Monkeys recognize the faces of group mates in photographs

Jennifer J. Pokorny1 and Frans B. M. de Waal1

Living Links, Yerkes National Primate Research Center, and Department of Psychology, Emory University, Atlanta, GA 30322

Contributed by Frans B. M. de Waal, October 22, 2009 (sent for review July 30, 2009)

Nonhuman primates possess a highly developed capacity for face recognition, which resembles the human capacity both cognitively and neurologically. Face recognition is typically tested by having subjects compare facial images, whereas there has been virtually no attention to how they connect these images to reality. Can nonhuman primates recognize familiar individuals in photographs? Such facial identification was examined in brown or tufted capuchin monkeys (Cebus apella), a New World primate, by letting subjects categorize facial images of conspecifics as either belonging to the in-group or out-group. After training on an oddity task with four images on a touch screen, subjects correctly identified one in-group member as odd among three out-group members, and vice versa. They generalized this knowledge to both new images of the same individuals and images of juveniles never presented before, thus suggesting facial identification based on real-life experience with the depicted individuals. This ability was unexplained by potential color cues because the same results were obtained with grayscale images. These tests demonstrate that capuchin monkeys, like humans, recognize whom they see in a picture.

face recognition | individual discrimination | oddity | primate | visual discrimination

Whereas human face recognition is well studied, we are only beginning to understand the extent of this capacity in other primates. Nonhuman primate face recognition seems highly accurate (1–6), sensitive to the configuration of facial elements as reflected in the so-called “inversion effect” (7–14) and “Thatcher effect” (15), and dependent on similar neural substrates as in humans (16–21). In one study, chimpanzees matched faces of unfamiliar individuals on the basis of kin resemblance (22), and in another they matched faces with a different bodily view of the same individual provided they were familiar with the depicted individual (23). The latter capacity has also been suggested for macaques (24), but absent rigorous methodology and successful replication, the data remain inconclusive. Together with cross-modal identity matching of conspecifics (24–27), the above results offer a first hint that nonhuman primates not only discriminate faces but connect two-dimensional facial representations with actual individuals that they know, the way we recognize whom we see in a photograph.

There is debate as to how exactly animals perceive pictorial stimuli (28–30). For monkeys, similar to humans, the eyes are the most salient feature of faces (31–33), and some studies report socially appropriate behavior, such as lip smacking, toward facial images (8, 34). Chimpanzees, furthermore, show differential heart rate responses (35) and are able to correctly assign names to depictions of known individuals (36). Evidence for such facial identification is much scarcer for monkeys, however, and to our knowledge has never been tested with faces per se. Monkeys are often assumed to process faces as complex stimuli that they merely match and compare, whereas individual identification is seen as part and parcel of human face recognition. This issue was addressed here by testing whether brown capuchin monkeys can apply real-life social knowledge to facial images of group members vs. outsiders.

In addition, this study addresses a larger issue of nonhuman primate understanding of the representational nature of two-dimensional images. Researchers use pictures for the amount of control it offers over the stimulus. Although widely used, there is debate as to what animals truly understand about pictures in general (29, 30), and specifically with respect to face perception (37). Fagot et al. (38) outline three ways in which animals can perceive a picture: (i) independence: there is no association between the image and the object it represents, so all that matters are features present in the image; (ii) confusion: the picture and the object remain undistinguished so that subjects respond to images the way in which they would respond to the real object, such as displaying an emotional reaction; or (iii) equivalence: subjects associate the picture with the object but do not confuse the two. A few studies have demonstrated picture equivalence in chimpanzees (36, 39) and capuchins (40), but for other nonhuman primates, results may fall under confusion (41, 42). If capuchin monkeys in the present study demonstrate differentiation between facial images of in-group and out-group conspecifics, as they do in real life (43–46), and fail to show confusion (e.g., hostile reactions to out-group members), we may conclude that this species interprets two-dimensional images as representing reality.

In-group/out-group distinctions are critical for the survival of many animals, because the out-group typically poses a threat to a group’s food and mating resources. Capuchins live in social groups of ~14–17 individuals and regularly come into both visual and physical contact with neighboring groups (43, 47). Both in captivity and in the wild, capuchin monkeys typically are hostile to outsiders (43–46). We do not know, however, what information they use to discriminate between in-group and out-group members. There is nothing visually apparent that would indicate an individual’s group membership. In humans, faces provide a rapid mechanism for determining the identity of individuals, as well as their age, sex, and emotional state. Because nonhuman primates, too, possess an extensive repertoire of facial expressions to communicate emotional states (48–51), the face is a highly salient stimulus that they probably also rely on for individual identification. Capuchin monkeys have presiously been shown to recognize faces (3) and to show the inversion effect (14) and hence seem an excellent candidate for tests of facial identification.

