*Mean*	*SEM*	*t*	*p (two-tailed)*
Mean luminance	76.96	1.795	73.20	2.015	1.399	0.167
Median luminance	74.30	2.193	67.22	2.651	2.072	0.043 *

Table 10. Independent samples t-test for the mean and median luminance levels of new photos in grayscale used in Experiment 3, the luminance intentionally reversed for the two groups (Group 1: N = 30; Group 2: N = 18)

	Group 1		Group 2			
Measure	*Mean*	*SEM*	*Mean*	*SEM*	*t*	*p (two-tailed)*
Mean luminance	57.50	1.273	91.39	1.354	-17.37	< 0.001 *
Median luminance	53.77	1.446	82.56	1.912	-12.07	< 0.001 *

b) Testing with grayscale images

The conclusion from the above analysis is that color cues could have been used by subjects to select the in-group or out-group member, as there were measurable differences in the level of red, green and blue, and in some cases, in the luminance as well. However, these differences disappear with the grayscale images. Therefore, if subjects were attending to color and luminance differences to help them solve the task correctly, they should perform at chance when presented with the grayscale images. Experiment 3 also presented the test trials as probes embedded within normal testing, therefore comparisons were made between the first 40 probe trials (grayscale) and the first 40 typical group membership trials that were performed in the same session.

For the first test, we presented the original stimulus set in grayscale. As a group, performance was above chance in all conditions (In-group Odd original: Gh = NS, Gp = 209.33, df =1, $p < 0.001$; In-group Odd grayscale original: Gh = NS, Gp = 110.93, df = 1, $p < 0.001$; Out-group Odd original: Gh = NS, Gp = 241.90, df = 1, $p < 0.001$; Out-group Odd grayscale original: Gh = NS, Gp = 123.23, df = 1, $p < 0.001$). While subjects performed significantly above chance on the probe trials, generalization was also assessed by comparing performance on probe trials compared to performance on typical trials presented within the same session (Table 11). Table 12 reports the means for each subject. Results revealed that Mason performed worse on both probe conditions, while Bias performed poorer on In-group Odd probes and Wilma was poorer on Out-group Odd probes (Figure 6). We also assessed whether subjects performed differently on In-group Odd and Out-group Odd conditions for either the typical trials or probe trials and found that no subject performed differently in either condition (Table 13). Performance differences are summarized in Figure 6.

Table 11. McNemar tests comparing performance on typical trials and probe trials of the original stimulus set converted to grayscale.

	In-group Odd		Out-group Odd	
Subject	*N*	*p*	*N*	*p*
Bias	40	0.049 *	40	0.065
Mason	40	0.035 *	40	0.004 *
Wilma	40	0.774	40	0.007 *

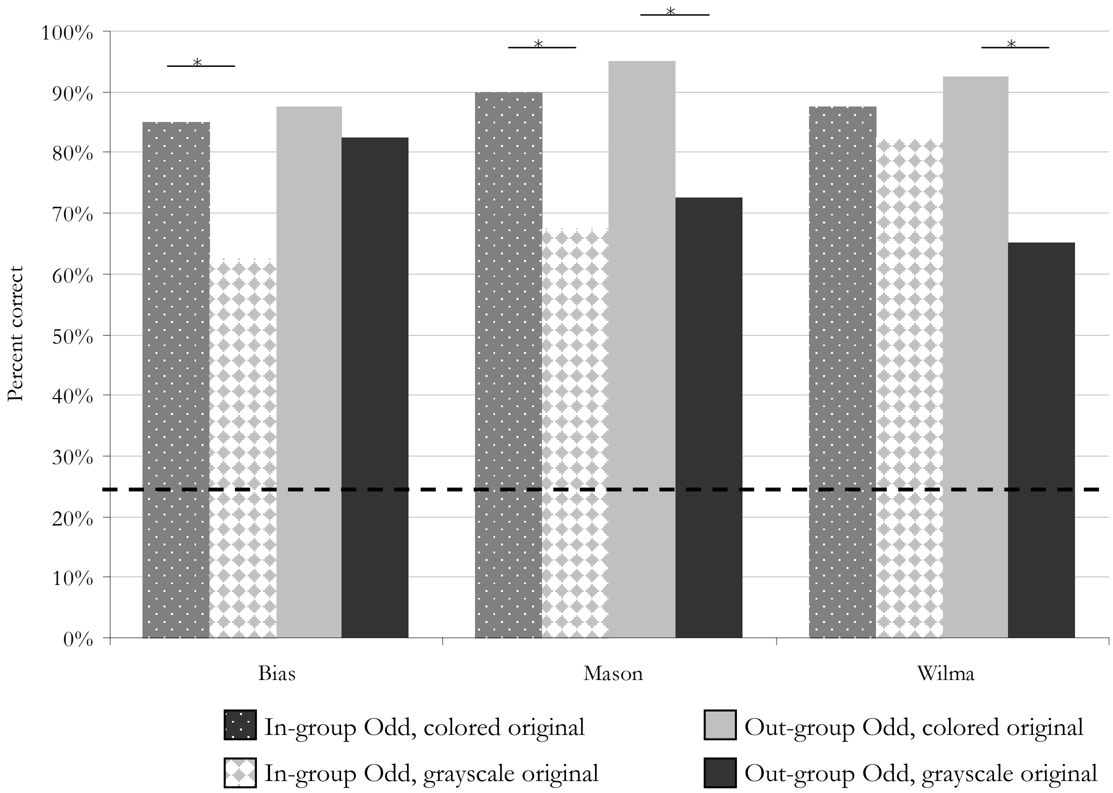
Table 12. Means of each condition by subject for typical trials and probe trials of the original stimulus set converted to grayscale.

	In-group Odd		Out-group Odd	
	Typical	Probe	Typical	Probe
Subject	*M*	*M*	*M*	*M*
Bias	77.5%	62.5%	87.5%	57.5%
Mason	90.0%	67.5%	95.0%	72.5%
Wilma	87.5%	82.5%	92.5%	65.0%

Table 13. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and grayscale probe trials. All comparisons are non-significant (a = 0.05, N = 40), indicating that performance did not differ depending on what condition subjects were performing, the In-group Odd or the Out-group Odd condition.

	Typical	Probe
	p	*p*
Bias	0.289	0.238
Mason	0.687	0.754
Wilma	0.727	0.118

*Figure 6. Individual's performance on Group Membership trials with original stimuli and on probe trials with the original stimulus set converted to grayscale in both the Out-group Odd and In-group Odd conditions. Grayscale original stimuli = 40 probe trials, Original stimuli = 40 trials during same sessions in which probe trials were presented. Subject is on the X-axis and the percentage correct is along the Y-axis. The dotted line indicates chance level (25%). All subjects performed significantly above chance in all trial types. A * indicates whether performance on the condition was significantly ($p < 0.05$) different than the comparison condition according to a McNemar test.*



The latency to respond in a given condition can also reveal differences in cognitive processing. All latencies were analyzed for normalcy as outlined in the methods and transformed if needed. First we assessed whether subjects responded differently to the probe trials than to the typical trials (Table 14). Mason was slower to respond to probe trials in the Out-group Odd condition compared to Out-group Odd trials with the original stimuli.

Next we looked for response differences depending on whether subjects were selecting the in-group or out-group (Table 15). Only Wilma demonstrated any difference, responding faster when selecting in-group members in typical trials.

Table 14. Comparison of latencies to respond to typical (N = 40) and grayscale original probe trials (N = 40) in each condition. Medians and IQRs are reported untransformed while t values may be from transformed data (see Methods).

		Typical		Probe			
Subject	Condition	*Mdn*	*IQR*	*Mdn*	*IQR*	*t*	*p (two-tailed)*
Bias	In-group Odd	2228.00	1487.75	2669.00	1692.25	-0.708	0.483
	Out-group Odd	2168.00	1770.50	2213.00	1228.50	0.405	0.688
Mason	In-group Odd	1532.00	531.00	1442.00	651.00	0.163	0.872
	Out-group Odd	1443.00	551.00	1658.00	573.25	-2.159	0.037*
Wilma	In-group Odd	1322.00	570.00	1392.00	759.00	-1.188	0.242
	Out-group Odd	1617.50	435.75	1442.00	611.00	1.842	0.073

Table 15. Comparison of the latency to respond by subject in the In-group Odd and Out-group Odd condition for typical trials (N = 40) and grayscale original probe (N = 40) trials.

		In-group Odd		Out-group Odd			
Subject	Condition	*Mdn*	*IQR*	*Mdn*	*IQR*	*t*	*p (two-tailed)*
Bias	Typical	2228.00	1487.75	2168.00	1770.50	0.460	0.648
	Probe	2669.00	1692.25	2213.00	1228.50	1.813	0.078
Mason	Typical	1532.00	531.00	1443.00	551.00	1.051	0.300
	Probe	1442.00	651.00	1658.00	573.25	-1.341	0.188
Wilma	Typical	1322.00	570.00	1617.50	435.75	-3.475	0.001*
	Probe	1392.00	759.00	1442.00	611.00	-0.070	0.945

The second grayscale set presented to subjects was the transfer set that was used previously, converted to grayscale. Subjects performed above chance in all conditions (In-group Odd original: Gh = NS, Gp = 186.11, *df* = 1, $p < 0.001$; In-group Odd grayscale transfer: Gh = NS, Gp = 26.27, *df* = 1, $p < 0.001$; Out-group Odd original: Gh = NS, Gp =228.32, *df* = 1, $p < 0.001$; Out-group Odd grayscale transfer: Gh = NS, Gp = 41.41, *df* = 1, $p < 0.001$). To assess individual transfer, subjects must perform at or above 40% in 40 trials to be significantly above chance (binomial test, $\alpha = 0.05$). In this case, Mason was not above chance in the Out-group Odd probe condition. While probe trials were significantly above chance as a group, generalization was also assessed by comparing probe performance to typical performance (Table 16). In this case, all subjects performed poorer on probe trials than on typical trials performed within the same session (Table 17). Performance differences between the In-group Odd and Out-group odd conditions were also assessed and no subject demonstrated a significant difference between condition types for either typical trials or probes (Table 18). Results are summarized in Figure 7.

Table 16. McNemar tests comparing performance on typical trials and probe trials of the transfer stimulus set converted to grayscale.

