CrossMark

ORIGINAL PAPER

Two-item same/different discrimination in rhesus monkeys (*Macaca mulatta*)

Benjamin M. Basile[1] · Emily J. Moylan[1] · David P. Charles[1] · Elisabeth A. Murray[1]

Abstract Almost all nonhuman animals can recognize when one item is the same as another item. It is less clear whether nonhuman animals possess abstract concepts of "same" and "different" that can be divorced from perceptual similarity. Pigeons and monkeys show inconsistent performance, and often surprising difficulty, in laboratory tests of same/different learning that involve only two items. Previous results from tests using multi-item arrays suggest that nonhumans compute sameness along a continuous scale of perceptual variability, which would explain the difficulty of making two-item same/different judgments. Here, we provide evidence that rhesus monkeys can learn a two-item same/different discrimination similar to those on which monkeys and pigeons have previously failed. Monkeys' performance transferred to novel stimuli and was not affected by perceptual variations in stimulus size, rotation, view, or luminance. Success without the use of multi-item arrays, and the lack of effect of perceptual variability, suggests a computation of sameness that is more categorical, and perhaps more abstract, than previously thought.

Keywords Categorization · Concept learning · Entropy · Perceptual variability

✉ Benjamin M. Basile
  benjamin.basile@nih.gov

[1] Section on the Neurobiology of Learning and Memory, Laboratory of Neuropsychology, National Institute of Mental Health, NIH, Building 49, Room 1B80, 49 Convent Drive MSC 4415, Bethesda, MD 20892-4415, USA

Introduction

It is well established that many nonhuman animals can identify when one item is the same as another item. For example, monkeys can be trained to respond to item repetitions, such as in matching paradigms (e.g., Mishkin and Delacour 1975), and they also naturally treat repeated objects differently than novel ones, as demonstrated in preferential viewing paradigms (e.g., Nemanic et al. 2004). The philosopher W.V. Quine said that "In this behavioral sense it can be said equally of other animals that they have an innate standard of similarity too. It is part of our animal birthright. And, interestingly enough, it is characteristically animal in its lack of intellectual status" (1969, p. 10).

It is less clear to what degree nonhuman animals possess something akin to an abstract concept of sameness that can be divorced from perceptual features. For example, we cannot only identify similarity in perceptually similar things (e.g., two dogs are the same breed), but also perceptually dissimilar things (e.g., a bullet is like an arrow), and we can conceptualize similarity as an abstract concept that can have value independent of specific objects (e.g., sameness might be good or bad depending on the context). Evidence has accumulated over the past few decades that some nonhuman animal species can be trained to selectively respond to visual displays as a function of whether those displays are composed of objects that are the same or different (e.g., Blaisdell and Cook 2005; Brown and Sayde 2013; Castro et al. 2010; Flemming et al. 2007; Giurfa et al. 2001; Wasserman et al. 2001; Wright and Katz 2007; Young et al. 1997). Because performance was controlled by the sameness or differentness of the displays, and not the identity of the individual stimuli, this has been taken as evidence that these species possess an abstract concept of sameness. In dissent, and as explored in the discussion,

Penn et al. ([2008](#)) have noted the same/different discrimination tasks used with nonhumans still rely almost entirely on the perceptual similarity between and among display items. Further, they note that the pattern of performance displayed by most species, in which performance scales continuously with incremental changes in perceptual similarity, does not suggest categorical boundaries between "same" and "different," as might be expected if subjects did possess these abstract concepts.

In support of the position that nonhuman animals do not possess abstract concepts of "same" and "different" that are categorical, is the finding that many organisms have difficulty in laboratory tasks that require them to use the sameness or differentness of two items as a discriminative cue. The majority of laboratory research on same/different judgments by nonhuman animals has been performed in pigeons (see Wasserman and Young [2010](#) for a review). One of the central findings from this research is that pigeons usually perform very well when making a same/different judgment about a large array of items but very poorly when making the same judgment about two items. For example, Castro et al. ([2010](#)) trained pigeons to peck on a computer screen to choose either an array of 16 identical clipart images or an array of 16 different clipart images depending on the background color. Pigeons showed near-perfect transfer of performance to novel stimuli when the stimulus arrays remained large (baseline = 91 %; first session transfer = 89 %). In contrast, when the arrays were tested with just two items in each, performance dropped to just above chance on the first transfer session (59 %; chance = 50 %). Similarly, Young et al. ([1997](#)) found that pigeons' performance fell to chance levels with two-item arrays. Wasserman and Young ([2010](#)) also discuss unpublished data in which pigeons trained from the outset with two-item arrays failed to learn the same/different discrimination. Even in one of the most successful demonstrations in pigeons, the birds only showed partial transfer, performing significantly above chance but significantly below baseline performance (baseline = 79 %; novel color transfer = 64 %; chance = 50 %; Blaisdell and Cook [2005](#)).