All subjects in this study live socially and therefore have experience interacting with a relatively small number of group members (~14 individuals over the course of testing). They were trained on an oddity task, which was used to let subjects select from among four two-dimensional photographic portraits: either one in-group face against three out-group faces or one out-group face against three in-group faces. Performance was assessed by

Author contributions: J.J.P. and F.B.M.d.W. designed research; J.J.P. performed research; J.J.P. and F.B.M.d.W. analyzed data; and J.J.P. and F.B.M.d.W. wrote the paper.

The authors declare no conflict of interest.

1To whom correspondence may be addressed: E-mail: jpkorn@emory.edu or dewaal@emory.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0912174106/DCSupplemental.
transferring subjects to a brand-new set of facial stimuli and comparing transfer performance with the 25% chance level.

If subjects perform above chance on this task, this strongly suggests that they classify facial images on the basis of their familiarity with and/or closeness to the depicted individuals, which task cannot be accomplished without facially identifying them. It is important, however, to rule out alternative explanations for positive performance. First, even though the transfer set contained new images, the individuals depicted were the same as those represented during training. Is it possible that subjects categorized the new images on the basis of previous associations learned during testing, thus making their choices independent of real-life experience with these individuals? To rule this out, subjects were presented with novel individuals whom they had never before seen in a two-dimensional image. Testing had been done using images of all available adult monkeys, but for this control test, images were introduced of in-group vs. out-group juveniles. Successful performance would confirm that subjects connect these images to real conspecifics with whom they are familiar.

A possible alternative explanation for positive transfer is that the two groups differ in hue—on average and perhaps only slightly—owing to the genetic influence of one or two breeding males. This might permit subjects to make an in-group vs. out-group distinction according to color cues rather than individual recognition. Therefore, subjects were presented with stimulus sets converted to grayscale, to remove color information. Successful performance with these stimulus sets would again confirm that subjects operate on the basis of individual recognition and are able to connect the facial images to actual conspecifics regardless of color information.

Results

Transfer Test. This task presented two conditions: (i) In-group Odd: three facial images represented individuals from the out-group, and one represented an individual from the in-group relative to the subject; and (ii) Out-group Odd: three images from the in-group and one from the out-group relative to the subject (Fig. 1). Conditions were presented in blocks, such that the In-group Odd condition was presented for several days, and then the Out-group Odd condition was presented for an equivalent number of days. Once subjects reached a performance criterion of 60% correct choices for two consecutive sessions on the given condition, they were transferred to a new set of stimuli, and performance was assessed on the first 40 transfer trials with this set.

As a group, the three subjects performed above 25% chance upon transfer under both conditions [In-group Odd: Gh = nonsignificant (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]. The significant heterogeneity in the Out-group Odd condition is due to one subject’s (Bias) poor performance [mean (M) = 39.4%] compared with the better performance of the other two, although even this one subject still performed significantly above chance (z = 1.67, P = 0.048, binomial test). 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 only Bias (McNemar test, n = 33, P = 0.049) showed a difference between conditions, performing better on the In-group Odd condition (In-group Odd: M = 68.4%; Out-group Odd: M = 39.4%; Mason: McNemar test, n = 40, P = 0.815; Wilma: McNemar test, n = 39, P = 1.000). See Fig. 2 for individual performances. Latencies were also analyzed and can be found in the supporting information (SI).

Juvenile Probe Trials. To determine whether the above performance could be explained by previous training and exposure to images of the same adult faces, even during transfer, subjects were presented with 40 probe trials containing novel images of juveniles within their normal testing sessions. They had never before been tested on images of these particular individuals. 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,

Fig. 1. Subjects need to select the odd facial image from among four. On this trial, the odd image is a member of group 1 (Top Left) compared with three members of group 2. For monkeys living in group 1 this trial represents the In-group Odd condition, but for those living in group 2 it is the Out-group Odd condition.

Fig. 2. Percentage correct choices during the last 40 trials of training and upon transfer to new images (first 40 trials) for both conditions (In-group Odd and Out-group Odd). Gray bars represent the mean performance, whereas each shape indicates the individual performance of three subjects. The horizontal dotted line specifies the chance level (25%).
df = 1, P < 0.001). The significant heterogeneity in the Out-group Odd condition was due to one subject’s (Mason) performance (M = 42.5%) being considerably lower than that of the other two, even though when tested individually this subject did perform significantly above chance (z = 2.26, P = 0.012, binomial test). We tested for performance differences between condition types and found that 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), whereas Mason showed the opposite result (McNemar test, n = 40, P = 0.049). Wilma (McNemar test, n = 40, P = 0.286) showed no difference. See Fig. 3 for individual performances on juvenile probe trials. Latency analyses can be found in SI Text.