	In-group Odd		Out-group Odd	
Subject	*N*	*p*	*N*	*p*
Bias	40	0.013 *	40	0.003 *
Mason	40	<0.001 *	40	<0.001 *
Wilma	40	<0.001 *	40	0.004 *

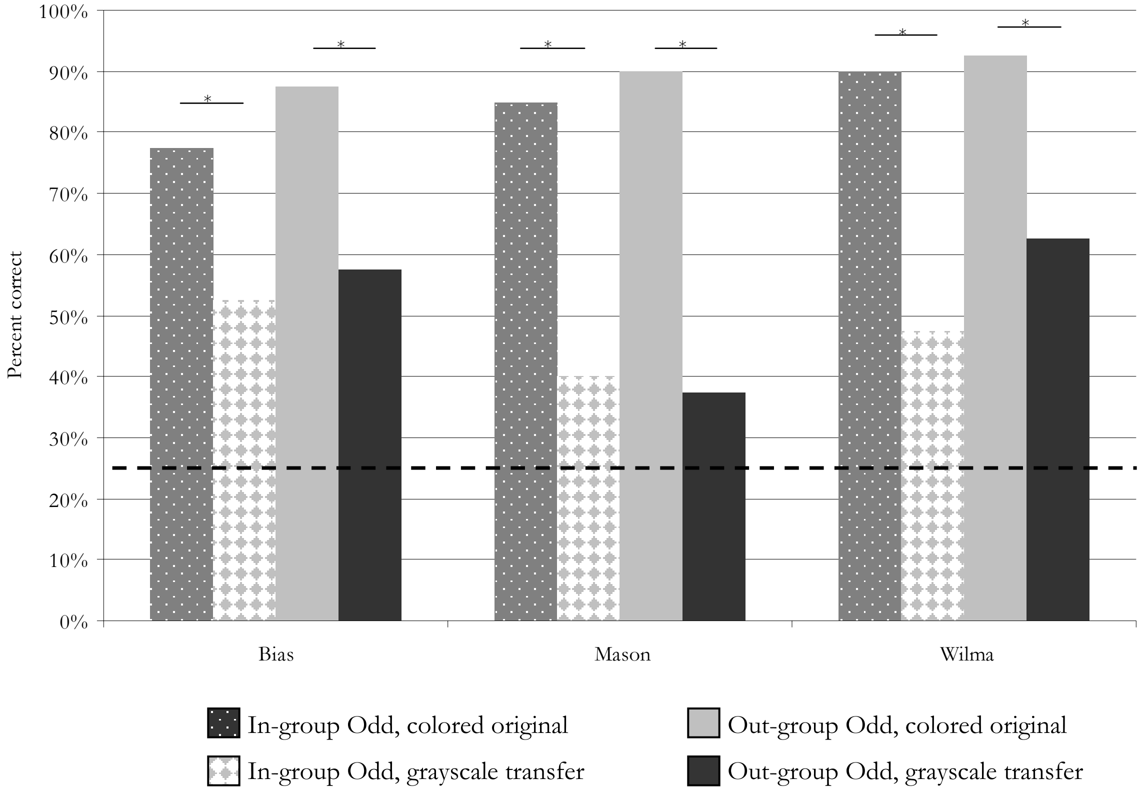
Table 17. Means of each condition by subject for typical trials and probe trials of the transfer stimulus set converted to grayscale.

	In-group Odd		Out-group Odd	
	Typical	Probe	Typical	Probe
Subject	*M*	*M*	*M*	*M*
Bias	85.0%	60.0%	87.5%	50.0%
Mason	85.0%	40.0%	90.0%	37.5%
Wilma	90.0%	47.5%	92.5%	62.5%

Table 18. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and probe trials. All comparisons are non-significant (a = 0.05, N = 40), indicating that performance did not differ depending on what condition subjects were performing, the In-group Odd or the Out-group Odd condition.

	Typical	Probe
	p	*p*
Bias	1.000	0.503
Mason	0.754	1.000
Wilma	1.000	0.286

*Figure 7. Individual's performance on Group Membership trials with original stimuli and on probe trials with the transfer stimulus set converted to grayscale in both the Out-group Odd and In-group Odd conditions. Grayscale transfer stimuli = 40 probe trials, Original stimuli = 40 trials during same sessions in which probe trials were presented. Subject is on the X-axis and the percentage correct is along the Y-axis. The dotted line indicates chance level (25%). All subjects performed significantly above chance in all trial types except for Mason in the Out-group Odd probe condition. A * indicates whether performance on the condition was significantly (p < 0.05) different than the comparison condition according to a McNemar test.*



Latency responses were again assessed to examine differences between typical versus probe trials (Table 19) and differences when selecting the in-group or out-group (Table 20). Wilma was the only subject to show any differences in her response time. She was faster on typical trials than probe trials when selecting the out-group individual, but no different when selecting the in-group member.

Table 19. Comparison of response latencies to typical (N = 40) and grayscale transfer probe trials (N = 40) in each condition.

		Typical		Probe			
Subject	Condition	*Mdn*	*IQR*	*Mdn*	*IQR*	*t*	*p (two-tailed)*
Bias	In-group Odd	2223.00	1363.75	2844.00	1072.00	-1.922	0.062
	Out-group Odd	2128.00	983.50	2258.00	1041.00	-0.548	0.587
Mason	In-group Odd	1312.00	552.75	1587.00	736.00	-1.506	0.140
	Out-group Odd	1612.00	556.00	1648.00	831.75	-1.676	0.102
Wilma	In-group Odd	1286.50	458.25	1387.00	438.50	-0.426	0.673
	Out-group Odd	1322.00	351.00	1522.00	292.50	-2.907	0.006*

Table 20. Comparison of the latency to respond by subject in the In-group Odd and Out-group Odd condition for typical trials (N = 40) and grayscale transfer probe trials (N = 40).

		In-group Odd		Out-group Odd			
Subject	Condition	*Median*	*IQR*	*Median*	*IQR*	*t*	*p (two-tailed)*
Bias	Typical	2223.00	1363.75	2128.00	983.50	-0.225	0.824
	Probe	2844.00	1072.00	2258.00	1041.00	1.220	0.230
Mason	Typical	1312.00	552.75	1612.00	556.00	-1.799	0.080
	Probe	1587.00	736.00	1648.00	831.75	-1.424	0.162
Wilma	Typical	1286.50	458.25	1322.00	351.00	0.506	0.616
	Probe	1387.00	438.50	1522.00	292.50	-1.177	0.246

Since this stimulus set was the same as the set used in the transfer session in Experiment 1, converted to grayscale, we were also able to assess whether removing the color information impaired performance by comparing performance when the trials were presented in color and then when they were presented in grayscale. Results revealed that only Mason performed significantly poorer when the images were converted to grayscale than when initially presented in color (Table 21).

Table 21. 2-way chi square contingency tests comparing performance on transfer trials presented in color and probe trials of the transfer stimulus set converted to grayscale.

	In-group Odd				Out-group Odd			
Subject	χ^2	*df*	*N*	*P*	χ^2	*df*	*N*	*p*
Bias	1.45	1	78	0.228	1.70	1	73	0.192
Mason	4.06	1	80	0.044 *	7.24	1	80	0.007 *
Wilma	0.81	1	80	0.370	0.01	1	79	0.928

The final grayscale stimulus set was deliberately modified to reverse luminance differences between the groups, such that we made Group 1 darker in luminance than Group 2 (refer to Table 8). For this test, we only have results from two subjects, Bias and Wilma, as Mason opted out of testing. Subjects performed significantly above chance on the In-group Odd condition (Typical trials: Gh = NS, Gp = 139.55, *df* = 1, $p < 0.001$; Probe trials: Gh = NS, Gp = 27.61, *df* = 1, $p < 0.001$) and the Out-group Odd condition (Original trials: Gh = NS, Gp = 166.00, *df* = 1, $p < 0.001$; Probe trials: Gh = NS, Gp = 30.06, *df* = 1, $p < 0.001$). Though performance was significantly above chance on probe trials, performance did drop compared to typical trials that presented the original stimulus set in color (Table 22). Means can be seen in Table 23. No differences were found depending on whether subjects were selecting the in-group or out-group member (Table 24). These findings are summarized in Figure 8. As far as finding any latency differences, there was only one – Wilma responded faster to the probes than to the typical trials in the Out-group Odd condition (Table 25).

Table 22. McNemar tests comparing performance on typical trials and probe trials of the new grayscale stimulus set intentionally reversed in luminance.

	In-group Odd		Out-group Odd	
Subject	*N*	*P*	*N*	*p*
Bias	40	<0.001 *	40	<0.001 *
Wilma	40	0.031 *	40	<0.001 *

Table 23. Means of each condition by subject for typical trials and probe trials of the novel grayscale stimulus set reversed in luminance.

	In-group Odd		Out-group Odd	
	Typical	Probe	Typical	Probe
Subject	*M*	*M*	*M*	*M*
Bias	92.5%	47.5%	97.5%	57.5%
Wilma	82.5%	57.5%	87.5%	50.0%

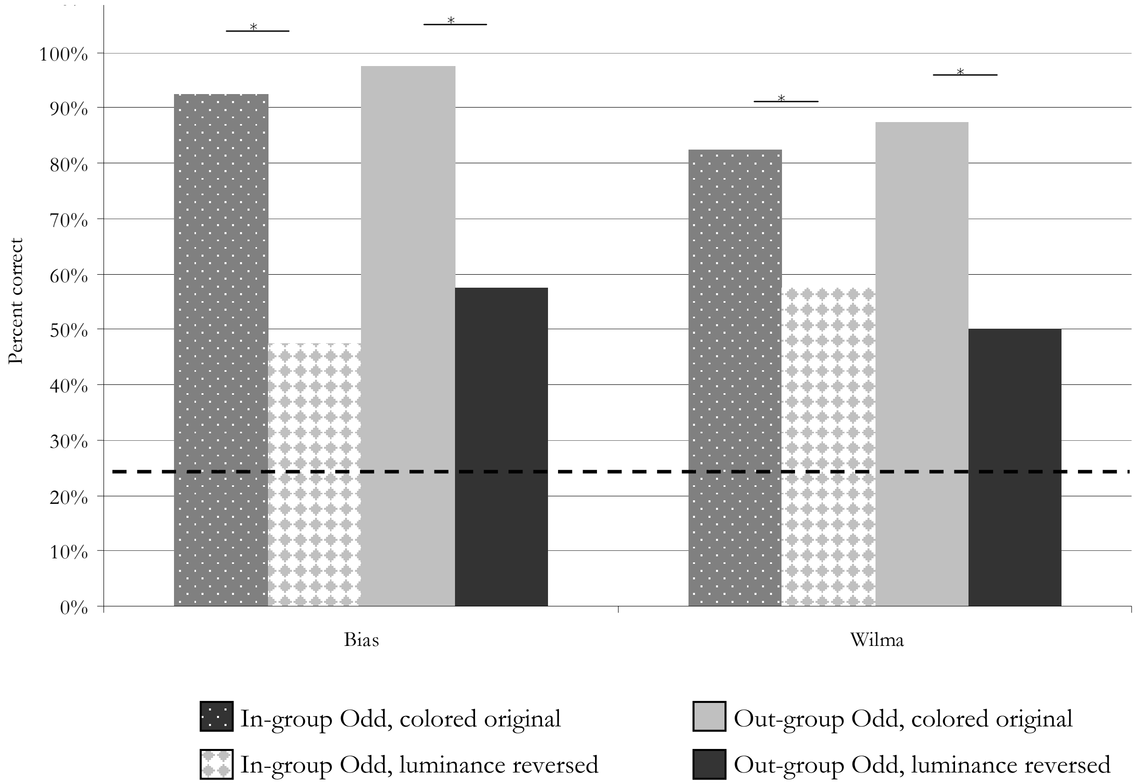
Table 24. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and probe trials. All comparisons are non-significant (a = 0.05, N = 40), indicating that performance did not differ depending on what condition subjects were performing, the In-group Odd or the Out-group Odd condition.