Monkeys have also shown surprising difficulty in performing same/different discriminations with two items. Wasserman et al. ([2001](#)) tested baboons using large arrays similar to those used with pigeons and found that baboons' performance declined systematically as the size of the array decreased, with performance falling to chance when the arrays consisted of either 4 or 2 items. Flemming et al. ([2007](#)) found that rhesus monkeys initially failed to learn a two-item same/different judgment. However, the monkeys were able to learn the task following three training manipulations. First, monkeys were trained to make the discrimination using arrays of eight items each. Second, a discriminative cue—the background color of the screen—

was introduced. This cue signaled whether the arrays of same items or the array of different items would be rewarded. Third, the number of items in each array was gradually reduced in increments of two as the monkeys met criterion at each step. Following this scaffolding training, the monkeys were now able to successfully make same/different judgments about the two-item arrays. Two of the original five monkeys were then tested for transfer to novel images and demonstrated full transfer. Thus, this evidence suggests that training with multi-item arrays may be necessary for rhesus monkeys to initially learn two-item same/different discriminations, but continued experience with multi-item arrays is not necessary to maintain that behavior.

Nonhuman animals' difficulty in using sameness as a discriminative cue with two items, and their increasing ease of doing so as the number of items increases, has led to the theory that they compute sameness as a continuous function of perceptual variability rather than as categorical abstract concepts of "same" and "different." Young and Wasserman ([1997](#)) calculated an "entropy" value based on the number of different items in the multi-item arrays and found that this measure of perceptual variability did predict pigeons' performance. More recently, Young et al. ([2003](#), [2007](#)) proposed the finding differences model as a more developed model that measures perceptual variability as a function of both the number of different items in the display and the degree of perceptual difference between and among those items. Another model, the optimal observer model, has been used to predict performance of humans on arrays of one-dimensional stimuli, but has not yet been applied to the type of multi-dimensional stimuli used in the majority of experiments with nonhumans (van den Berg et al. [2012](#)). Note that others have challenged the validity of accounts based solely on entropy or perceptual variability (Katz et al. [2007](#)), and we consider this more in the Discussion.

One prediction of models that assume sameness judgments are based on a continuum of perceptual variability is that "any manipulation that affects the similarity of the presented items … is predicted to affect discriminative performance" (Wasserman and Young [2010](#), p. 18). Tests of this prediction have produced mixed results. When the similarity of the *different* items was increased by presenting them in the same color (Blaisdell and Cook [2005](#)) or by applying a Gaussian blur (Young et al. [2007](#)), pigeons' performance decreased. In contrast, decreasing the similarity of the *same* items by manipulating the rotation (Young and Wasserman [2001](#)) or size (unpublished data cited in Young et al. [2007](#)) of individual items did not substantially change pigeons' performance. Thus, not all manipulations of perceptual variability affect same/different performance, and the interpretation of these results is

further complicated by other methodological differences between studies. In other species, such as rhesus monkeys, the effect of most of these perceptual manipulations on same/different performance is currently unknown. This study addresses this gap by manipulating perceptual variability across several dimensions and thus serves as a test of the hypothesis that organisms compute sameness as a continuous measure of perceptual variability.

In the current study, we tested the ability of rhesus monkeys to acquire the abstract concept "same." Monkeys were trained to make same/different discriminations using two-item arrays (Fig. 1a), such as those that have been problematic in the past. Unlike a previous study with rhesus monkeys (Flemming et al. 2007), we did not provide initial training with multi-item arrays. On each trial, monkeys saw two pairs of images on a touchscreen and earned food by touching the pair for which the two items were the same. To encourage learning of the abstract concept "same," we took two main approaches. First, to minimize learning about particular items and to promote learning of the abstract concept, we selected same and different pairs from a large stimulus set. This is the same approach that has been used successfully in other settings to train monkeys to perform according to abstract rules or concepts (Katz et al. 2002; Mishkin and Delacour 1975). Second, monkeys were challenged with identifying like pairs of images across several perceptual manipulations. Specifically, we manipulated the size, rotation, and view of the stimuli within each pair (Fig. 1b). To perform the task, monkeys had to identify two items as same not only when they were perceptually identical (normal condition), but

also when they appeared in different degrees of rotation (rotation condition), different overall size (size condition), and from different vantage points (view condition). As a measure of learning, we tested the degree to which performance transferred to novel images.