Grayscale Original Probe Trials. Subjects were tested with the original stimulus set converted to grayscale to assess whether prior performance could have been due to color differences between the two groups. We report the results of measured hue differences between the two groups for the initial stimulus set and the transfer stimulus set in SI Text. Forty trials were presented as probes within normal testing sessions with the original colored stimulus set. As a group, performance was above chance in both conditions on the grayscale probe trials (In-group Odd: Gh = NS, Gp = 110.93, df = 1, P < 0.001; Out-group Odd: Gh = NS, Gp = 123.23, df = 1, P < 0.001). No subject performed differently according to condition type (Bias: McNemar test, n = 40, P = 0.238; Mason: McNemar test, n = 40, P = 0.286; Wilma: McNemar test, n = 40, P = 0.118). See also Fig. 4. SI Text contains the results of latency analyses.

Grayscale Transfer Probe Trials. To further rule out the possible role of color cues, subjects were presented with the transfer stimulus set converted to grayscale within their normal testing sessions. Subjects performed above chance in both conditions on the 40 grayscale transfer probe trials (In-group Odd: Gh = NS, Gp = 26.27, df = 1, P < 0.001; Out-group Odd: Gh = NS, Gp = 41.41, df = 1, P < 0.001). Again, subjects showed no difference depending on condition type (Bias: McNemar test, n = 40, P = 0.503; Mason: McNemar test, n = 40, P = 1.000; Wilma: McNemar test, n = 40, P = 0.286). See also Fig. 4. Latency analyses can be found in SI Text.

Because this stimulus set was the same as the set used in the transfer session but converted to grayscale, we were able to assess whether removing the color information impaired performance by comparing performance when the trials were initially presented in color (Transfer test) and then when they were presented here in grayscale. Results revealed that only Mason performed significantly poorer when the images were converted to grayscale than when initially presented in color [χ²(1, n = 80) = 4.06, P = 0.007].

Discussion

Socially living capuchin monkeys successfully selected the image of an individual who did not belong to the same social group as three other individuals depicted on the same computer screen. This required them to discriminate among similar-looking faces and apply their personal knowledge of group membership. Given their
high performance under a variety of experimental conditions, we conclude that capuchins can individually identify known conspecifics from two-dimensional images of their faces, similar to how humans identify friends and family in photographs.

Presenting the images in grayscale ruled out that slight color differences among groups would take into account individual contributions to the subjects’ ability to categorize the facial images as in-group vs. out-group. This strengthens the claim that subjects were selecting the odd individual on the basis of personal knowledge obtained through interactions with in-group members rather than according to visual cues. Our experiment also ruled out that knowledge of group membership was built up during the training and testing itself, because when subjects were presented with a brand new set of images of individuals never presented before (i.e., juveniles born in both groups), they still performed above chance on the task.

Group membership can be known only through personal experience, because it has no obvious facial correlates as opposed to other dimensions, such as sex or age (52–55). Our results strongly suggest a role of personal experience, thus ascribing the outcome to the same learned associations thought to underlie performance in cross-modal studies in which subjects need to match the voice and face of familiar individuals (24–26), which also requires knowledge of familiar individuals.

Many researchers use two-dimensional images for the control task, as there are no obvious facial correlates. The underlying assumption is that subjects understand what the image represents. Direct tests of picture–item equivalence have typically been performed using tasks such as food vs. nonfood items (41, 42), and in some cases, primates were found to confuse a food image with the real object. In the present study, we did not observe subjects reacting socially to the images, which would suggest confusion. Furthermore, the task required subjects to touch the face image, a behavior not shown toward live conspecifics, further ruling out confusion. We also can rule out independence, whereby animals do not have any association between the image and the object, because there is nothing visibly different about the images that would indicate group membership and allow them to categorize the images appropriately. Given that our subjects successfully categorized the visually diverse facial images as either in-group or out-group, corresponding to their experience in daily life, and that we ruled out alternatives for how they could have perceived the images, we conclude that capuchins understand the representational nature of two-dimensional facial images. They process images of faces not just as complex stimuli but in the same way we do, as faces of individuals they know or do not know.