	Typical	Probe
	p	*p*
Bias	0.625	0.424
Wilma	0.727	0.629

Table 25. Comparison of response latencies to typical (N = 40) and grayscale luminance reversed probe trials (N = 40) in each condition.

		Typical		Probe			
Subject	Condition	*Mdn*	*IQR*	*Mdn*	*IQR*	*t*	*p (two-tailed)*
Bias	In-group Odd	2243.00	1178.50	2283.50	1365.00	0.496	0.622
	Out-group Odd	2263.00	1532.00	2398.00	1007.50	0.594	0.556
Wilma	In-group Odd	1347.00	347.75	1382.00	468.50	-1.634	0.110
	Out-group Odd	1252.00	302.75	1547.50	631.00	4.582	<0.001*

*Figure 8. Individual's performance on Group Membership trials with original stimuli and on probe trials with the novel grayscale stimulus set reversed in luminance. Grayscale luminance reversed stimuli = 40 probe trials, Original stimuli = 40 trials during same sessions in which probe trials were presented. Subject is on the X-axis and the percentage correct is along the Y-axis. The dotted line indicates chance level (25%). All subjects performed significantly above chance in all trial types. A * indicates whether performance on the condition was significantly ($p < 0.05$) different than the comparison condition according to a McNemar test.*



Discussion Experiment 3

We designed Experiment 3 to examine the role that color information in the images may have played in the capuchins' ability to categorize the images previously presented. To this end, images were presented in grayscale, removing color information. The first stimulus set was the same set that was used throughout all experiments, converted to grayscale. All subjects demonstrated some decrease in performance, whether it was on In-group trials (Bias), Out-group trials (Wilma) or both (Mason). However, subjects did perform relatively well on these probe trials, above 60% which is well above chance (25%). The second stimulus set was the transfer set that was used in Experiment 1. Subjects only were presented with that set of images for two sessions six months prior, so the stimuli were still rather novel. In this case, subjects performed worse on all conditions, though they were still significantly above chance. We were also able to compare performance on the stimuli when presented in color (Experiment 1) and in grayscale (Experiment 3) to determine if removing the color had a detrimental effect on their performance. Only Mason appeared to have been affected by removing the color, as Bias and Wilma performed the same on the grayscale images as when they were presented in color. From these two tests, we would conclude that subjects were still able to recognize the individuals depicted in the images and categorize them based on their group membership. However, removing the color did appear to have some effect on their performance, though not enough to drop to chance levels. Color differences alone cannot therefore explain subjects' successful performance in Experiments 1 and 2.

The third grayscale stimulus set was created to intentionally reverse the observed luminance differences. During experimental sessions, if subjects were solely selecting based on color and luminance differences (i.e. the rule would be to select the lightest individual), subjects would be selecting the lighter individual from Group 1 on typical trials, and then on probes would also select the light individual, which in this case would actually be an individual from Group 2 which would be incorrect. That did not appear to be the case because subjects still performed significantly above chance on these probe trials, though not to the same level as typical trials. If anything, the luminance differences were quite apparent with this third grayscale set, to the point that the odd image should be more obvious (e.g. pop out) to the subject and they should perform exceedingly well on these trials if they were only attending to luminance differences. Performance was actually lower than on typical trials presented in color, which strongly suggests that subjects were not relying on luminance cues to guide their responses.

Methods Experiment 4: Alternatives for group membership categorization – Color cues – self

As just noted, color cues may have been present in the portraits presented to subjects. Aside from presenting the portraits in grayscale, we also presented subjects with a portrait of themselves. Subjects should be a similar color to their group mates, though they have not seen images of themselves. Therefore, if subjects used color cues to categorize the images and not the identity of the individual depicted, then subjects should categorize their own image as an in-group member. Alternatively, if subjects are using the identity and group membership of the individual depicted in the portrait, then subjects should categorize their own image as an out-group member.

Procedure

We conducted four test sessions of probe trials that include a portrait of the subject itself with either in-group or out-group members, resulting in 40 self/in-group and 40 self/out-group probe trials in both the In-group Odd and the Out-group Odd conditions. This allowed us to assess whether subjects perceived their own image as an in-group or out-group individual. None of our subjects have had extensive experience with their own image though all served as subjects in a mirror study conducted 3-4 years prior (de Waal, Dindo, Freeman, & Hall, 2005), providing them with less than one hour of total mirror exposure. A year after subjects were initially tested with their own image, we installed a body-sized mirror (3' x 3') in both animal enclosures for a period of 1-2 months and then tested subjects again on the same task. This was to assess whether subjects differently categorized their own image after having had extensive exposure to a mirror. One subject, Mason, was unable to be retested because he obtained alpha status and refused to test any longer. Therefore, his results are only reported for the initial self test.

Stimuli

The stimulus set for probe trials contained two new portraits for every adult or subadult in the two groups. The initial stimulus set contained 38 images (20 from Group 1 and 18 from Group 2). Three individuals were removed from Group 2 after the initial test and before re-testing, so those individuals were removed from the stimulus set. Therefore, the re-test stimulus set contained 32 images (20 from Group 1 and 12 from Group 2).

Specific predictions for before and after mirror exposure (predictions are the same):

1 – Know thyself: Subjects perceived their own image as familiar and belonged to their own group.

2 – Self as stranger: Subjects saw their own image as an out-group/unfamiliar individual. When subjects were initially presented with their own image, and if they were making choices based on common color, then they would categorize their own portrait as an in-group individual. However, if subjects were categorizing based on the identity of the depicted individual, then they would likely categorize their own portrait as an out-group member as they have had no experience seeing "self" in the group. Following mirror exposure, we predicted that subjects would categorize their own portrait as an in-group member. For this experiment we compared performance before and after mirror exposure to determine if subjects altered their responding following exposure to a mirror, essentially after exposure to "self."

Our ultimate prediction was that subjects would initially categorize their own image as out-group and then following mirror exposure, they would categorize it as an in-group member. Predictions are also presented numerically in Table 26.

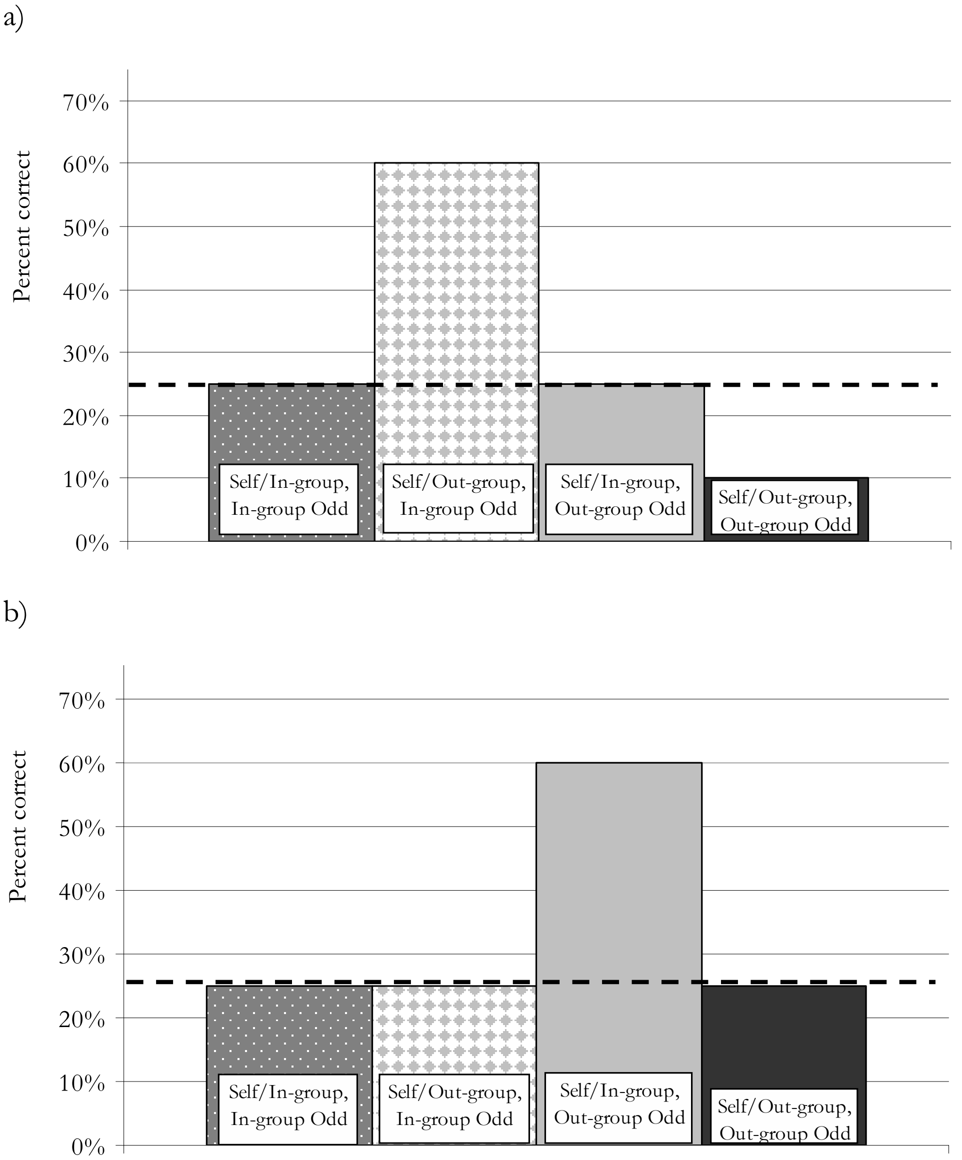
Table 26. Specific predictions for Self Group Membership test.