Methods

Subjects

Thirteen rhesus macaques (*Macaca mulatta*) initially began this study, but six were removed early for failure to learn either the pretraining stages or the main task (see Discussion for details). Thus, seven adult males (mean age at start of testing = 4.2 years) served as subjects in the experiment. Monkeys were housed singly or in pairs, kept on a 12-hour light–dark cycle, and had visual and auditory contact with conspecifics. To ensure that monkeys worked consistently in the test apparatus, we controlled their access to food such that their weights were maintained at 85 % or more of their free-feeding weight. Water was available ad libitum in the home cages. Prior to this study, we shaped each monkey to use a touchscreen. All procedures were reviewed and approved by the NIMH Animal Care and Use Committee and complied with US law.

Apparatus

We tested each monkey in a sound-attenuating chamber equipped with a 15″ color LCD touch-sensitive screen
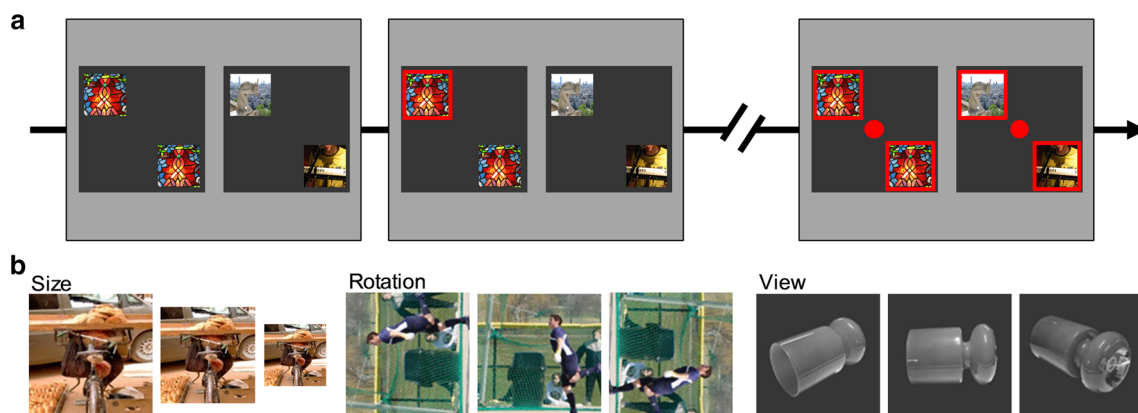

Fig. 1 Trial sequence and perceptual manipulations. **a** Example screens from the same/different discrimination. Monkeys saw two pairs of images, each pair contained within a *dark gray square* (first screen). To ensure that monkeys viewed and attended to each stimulus before making a choice, we required them to touch each image. After monkeys touched an image, a *red border* appeared around that image (second screen). Monkeys had to touch all four images in any order (not shown, as indicated by the gap). Once monkeys had touched all images, *two red dots* appeared at the center of each *dark gray square* (third screen). Monkeys earned food by touching the *red dot* between images that comprised the *same* pair. **b** There were four types of trials: normal, size, rotation, and view. Normal trials used identical images for the *same* pair, as shown in (**a**). For size trials, images were resized to 110, 85, or 55 pixels squared. For rotation trials, images were rotated around the center −90, 0, or 90°. For view trials, objects were photographed from −45, 0, and 45° around the *horizontal* axis. Objects for view trials appeared on a *dark gray* background that matched the *dark gray square* on the testing screen, effectively presenting the object alone without any border

(3M, Maplewood, MN) running at a resolution of 800 × 600 pixels and one automated food dispenser (BRS/LVE, Laurel, MD) that delivered nutritionally complete banana-flavored pellets (BioServ, Flemington, NJ) into a food cup mounted to the lower right of the screen. During testing, monkeys sat approximately 28 cm from the screen in a plexiglass primate chair that allowed them to reach out to touch the screen and to retrieve food. We used Gramalkn (Ryklin Software) running on Windows-based personal computers to control stimulus presentation, record responses to the touchscreen, and activate the food dispensers.