Materials and Methods

Subjects. Subjects were three adult brown capuchin monkeys, aged 8 through 19 years, from two separately housed social groups at Yerkes National Primate Research Center. All three subjects (Bias, Mason, and Wilma) participated in earlier studies on face recognition (3). Training and testing began in September 2006, at which time both groups had 15 individuals, and lasted until early 2008, when images were related according to group membership (e.g., all three were from group 1), and one was from the other group (e.g., group 2), this being the correct choice. Typical trials consisted of presenting subjects with either (i) three different in-group individuals and one out-group individual (Out-group Odd), or (ii) three different out-group individuals and one in-group individual (In-group Odd; Fig. 2). The two trial types were presented in blocked sessions such that the Out-group Odd condition was presented for three consecutive days before switching to the In-group Odd condition for an equivalent number of days. The layout and 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 subjects received a food reward via the pellet dispenser. When the incorrect image was selected, all images disappeared, a low tone was played, and four seconds was added to the intertrial interval. Subjects were given 30 s to make their selection. If the trial ended without a response, all images reappeared and were not included in the data analysis. A correction procedure was used 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, whichever occurred first. Only the first presentation was included in data analysis.

Procedure. This experiment used an oddity paradigm. Subjects were to start the trial by making contact with a colored square located at the center of the screen. After starting the trial, the center square was removed and four images appeared simultaneously on the screen in either a diamond or square layout. Each trial consisted of three images related according to group membership (e.g., all three were from group 1), and one was from the other group (e.g., group 2), this being the correct choice. Typical trials consisted of presenting subjects with either (i) three different in-group individuals and one out-group individual (Out-group Odd), or (ii) three different out-group individuals and one in-group individual (In-group Odd; Fig. 2). The two trial types were presented in blocked sessions such that the Out-group Odd condition was presented for three consecutive days before switching to the In-group Odd condition for an equivalent number of days. The layout and 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 subjects received a food reward via the pellet dispenser. When the incorrect image was selected, all images disappeared, a low tone was played, and four seconds was added to the intertrial interval. Subjects were given 30 s to make their selection. If the trial ended without a response, all images reappeared and were not included in the data analysis. A correction procedure was used 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, whichever occurred first. Only the first presentation was included in data analysis.

Transfer Test. After training on the group membership oddity (subjects attained a performance criterion of 60% correct for two consecutive sessions on the given condition type), subjects were presented with a transfer test using a new stimulus set. This set consisted of three new photographs taken of all adults and subadults in the two groups, prepared as was described for the initial stimulus set. Sessions included 40 transfer trials and 35 clip art trials, used as a baseline measure of attention.

Probe Tests: Juvenile, Grayscale Original, and Grayscale Transfer. After the transfer test, three more experiments were conducted. The first one (Juvenile) used a new stimulus set consisting of four photographs of eight juveniles each, four from each group for a total of 32 images. The second test (Grayscale original) used the original stimulus set, but converted to grayscale. The last test (Grayscale transfer) used the transfer stimulus converted to grayscale. All experimental trials followed the same procedure, but no correction procedure was used. Instead, subjects were rewarded for any response on probe trials so we could assess performance without contamination of reinforcement history. Sessions consisted of 20 probe trials, 35 original trials (original stimulus set), and 20 clip art trials.

Data Collection and Analysis. The computer controlling the stimulus presentation also was responsible for data collection. Data recorded for each trial included subject, experimenter, date, type of test, trial condition (In-group Odd/Out-group Odd), trial number, the images presented, and where they were patched to the screen. After starting the trial, the center square was removed and four images appeared simultaneously on the screen in either a diamond or square layout. The layout and location of the odd stimulus was randomly chosen at the beginning of each trial. The latency to start the trial and make a response (in milliseconds), the image and location selected by the subject, and whether the trial was correct, incorrect, or aborted. Data were analyzed by using SPSS 16.0. To assess performance above chance level (25%), heterogeneity G tests were conducted. Heterogeneity G tests compare performance with random chance, similar to a t², but the G test takes into account individual contributions. Reported results of the G test are Gh, which measures whether the data are homogeneous, and Gp, which indicates whether the pooled data fit the expected ratio. A significant Gh indicates that data are not homogeneous and
deviations from our expectation may be in different directions (e.g., two subjects are well above chance whereas one is below chance). If Gp is significant, it indicates that the data are significantly different from the indicated chance level (25%).

ACKNOWLEDGMENTS. We thank Christine Webb, Kristin Leimgruber, Amanda Greenberg, Eva Kennedy, Charine Tabbah, Daniel Brubaker, Kari-anne Chung, and Tara McKenney for technical assistance; William Hopkins, Robert Hampton, Philippe Rochat, and Kim Wallen for helpful discussions; and the animal care and veterinary staff at the Yerkes National Primate Research Center (YNPRC) for the maintenance and care of our research subjects. Research was supported by Grant IOS-0718010 from the National Science Foundation (to F.B.M.D.W.) and by a base grant (RR-00165) from the National Institutes of Health to YNPRC. The YNPRC is fully accredited by the American Association for Accreditation for Laboratory Animal Care.