	Trial type	Same	Odd	Prediction select self as odd1) self as in-group	Predictions select self as odd 2) self as stranger
In-group Odd condition	Typical	3 out-group	1 in-group	> 25%	> 25%
	Probe	3 in-group	Self	= 25%	= 25%
		3 out-group	Self	> 25%	= 25%
Out-group Odd condition	Typical	3 in-group	1 out-group	> 25%	> 25%
	Probe	3 in-group	Self	= 25%	> 25%
		3 out-group	Self	< 25%	= 25%

Results Experiment 4

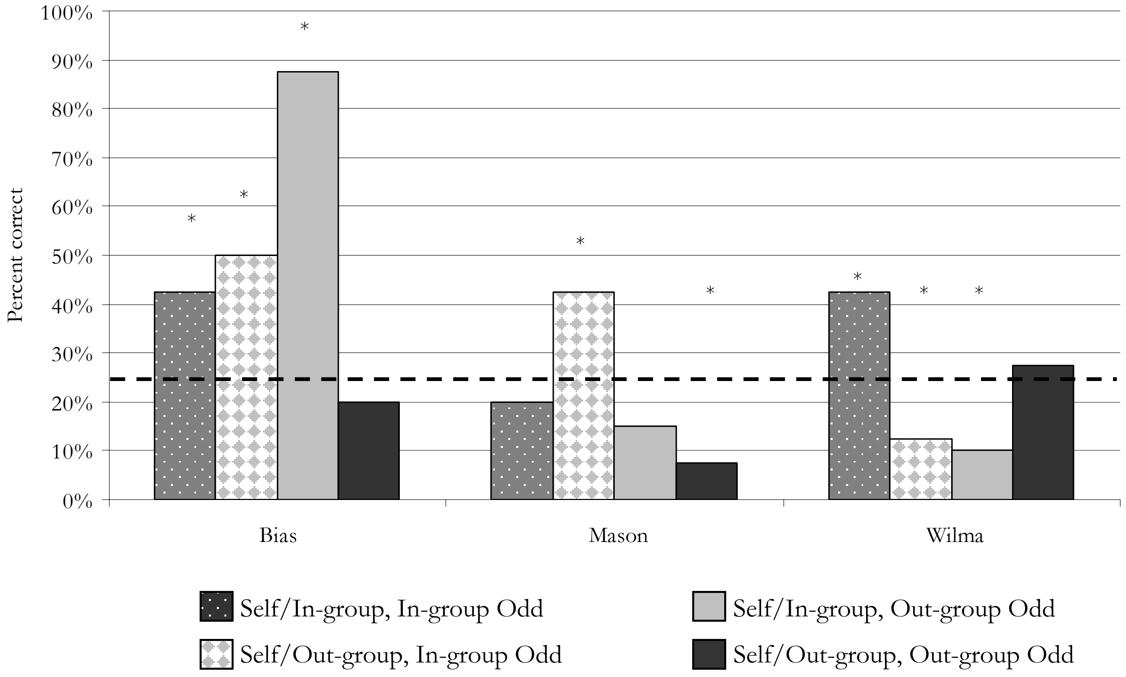
This experiment was designed to follow-up whether color cues were guiding the subjects' behavior. Our monkeys have limited exposure to their own reflection and therefore should not be familiar with their own image. When presented with their own image in the group membership oddity task, if their selection is based on knowledge of the individuals depicted, they should categorize their own image as an out-group member. If categorization is based on color similarities, then they should categorize their own image as an in-group member, as they resemble other individuals from their in-group. Figure 9 shows the ideal pattern that subjects should follow if they categorize their own image as a) an in-group member or b) as an out-group member.

Figure 9. Displays the ideal pattern that should be found if subjects categorize their own image as either a) an in-group member or b) an out-group member.



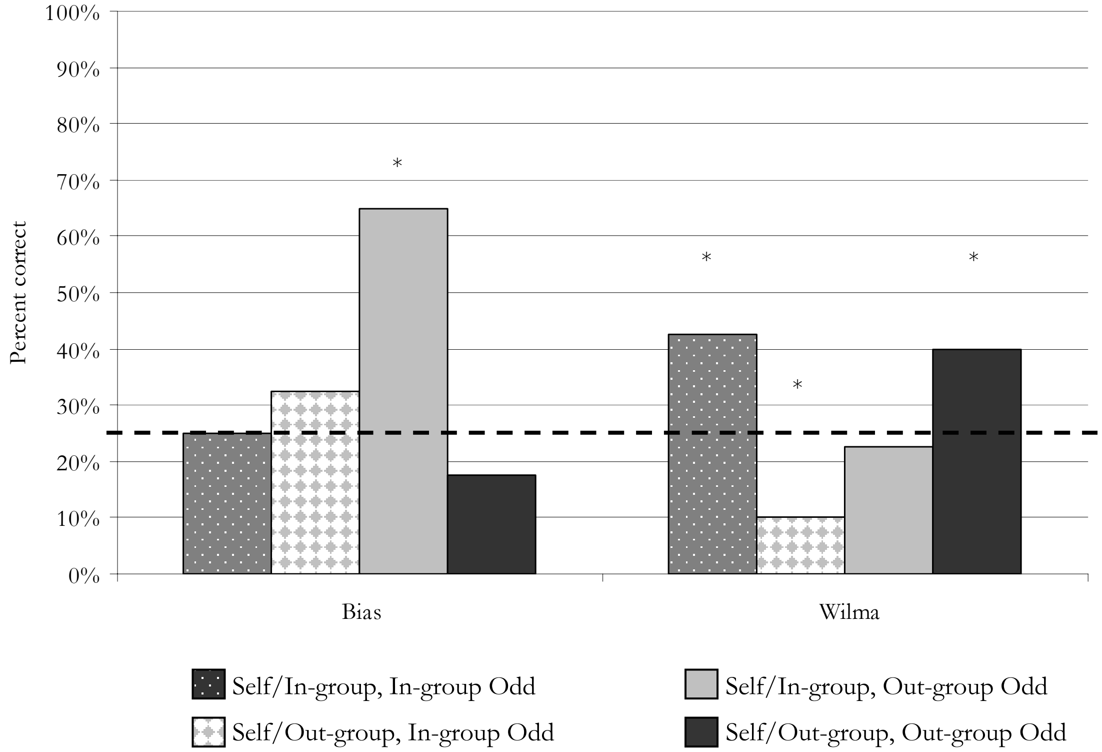
Subjects must perform at or above 40% in 40 trials for performance to be significantly above chance (binomial test, α = 0.05). Looking at the results from the first self test, before subjects had any exposure to a mirror, Mason appears to follow the pattern we predicted if he categorized his own image as an in-group member, or as being familiar (Figure 10). Bias, on the other hand, follows categorizing her own image as an out-group, or unfamiliar, member. Wilma did not follow either predicted pattern.

*Figure 10. Individual's performance on the Group Membership task when images of own portrait were presented prior to mirror exposure. Self image was presented either with 3 in-group members or 3 out-group members in the In-group Odd and Out-group Odd conditions. The first 40 trials of each condition are included in the graph. The dotted line indicates chance (25%) level. A * indicates significantly different (higher or lower) than chance performance on that condition according to a binomial test.*



After a mirror was installed in the home enclosure for at least a month, subjects were re-tested on the test task. Unfortunately, we were unable to test Mason again. Therefore, Figure 11 displays the results for only Bias and Wilma. Both subjects displayed the same pattern as before any mirror exposure, Bias categorized her image as out-group and Wilma did not following either predicted pattern.

*Figure 11. Individual's performance on the Group Membership task when images of own portrait were presented after mirror exposure. Self image was presented either with 3 in-group members or 3 out-group members in the In-group Odd and Out-group Odd conditions. The first 40 trials of each condition are included in the graph. The dotted line indicates chance (25%) level. A * indicates significantly different (higher or lower) than chance performance on the given condition.*



Discussion Experiment 4

Following up on whether color cues could have guided performance on previous tasks, this task presented subjects with an image of themselves, along with either three in-group members or three out-group members. Based on our specific predictions whether subjects would categorize their image as an in-group or out-group member, Bias categorized her image as out-group while Mason categorized his as in-group. Wilma did not follow either predicted pattern. A year after completing the first self test, we installed a large mirror in each enclosure for at least a month before re-testing subjects on the same task to assess whether subjects would change their responses based on now experiencing their own image in the space that other in-group members share. Unfortunately, we were unable to test Mason again. As for Bias and Wilma, they followed the same pattern of behavior as they displayed during the first test. Bias categorized her image as in-group while Wilma did not follow any predicted pattern. These findings are difficult to resolve as there was no concordance among subjects. We can conclude again that color differences alone cannot explain prior performance, at least not for Bias or Wilma because neither categorized their image as an in-group member, and their own coloration should match that of the group.

It should be noted that we do not believe that Mason recognized his own image as being "himself," as capuchin monkeys have not been shown to have mirror self-recognition (for a review Anderson & Gallup, 1999). However, in a previous study our capuchin monkeys did not respond to their mirror image as being a stranger either (de Waal et al., 2005). A critical difference between mirror studies and our experiment here is that here the self image is a static image while mirror studies have dynamic stimuli, allowing subjects to use temporal information from their own body movements and visual feedback of the image's movements. Therefore, Experiment 4 was not examining self-recognition per se. Should subjects have responded differently after mirror exposure, namely by categorizing their own image as familiar, our interpretation would have been that subjects recalled seeing this "new" individual with other individuals in their home enclosure, therefore the image is an in-group member or at least familiar. However, this was not the case as neither Bias nor Wilma changed their response after mirror exposure.

Methods Experiment 5: Oddity or categorization?

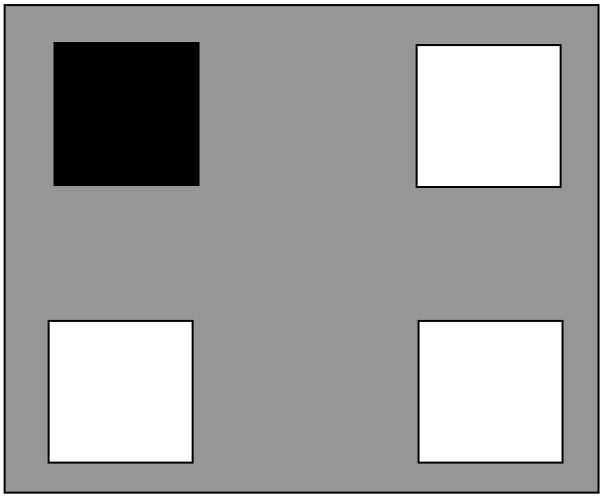
Experiment 5 was designed to determine whether subjects were applying the oddity concept to the dimension of group membership. In Experiment 1 we found that subjects appeared to have difficulty performing the task if the In-group Odd and Out-group Odd condition were presented together within a session (Appendix A). In addition, when switching from one condition to the other, subjects would drop to below chance (25%) levels for approximately two sessions before performance would rise back up to previous levels, typically above 85% (Figure 3). If subjects were applying the oddity concept, then they should be unable to perform the task if only two images were displayed, because the principle behind the oddity task is that it is the relationship between the images that determines which image is odd. With only two images there is no relationship and no way to determine which image is odd. However, if subjects are applying two categorization rules: 1) select the in-group member, or 2) select the out-group member, then subjects would be able to perform the task when only two images are presented as probe trials embedded within typical trials (four images presented).

Procedure

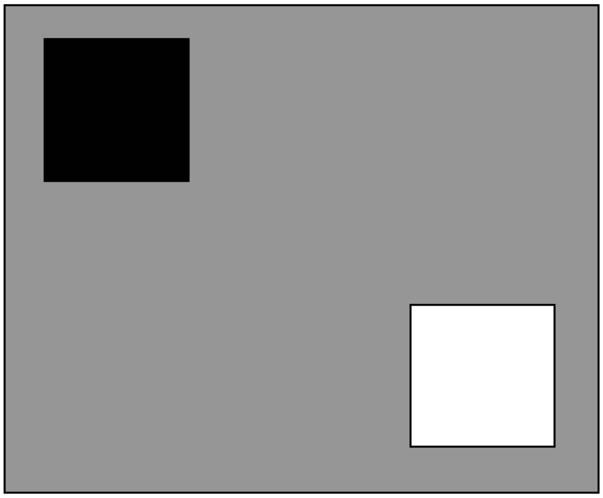
Test trials displayed two stimuli, one in-group and one out-group member, and were presented as probe trials. One session consisted of 20 typical trials (four images from original stimulus set, Figure 12a), 20 probe trials (two images from original stimulus set, Figure 12b), and 35 clip art images. Subjects performed both conditions, In-group Odd and Out-group Odd. Only Bias and Wilma participated in this experiment as Mason refused to test once he obtained alpha status.

Figure 12. Representation of Experiment 5 trial types. The black squares represent portraits from Group 1 and the white squares represent portraits from Group 2. Two trial types were presented: a) typical trials that presented four images, three from the same group and one from the other group and b) probe trials that presented two images, one from each group.

a)



b)



Results Experiment 5

Experiment 5 examined whether subjects were applying the oddity concept to the Group Membership Oddity Task we employed throughout this study. Subjects performed significantly above chance in all conditions (In-group Odd typical trials: Gh = NS, Gp = 158.95, *df* = 1, $p < 0.001$; Out-group Odd typical trials: Gh = NS, Gp = 145.76, *df* = 1, $p < 0.001$; In-group Odd probe trials: Gh = NS, Gp = 173.41, *df* = 1, $p < 0.001$; Out-group Odd probe trials: Gh = NS, Gp = 173.41, *df* = 1, $p < 0.001$). In addition, there was no difference in performance between typical trials and probe trials in either condition (Table 27). Subjects also performed the same on In-group Odd and Out-group Odd trials regardless if they were typical trials or probe trials (Table 28). Performance results are summarized in Figure 13.

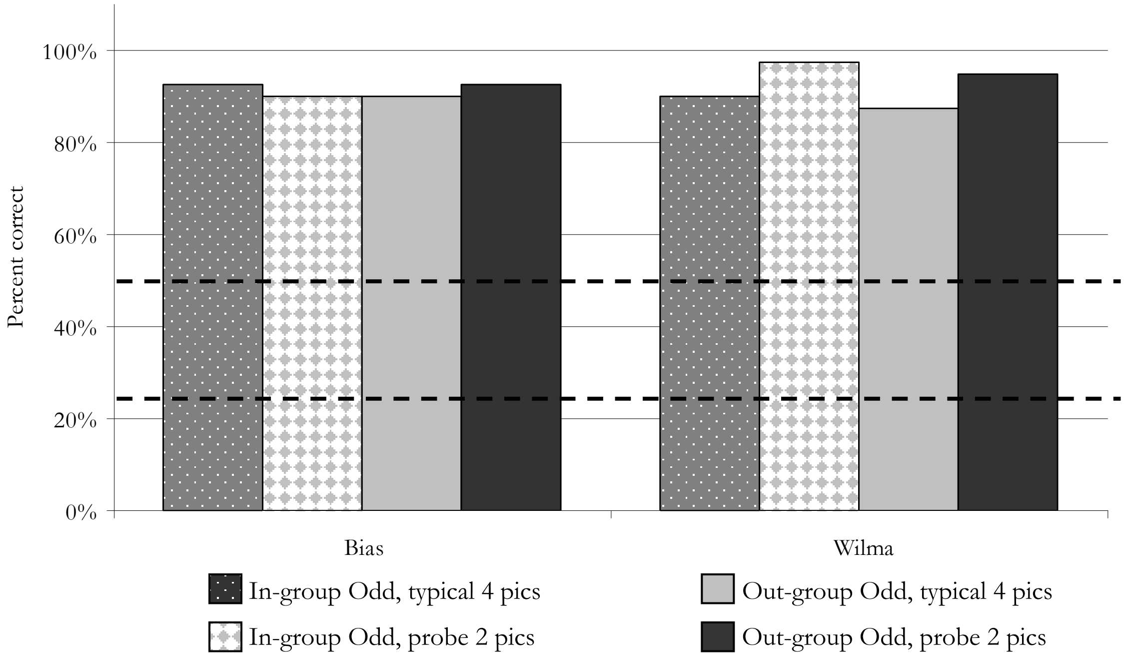
Table 27. McNemar tests comparing performance on typical trials presenting four pictures and probe trials presenting two pictures. All comparisons are non-significant (a = 0.05, N = 40).

	In-group Odd	Out-group Odd
	p	*p*
Bias	1.000	1.000
Wilma	0.250	0.453

Table 28. Results of McNemar tests comparing In-group Odd and Out-group Odd performance for typical and probe trials. All comparisons are non-significant (a = 0.05, N = 40), indicating that performance did not differ depending on what condition subjects were performing, the In-group Odd or the Out-group Odd condition.

	Typical	Probe
	p	*p*
Bias	1.000	1.000
Wilma	1.000	1.000

Figure 13. Individual performance on typical and probe trials for In-group Odd and Out-group Odd conditions in Experiment 5. Probe trials presented two pictures (one in-group member and one out-group member). Each bar represents the first 40 trials performed in each condition. The lower dotted line indicates chance (25%) level for typical trials with 4 images and the upper dotted line indicates chance (50%) level for probe trials with 2 images. Subjects were significantly above the respective chance level in each condition according to a binomial test.



Given that there was no decrease in performance when two pictures were presented as opposed to four, it appeared the task was not cognitively more difficult. Latencies can be sensitive to processing differences, so we examined whether there were latency differences when subjects performed the probe trials compared to the typical trials (Table 29). Bias responded the same, but Wilma responded faster to both probe conditions, suggesting faster cognitive processing when only two images appeared.

Table 29. Comparison of response latencies to typical trials in which four images appeared (N = 40) and probe trials in which two images appeared (N = 40)

		Typical		Probe			
Subject	Condition	*Mdn*	*IQR*	*Mdn*	*IQR*	*t*	*p (two-tailed)*
Bias	In-group Odd	2538.50	204.62	2634.00	1414.50	1.099	0.279
	Out-group Odd	2188.00	1032.25	2323.00	962.00	0.430	0.670
Wilma	In-group Odd	1447.00	530.25	1052.00	250.00	-5.266	<0.001*
	Out-group Odd	1247.00	350.75	951.50	140.00	-5.522	<0.001*

Discussion Experiment 5

Experiment 5 followed up on an earlier finding from Experiment 1, namely that subjects were unable to perform the Group Membership Oddity task when both conditions were presented mixed within a session. This suggested to us that subjects may not have been applying the oddity concept to the dimension of group membership. Our subjects are able to perform the oddity task using images of clip art and faces, when three faces are of the same individual and one individual is different (Pokorny & de Waal, in press), so they do understand the oddity concept and apply it to faces when presented in an individual recognition task. However, it appears they did not apply the oddity concept in this study. In principle the oddity task is ideal for this research question, as it presents three individuals are the same along a given dimension (in-group/out-group), along with another individual that does not fit in the same category. Solving the task in this manner would require that subjects apply knowledge of third-party relationships. Alternatively, subjects could still solve the task by taking an egocentric approach, essentially categorizing each individual based on the relationship the subject had with that individual (in-group/out-group or familiar/unfamiliar). The task then becomes one of categorization and the application of two rules: either 1) select the in-group (In-group Odd) or 2) select the out-group (Out-group Odd). Application of either of these strategies in the previous tasks would appear the same. However, in Experiment 5, when only two images appeared on the screen in probe trials, subjects should have been unable to perform the task if strictly applying the oddity concept. Given that this is not what we found, we conclude that subjects were not applying the oddity concept along the group membership dimension and instead applied two rules, likely from an egocentric perspective. While this does not change our conclusions from previous experiments, it does give us some insight as to how they approached the task and it will likely influence how we design experiments in the future that investigate other categorical dimensions.

General Discussion

The motivation behind this study was to examine social cognition in capuchin monkeys. What do they know about the individuals they live with in their group? One of the basic social categories to investigate is group membership, in-group vs. out-group, as outsiders pose a threat to a group's mating and food resources and we know that capuchins respond agonistically to outsiders in both natural and artificial environments (Cooper et al., 2001; Crofoot, 2007; Fragaszy et al., 1994; Spironello, 2001). This provides a behavioral indication that capuchins can and do discriminate in-group from out-group members. We sought to design a task that enabled subjects to specify which individuals belonged to which category. Previously we demonstrated that these same subjects were able to discriminate both familiar and unfamiliar conspecific faces (Pokorny & de Waal, in press), which is one of the criteria of categorization, the ability to discriminate between objects or individuals within a category (Zayan & Vauclair, 1998). The second criterion is the ability to discriminate between exemplars belonging to different categories, which we addressed in the current study.

Capuchins performed a task in which they selected the one individual who did not belong to the same social group at the other three individuals presented on the screen. The task required that subjects select among similar looking conspecifics that ideally should only differ based on group membership. This is different from other categorization tasks that could be solved based on perceptual similarity (e.g. Neiworth et al., 2004; Schrier et al., 1984; Schrier & Brady, 1987; Vonk & MacDonald, 2002). Instead, it is similar to studies that required subjects to categorize along a functional associative class, such as food (Bovet & Vauclair, 1998; Deputte et al., 2001; Inoue et al., 2008; Santos et al., 2001) or gender (Inoue et al., 2008; Koba & Izumi, 2006). In the case of food, there is no one feature or characteristic that indicates if an item is food or not, it is the function, whether the item is edible or not. Likewise, there are no specific features that signify individuals are from Group 1 as opposed to Group 2. In fact, it is unlikely that humans unfamiliar with our colony would be able to categorize the images correctly, though we did not explicitly test this in the current study.

As for detecting group membership in the current study, there should be nothing inherently different about the images from Group 1 compared to Group 2. The only thing that should differentiate Group 1 from Group 2 is the relationship relative to one another and to the subject. By using an oddity paradigm, we wanted subjects to take into account the relationship between individuals presented on the screen to determine who was "odd." However, it was entirely possible that subjects could rely on knowledge of their own relationship with each individual when selecting the odd individual. It appears that subjects did view the task from an egocentric viewpoint, given that they were able to perform the task, with no drop in performance, when only two images, one from each social group, appeared instead of four (Experiment 5). This implies that subjects knew which category they were to select and would pick the individual that fit in the given category, either in-group or out-group.

While we did say that there should be nothing visibly different about the images from Group 1 compared to Group 2, we devoted several tests to exclude the possibility that color differences may have been present between both groups. We presented subjects with several different stimulus sets converted to grayscale and controlled in luminance. In doing so, subjects were still able to correctly categorize the images, as they performed above chance, though the percent correct was less than how they performed on the task with the original colored set of stimuli. In the case where we could compare performance on colored and grayscale images while the stimuli were still relatively novel (the transfer set), Bias and Wilma performed the same whether the color information was present or not. While removing color cues was important to demonstrate that behavior was not driven solely by possible color differences, color does play a role in how individuals select food (Sands, Lincoln, &

Wright, 1982; Santos et al., 2001) and mating partners (e.g. Waitt, Gerald, Little, & Kraiselburd, 2006; Waitt et al., 2003), both of which are critical to one's survival and should not be entirely disregarded.

In general, Experiments 1-3 support our hypothesis that capuchins are able to categorize conspecifics based on the dimension of group membership. This does not necessarily mean that capuchins have a concept of group membership as the only knowledge one needed was knowledge about in-group, or familiar, individuals. We used the term in-group and out-group as opposed to familiar and unfamiliar because it is possible that some individuals may have been familiar at some point in the past and the individuals depicted in the typical stimulus set were presented for several years, so it would be difficult to argue that the images themselves, and those represented in the images, were "unfamiliar." We also suggested that the capuchins were following two rules to solve the task: either 1) select the in-group individual, or 2) select the out-group individual. It is entirely possible that they did this with reference only to in-group, or familiar, individuals, such that the rules would instead be: either 1) select the in-group individual, or 2) avoid the in-group individuals. We cannot determine from this study alone which of these rules subjects were using, though one may argue that it is more likely that subjects operated completely on knowledge of the in-group alone.

It would be difficult to assess whether capuchins have a concept of out-group, or unfamiliar, as this category has infinite possible members that have little in common aside from having no meaning to the subject. In-group, or familiar, may be a more concrete category for animals to utilize as experience allows them to acquire information necessary to group the items together. In fact, chimpanzees have been found to spontaneously sort items based on familiarity, but not on the dimension of unfamiliar, or novelty (Tanaka, 1995). In studies that measured heart rate responses of chimpanzees as they viewed conspecific or human images with whom the subject had varied relationships with, Boysen and Berntson (1986, 1989) found differential responses depending on the quality of the relationship of the individual depicted. While this does not tell us whether the chimpanzee had a concept of "familiar individual" and "unfamiliar individual" it does indicate that she was able to discriminate between those categories.

Likewise, in our study we can conclude that capuchins discriminate between the categories of in-group and out-group under various experimental manipulations. These two categories were defined by the human experimenter, though they correspond to the living situation of our capuchins making it likely that subjects related on some level to our category definitions. The fact that subjects correctly categorize previously unseen individuals (Experiment 3) strongly suggests that our categories were similar.

We did not find any consistent differences based on which condition was being performed, such as whether subjects performed better when selecting the in-group or out-group member. Likewise, latency differences were not consistent between experiments even for the same subject. Therefore we are unable to make broad statements regarding whether subjects selected the in-group or the out-group individual more accurately, or whether they responded faster to one group than the other, as performance varied depending on the experiment, and often there was no difference. In general, performance typically dropped when new images were presented, though it did remain above chance.

Our results also provide supporting evidence that nonhuman primates are able to recognize the representational nature of two-dimensional images. Many researchers utilize two-dimensional images given the amount of control one can have over the stimulus. The underlying assumption is that subjects understand what the image represents. Many studies, including the current one, rely on picture-to-picture transfer meaning that subjects are trained to respond to images in some manner, then are presented with new images to assess generalization. More convincing evidence comes from studies in which subjects responded with species-typical behaviors, such as lip-smacking or eye-avoidance, when images of conspecifics were presented (Mendelson, Haith, & Goldman-Rakic, 1982; Overman & Doty, 1982; Plimpton, Swartz, & Rosenblum, 1981; Rosenfeld & van Hoesen, 1979; Sackett, 1965, 1966). This behavior did eventually disappear, presumably once subjects realized that the image was not a real conspecific.

Other studies have directly assessed equivalence between photos and real objects (for a review see Fagot, Martin-Malivel, & Depy, 2001). Bovet and Vauclair (1998) trained baboons to categorize real food and non-food items. After performing the task successfully and transferring to novel stimuli, subjects were presented with pictures of the items. Categorization of the pictures was similar to previous categorization with the physical items. However, there were some instances in which baboons would grab at the photos, perhaps confusing the image with the real item. In fact, another study found that gorillas ate pictures of bananas, indicating picture-confusion (Parron, Call, & Fagot, 2008).

This finding suggests that experience with two-dimensional items may be necessary before understanding the representational nature of photos. Humans too need exposure to photos for proper recognition and understanding (for a review see Bovet & Vauclair, 2000). Animals that have been given experience with photos appear to comprehend that the photo represents a real item. A chimpanzee that was language trained, and familiarized with photos, was able to name items in photographs (Gardner & Gardner, 1984). Another chimpanzee, Ai, labeled familiar humans and conspecifics whether they were presented in person or in pictures (Matsuzawa, 1990). Bovet and Vauclair (2000) do advise that recognition may be facilitated by using images of familiar items or conspecifics, which was done in the current study. Given that our subjects correctly categorized novel images into in-group versus out-group members suggests that they connected the images to real-life individuals, because otherwise it would be hard to explain their successful categorization.

Overall, this study contributes to our understanding of how capuchins may organize their social world into categories such as in-group and out-group. It also offers a way to examine social categorization in an experimental setting, though the paradigm was not followed as intended by the researchers. Again, this provides us with a greater appreciation of how capuchins approach a task such as this and that they may not solve it in the manner we as humans anticipated.

References

Adams-Curtis, L. E. (1990). Conceptual learning in capuchin monkeys. *Folia Primatologica, 54*, 129-137.

Anderson, J. R., & Gallup, G. G. (1999). Self-recognition in nonhuman primates: past and future challenges. In M. Haug & R. E. Whalen (Eds.), *Animal Models of Human Emotion and Cognition* (pp. 175-194). Washington, DC: American Psychological Association.

Andrew, R. J. (1963). Evolution of facial expression. *Science, 142*, 1034-1041.

Aureli, F., Cozzolino, R., Cordischi, C., & Schucchi, S. (1992). Kin-oriented redirection among Japanese macaques: An expression of a revenge system? *Animal Behaviour, 44*, 283-291.

Beran, M. J., Klein, E. D., Evans, T. A., Chan, B., Flemming, T. M., Harris, E. H., et al. (2008). Discrimination reversal learning in capuchin monkeys (*Cebus apella*). *The Psychological Record, 58*, 3-14.

Bernstein, I. S. (1961). The utilization of visual cues in dimension-abstracted oddity by primates. *Journal of Comparative and Physiological Psychology, 54*(3), 243-247.

Berryman, R., Cumming, W. W., Cohen, L. R., & Johnson, D. F. (1965). Acquisition and transfer of simultaneous oddity. *Psychological Reports, 17*, 767-775.

Bovet, D., & Vauclair, J. (1998). Functional categorization of objects and of their pictures in baboons (*Papio anubis*). *Learning and Motivation, 29*, 309-322.

Bovet, D., & Vauclair, J. (2000). Picture recognition in animals and humans. *Behavioral Brain Research, 109*, 143–165.

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

Boysen, S. T., & Berntson, G. G. (1986). Cardiac correlates of individual recognition in the chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology, 100*, 321-324.

Boysen, S. T., & Berntson, G. G. (1989). Conspecific recognition in the chimpanzee (*Pan troglodytes*): Cardiac responses to significant others. *Journal of Comparative Psychology, 103*(3), 215-220.

Brown, D. A., & Boysen, S. T. (2000). Spontaneous discrimination of natural stimuli by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 114*(4), 392-400.

Bruce, C. (1982). Face recognition by monkeys: Absence of an inversion effect. *Neuropsychologia, 20*, 515-521.

Buckley, M. J., Booth, M. C. A., Rolls, E. T., & Gaffan, D. (2001). Selective perceptual impairments after perirhinal cortex ablation. *Journal of Neuroscience, 21*(24), 9824-9836.

Cheney, D. L., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour, 28*, 362-367.

Cheney, D. L., & Seyfarth, R. M. (1986). The recognition of social alliances by vervet monkeys. *Animal Behaviour, 34*, 1722-1731.

Cheney, D. L., & Seyfarth, R. M. (1990a). *How Monkeys See the World: Inside the Mind of Another Species.* Chicago: University of Chicago Press.

Cheney, D. L., & Seyfarth, R. M. (1990b). Reconciliation and redirected aggression in vervet monkeys (*Cercopithecus aethiops*). *Behaviour, 110*, 258-275.

Cheney, D. L., & Seyfarth, R. M. (1990c). The representation of social relations by monkeys. *Cognition, 37*, 167-196.

Cooper, M. A., Bernstein, I. S., Fragaszy, D. M., & de Waal, F. B. M. (2001). Integration of new males into four social groups of tufted capuchins (*Cebus apella*). *International Journal of Primatology, 22*(4), 663-683.

Crofoot, M. C. (2007). Mating and feeding competition in white-faced capuchins (*Cebus capucinus*): the importance of short- and long-term strategies *Behaviour, 144*, 1473-1495.

D'Amato, M., & Van Sant, P. (1988). The person concept in monkeys (Cebus apella). *Journal of Experimental Psychology: Animal Behavior Processes, 14*(1), 43-55.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals* (1998, 3rd ed.). New York: Oxford University Press.

Dasser, V. (1987). Slides of group members as representations of the real animals (*Macaca fascicularis*). *Ethology, 76*, 65-73.

Dasser, V. (1988). A social concept in Java monkeys. *Animal Behaviour, 36*, 225-230.

De Lillo, C., & Visalberghi, E. (1994). Transfer Index and mediational learning in tufted capuchins (*Cebus apella*). *International Journal of Primatology, 15*, 275-287.

de Waal, F. B. M. (1982). *Chimpanzee Politics: Power and Sex among Apes*. London: Jonathan Cape.

de Waal, F. B. M., Dindo, M., Freeman, C. A., & Hall, M. J. (2005). The monkey in the mirror: Hardly a stranger. *Proceedings of the National Academy of Sciences, USA, 102*(32), 11140-11147.

de Waal, F. B. M., & Pokorny, J. J. (2008). Faces and behinds: Chimpanzee sex perception. *Advanced Science Letters, 1*(1), 9-13.

de Waal, F. B. M., & van Hooff, J. A. R. A. M. (1981). Side-directed communication and agonistic interactions in chimpanzees. *Behaviour, 77*(3), 164-198.

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.

Defler, T. R. (1982). A comparison of intergroup behavior in *Cebus albifrons* and *C. apella. Primates, 23*(3), 385-392.

Deputte, B. L., Pelletier, S., & Barbe, S. (2001). Visual categorization of natural and abstract items in forest monkeys and humans. *Behavioural Processes, 55*(1), 51-64.

Dufour, V., Pascalis, O., & Petit, O. (2006). Face processing limitation to own species in primates: a comparative study in brown capuchins, Tonkean macaques and humans. *Behavioural Processes, 73*(1), 107-113.

Fagot, J., Martin-Malivel, J., & Depy, D. (2001). What is the evidence for an equivalence between objects and pictures in birds and nonhuman primates? In J. Fagot (Ed.), *Picture Perception in Animals.* Philadelphia, PA: Psychology Press Ltd.

Fragaszy, D. M., Baer, J., & Adams-Curtis, L. E. (1994). Introduction and integration of strangers into captive groups of tufted capuchins (*Cebus apella*). *International Journal of Primatology, 15*(3), 399-420.

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The Complete Capuchin: The Biology of the Genus Cebus.* Cambridge: Cambridge University Press.

Gardner, R. A., & Gardner, B. T. (1984). A vocabulary test for chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 98*, 381-404.

Gothard, K. M., Erickson, C. A., & Amaral, D. G. (2004). How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Animal Cognition, 7*, 25-36.

Gouzoules, H., Gouzoules, S., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: representational signaling in the recruitment of agonistic aid. *Animal Behaviour, 32*, 182-193.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science, 146*(3643), 549-551.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 2*(4), 285-302.

Heywood, C. A., & Cowey, A. (1992). The role of the 'face-cell' area in the discrimination and recognition of faces by monkeys. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 335*(1273), 31-37; discussion 37-38.

Inoue, T., Hasegawa, T., Takara, S., Lukats, B., Mizuno, M., & Aou, S. (2008). Categorization of biologically significant objects, food and gender, in rhesus monkeys I. Behavioral study. *Neuroscience Research, 61*, 70-78.

Jack, K. M., & Fedigan, L. M. (2004). Male dispersal patterns in white-faced capuchins, *Cebus capucinus* Part 1: patterns and causes of natal emigration. *Animal Behaviour, 67*, 761-769.

Janson, C. H. (1986). The mating system as a determinant of social evolution in intergroup encounters. In J. G. Else & P. C. Lee (Eds.), *Primate Ecology and Conservation* (pp. 169-179). Cambridge: Cambridge University Press.

Jitsumori, M. (2004). Categorization and formation of equivalence classes in animals: Studies in Japan on the background of contemporary developments. *Japanese Psychological Research, 46*(3), 182-194.

Judge, P. G. (1991). Dyadic and triadic reconciliation in pigtail macaques (*Macaca nemestrina*). *American Journal of Primatology, 23*, 225-237.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*(11), 4302-4311.

Koba, R., & Izumi, A. (2006). Sex categorization of conspecific pictures in Japanese monkeys (*Macaca fuscata*). *Animal Cognition, 9*, 183-191.

Kummer, H. (1971). *Primate Societies.* Arlington Heights, Il: Harlan Davidson, Inc.

Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roithblat, T. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 263-276). Hillsdale, NJ: Erlbaum.

Matsuzawa, T. (1990). Form perception and visual acuity in a chimpanzee. *Folia Primatologica, 55*, 24-32.

Mendelson, M. J., Haith, M. M., & Goldman-Rakic, P. S. (1982). Face scanning and responsiveness to social cues in infant rhesus monkeys. *Developmental Psychology, 18*(2), 222-228.

Moon, L. E., & Harlow, H. F. (1955). Analysis of oddity learning by rhesus monkeys. *Journal of Comparative Psychology, 48*, 188-194.

Neiworth, J. J., Hassett, J. M., & Sylvester, C. J. (2007). Face processing in humans and new world monkeys: the influence of experiential and ecological factors. *Animal Cognition, 10*(2), 125-134.

Neiworth, J. J., Parsons, R. R., & Hassett, J. M. (2004). A test of the generality of perceptually based categories found in infants: attentional differences toward natural kinds by New World monkeys. *Developmental Science, 7*(2), 185-193.

Noble, C. S., & Thomas, R. K. (1970). Oddity learning in the squirrel monkey. *Psychonomic Science, 19*(5), 305-307.

Noble, L. M., & Thomas, R. K. (1985). Oddity and dimension-abstracted oddity (DAO) in humans. *American Journal of Psychology, 98*(4), 549-557.

Overman, W. H., & Doty, R. W. (1982). Hemispheric specialization displayed by man but not macaques for analysis of faces. *Neuropsychologia, 20*, 113-128.

Parr, L. A., Dove, T., & Hopkins, W. D. (1998). Why faces may be special: Evidence of the inversion effect in chimpanzees. *Journal of Cognitive Neuroscience, 10*(5), 615-622.

Parr, L. A., Winslow, J. T., & Hopkins, W. D. (1999). Is the inversion effect in rhesus monkeys face-specific? *Animal Cognition, 2*(3), 123 - 129.

Parr, L. A., Winslow, J. T., Hopkins, W. D., & de Waal, F. B. M. (2000). Recognizing facial cues: Individual discrimination by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Journal of Comparative Psychology, 114*(1), 47-60.

Parron, C., Call, J., & Fagot, J. (2008). Behavioural responses to photographs by pictorially naïve baboons (*Papio anubis*), gorillas (*Gorilla gorilla*) and chimpanzees (*Pan troglodytes*). *Behavioural Processes, 78*, 351-357.

Pascalis, O., & Bachevalier, J. (1998). Face recognition in primates: A cross-species study. *Behavioural Processes, 43*, 87-96.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A., Potter, D. D., Broennimann, R., et al. (1988). Specialized face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies. *Behavioural Brain Research, 1988*(29), 3.

Perry, S. (1996). Intergroup encounters in wild white-faced capuchins (*Cebus capucinus*). *International Journal of Primatology, 17*(3), 309-330.

Perry, S., Barrett, H. C., & Manson, J. H. (2004). White-faced capuchin monkeys show triadic awareness in their choice of allies. *Animal Behaviour, 67*, 165-170.

Phillips, K. A. (1996). Natural conceptual behavior in squirrel monkeys (*Saimiri sciureus*): An experimental investigation. *Primates, 37*(3), 327-332.

Pineda, J. A., Sebestyen, G., & Nava, C. (1994). Face recognition as a function of social attention in non-human primates: an ERP study. *Cognitive Brain Research, 2*, 1-12.

Plimpton, E. H., Swartz, K. B., & Rosenblum, L. A. (1981). Responses of juvenile bonnet macaques to social stimuli presented through color videotapes. *Developmental Psychobiology, 14*(2), 109-115.

Pokorny, J. J., & de Waal, F. B. M. (in press). Face recognition in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology.*

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience, 16*(16), 5205-5215.

Roberts, W. A., & Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior Processes, 14*(3), 247-260.

Roitberg, E., & Franz, H. (2004). Oddity learning by African dwarf goats (*Capra hircus*). *Animal Cognition, 7*, 61-67.

Rose, L. M. (1997). Vertebrate predation and food-sharing in *Cebus* and *Pan*. *International Journal of Primatology, 18*(5), 727-765.

Rosenfeld, S. A., & van Hoesen, G. W. (1979). Face recognition in the rhesus monkey. *Neuropsychologia, 17*, 503-509.

Rumbaugh, D. M. (1970). Learning skills of anthropoids. In L. A. Rosenblum (Ed.), *Primate Behavior: Developments in Field and Laboratory Research* (pp. 1-70). New York: Academic Press.

Rumbaugh, D. M. (1971). Evidence of qualitative differences in learning processes among primates. *Journal of Comparative and Physiological Psychology, 76*, 250-255.

Rumbaugh, D. M., & Pate, J. L. (1984). The evolution of cognition in primates: A comparative perspective. In H. L. Roitblat, T. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 569-587). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sackett, G. P. (1965). Responses of rhesus monkeys to social stimulation presented by means of colored slides. *Perceptual and Motor Skills, 20*, 1027-1028.

Sackett, G. P. (1966). Monkeys reared in isolation with pictures as visual input: Evidence for an innate releasing mechanism. *Science, 154*(3755), 1468-1473.

Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *Journal of Experimental Psychology: General, 111*(4), 369-389.

Santos, L. R., Hauser, M. D., & Spelke, E. S. (2001). Recognition and categorization of biologically significant objects by rhesus monkeys (*Macaca mulatta*): the domain of food. *Cognition, 82*, 127-155.

Schrier, A. M., Angarella, R., & Povar, M. L. (1984). Studies of concept formation by stumptailed monkeys: concepts humans, monkeys, and letter *A*. *Journal of Experimental Psychology: Animal Behavior 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. *Journal of Experimental Psychology: Animal Behavior Processes, 13*(2), 136-143.

Shettleworth, S. J. (1998). *Cognition, Evolution, and Behavior*. New York: Oxford University Press.

Sidman, M. (1994). *Equivalence Relations and Behavior: A Research Story*. Boston, MA: Author's Cooperative.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5-22.

Silk, J. B. (1999). Male bonnet macaques use information about third-party rank relationships to recruit allies. *Animal Behaviour, 58*(1), 45-51.

Spironello, W. R. (2001). The brown capuchin monkey (Cebus apella): Ecology and home range requirements in central Amazonia. In R. Bierregaard, C. Gascon, T. E. Lovejoy & R. Mesquita (Eds.), *Lessons from Amazonia: The ecology and conservation of a fragmented forest* (pp. 271-284). New Haven, CT: Yale University Press.

Strong, P. N., Jr., Drash, P., & Hedges, M. (1968). Solution of dimension abstracted oddity as a function of species, experience, and intelligence. *Psychonomic Science, 11*(9), 337-338.

Strong, P. N., Jr., & Hedges, M. (1966). Comparative studies in simple oddity learning: I. Cats, raccoons, monkeys, and chimpanzees. *Psychonomic Science, 5*(1), 13-14.

Swartz, K. B. (1983). Species discrimination in infant pigtail macaques with pictorial stimuli. *Developmental Psychobiology, 16*(3), 219-231.

Tanaka, M. (1995). Object sorting in chimpanzees (Pan troglodytes): Classification based on physical identity, complementarity, and familiarity. *Journal of Comparative Psychology, 109*(2), 151-161.

Tanaka, M. (2001). Discrimination and categorization of photographs of natural objects by chimpanzees (*Pan troglodytes*). *Animal Cognition, 4*, 201-211.

Thomas, R. K. (1996). Investigating cognitive abilities in animals: Unrealized potential. *Cognitive Brain Research, 3*, 157-166.

Thomas, R. K., & Boyd, M. G. (1973). A comparison of *Cebus albifrons* and *Saimiri sciureus* on oddity performance. *Animal Learning & Behavior, 1*(2), 151-153.

Thomas, R. K., & Frost, T. (1983). Oddity and dimension-abstracted oddity (DAO) in squirrel monkeys. *American Journal of Psychology, 96*(1), 51-64.

Tomasello, M., & Call, J. (1997). *Primate Cognition*. New York: Oxford University Press.

Tomonaga, M. (1994). How laboratory-raised Japanese monkeys (*Macaca fuscata*) perceive rotated photographs of monkeys: Evidence for an inversion effect in face perception. *Primates, 35*(2), 155-165.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471-491.

van Hooff, J. A. R. A. M. (1962). Facial expressions in higher primates. *Symposia of the Zoological Society of London, 8*, 97-125.

van Hooff, J. A. R. A. M. (1967). The Facial Displays of the Catarrhine Monkeys and Apes. In D. Morris (Ed.), *Primate Ethology* (pp. 7-68). London: Weidenfeld & Nicolson.

Vauclair, J. (2002). Categorization and conceptual behavior in nonhuman primates. In M. Bekoff, C. Allen & G. Burghardt (Eds.), *The Cognitive Animal* (pp. 239-245). Cambridge, MA: MIT Press.

Vauclair, J., & Fagot, J. (1996). Categorization of alphanumeric characters by Guinea baboons: within-and between-class stimulus comparison. *Current Psychology of Cognition, 15*, 449-462.

Visalberghi, E. (1997). Success and understanding in cognitive tasks: A comparison between *Cebus apella* and *Pan troglodytes*. *International Journal of Primatology, 18*(5), 811-830.

Visalberghi, E., & McGrew, W. C. (1997). *Cebus* meets *Pan*. *International Journal of Primatology, 18*(5), 677-681.

Vonk, J. (2002). Can orangutans and gorillas acquire concepts for social relationships? *International Journal of Comparative Psychology, 15*, 257-277.

Vonk, J., & MacDonald, S. E. (2002). Natural concepts in a juvenile gorilla (*Gorilla gorilla gorilla*) at three levels of abstraction. *Journal of the Experimental Analysis of Behavior, 78*(3), 315-332.

Waitt, C., Gerald, M. S., Little, A., C., & Kraiselburd, E. (2006). Selective attention toward secondary sexual color in male rhesus macaques. *American Journal of Primatology, 68*(7), 738-744.

Waitt, C., Little, A., C., Wolfensohn, S., Honess, P., Brown, A. P., Buchanan-Smith, H. M., et al. (2003). Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. *Proceedings of the Royal Society of London B: Biological Sciences, 270*(Suppl 2), S144-S146.

Wasserman, E. A., & Zentall, T. R. (Eds.). (2006). *Comparative Cognition: Experimental Explorations of Animal Intelligence*. New York: Oxford University Press.

Wolin, L. R., & Massopust, L. C., Jr. (1972). A Comparison of performance of cebus and rhesus monkeys on a battery of behavioral tests. *Primates, 13*(1), 71-76.

Wright, A. A., & Delius, J. D. (1994). Scratch and match: Pigeons learn matching and oddity with gravel stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 20*(1), 108-112.

Wright, A. A., & Delius, J. D. (2005). Learning processes in matching and oddity: the oddity preference effect and sample reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes, 31*(4), 425-432.

Yoshikubo, S. (1985). Species discrimination and concept formation by rhesus monkeys (*Macaca mulatta*). *Primates, 26*(3), 285-299.

Zayan, R., & Vauclair, J. (1998). Categories as paradigms for comparative cognition. *Behavioural Processes, 42*, 87-99.

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*(3), 237-248.

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K. R., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews, 3*, 13-45.

Appendix A

Results from Mixed Sessions

Initially, we presented both the In-group Odd and out-group Odd condition mixed within a session. Two subjects (Bias and Mason) completed 19 sessions while one subject (Wilma) completed 13 sessions. Performance on mixed sessions was less than expected as the percent correct was low relative to previous tasks these subjects have performed (In-group Odd: 33.28%, Out-group Odd: 33.85%, Figure 14) and was consistent across subjects (Table 30). A Heterogeneity G-test showed that as a group, performance was in fact above chance levels (In-group Odd: Gh = NS, Gp = 42.37, $p < 0.001$, two-tailed; Out-group Odd: Gh = NS, Gp = 45.45, $p < 0.001$, two-tailed) though performance was lower than previously demonstrated (Pokorny & de Waal, in press).

Figure 14. Group performance on initial sessions where In-group Odd and Out-group Odd conditions were both presented within sessions (mixed sessions). Session number is along the x-axis and the percent correct is along the y-axis. In-group Odd trials are indicated by the dark diamonds while the Out-group Odd trials are signified by the lighter squares. The horizontal dotted line indicates chance level (25%).

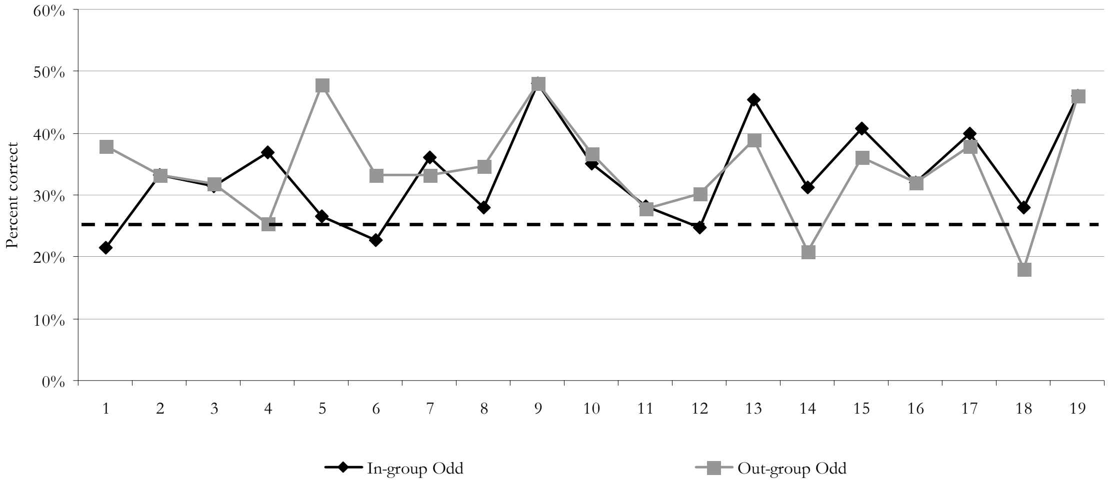


Table 30: Individual's correct performance on Group Membership trials when In-group Odd and Out-group Odd conditions were both presented within sessions, 13 (Wilma) or 19 sessions (Bias and Mason) per subject

	In-group Odd	Out-group Odd
Subject	*Mean*	*Mean*
Bias	32.7%	35.3%
Mason	32.7%	30.5%
Wilma	34.9%	36.8%

Given the less than expected performance, we discontinued mixed sessions and presented only one condition in each session, performing one condition type over several consecutive sessions before switching to the other condition.