Stimuli

We used two types of stimuli. The first consisted of color photographs (110 × 110 pixels) gathered online and from personal collections (Fig. 1a). Different sets of images were used for pretraining and the main task. The photographs contained a wide variety of subject matter, but did not contain photos of other nonhuman primates. The second consisted of black and white photographs of objects (Fig. 1b, right). We selected objects that were irregularly shaped, such that they presented different outlines when viewed from different angles. Each object was photographed three times. Photos were taken from an angle of approximately 30° above the horizontal plane and from three different points around a circular periphery: −45°, 0°, and 45°. Objects were then separated from their background, desaturated, and pasted onto a dark gray background using Adobe Photoshop (Adobe Systems, San Jose, CA). The gray background of each object matched the gray background of the screen during testing, effectively presenting each object without background.

Procedure

The final same/different discrimination task, described in detail below, required monkeys to make observing responses to four stimuli and then select a red dot between the pair of stimuli that comprised the *same* pair (Fig. 1a). Because prior studies have reported that monkeys have difficulty learning similar tasks, we trained monkeys to perform this task through a series of five pretraining phases.

Pretraining 1

First, we trained monkeys to perform a standard single-pair visual discrimination. On each trial, two images (a train and a sailboat) appeared on the screen, one to the left and one to the right of center. One image was designated the S+ and the other the S−. Location of the S+ and S− was

pseudorandomized according to a Gellermann schedule (Gellermann 1933). A touch to either stimulus outlined it in red and produced a red response dot in the center of that image. A subsequent touch to that red response dot selected the image, whereas touches to the image without the red response dot produced no effect. In other words, the first stimulus touched was recorded as the choice on that trial, pending the monkey's completion of the touch requirements. Selection of the S+ was correct, causing the red response dot and the S− image to disappear. The red-outlined S+ then remained on the screen while the monkey was rewarded with one food pellet. Selection of the S− was incorrect, causing all stimuli to disappear from the screen. Trials were separated by a five-second interval following correct responses and a ten-second interval following incorrect responses. Touches to the screen during the intertrial interval reset the interval timer. We trained monkeys at the rate of one 60-trial session per day, 5 or 6 days per week. Each monkey progressed to the next phase if combined accuracy over three consecutive sessions was 90 % or above.

Pretraining 2

Phase 2 was run like Phase 1 with the exception that we added nineteen novel pairs of images, for twenty total pairs. Assignment of the novel images into pairs and designation of the S+ within each pair was randomized for each monkey. Each monkey progressed to the next phase if combined accuracy over five consecutive sessions was 90 % or above.

Pretraining 3

Phase 3 was run like Phase 2 with the exception that the red response dots now appeared next to the images. For each image, the associated response dot was located adjacent to the image on the side closest the center of the screen. To respond to a trial, the monkey had to touch the correct image, which outlined it in red and made the associated red dot appear, and then touch the red response dot next to that image. Each monkey progressed to the next phase if combined accuracy over three consecutive sessions was 90 % or above.

Pretraining 4

Phase 4 was identical to Phase 3 except that now we required the monkey to touch both images in either order, thus highlighting both images in a red border, before making a choice. The two red response dots did not appear until both images had been touched and outlined in red.

Pretraining 5

Phase 5 was identical to Phase 4 except that the number of discrimination pairs was increased from 20 to 40, by adding 20 novel pairs, and session length was increased from 60 trials to 120 trials.

Double touch training

During training, we noticed that a subset of monkeys was not visually attending to the stimuli. We therefore introduced a requirement that monkeys touch each image twice before the red outline appeared. To prevent further training issues, we introduced this requirement to all monkeys regardless of their phase of pretraining. In all tests after pretraining, valid responses to an image required two consecutive touches.

Same/different task

Now, for the first time, we presented monkeys with two pairs of images per trial. Each pair was presented within a dark gray square (Fig. 1a). The two images that comprised a pair appeared equally often in the four quadrants of the dark gray square. Figure 1a depicts a trial on which images from both pairs appeared in the upper-left and lower-right quadrants. For the *same* pair, the two images were the same photograph. For the *different* pair, the two images were different photographs. Within a trial, neither of the images in the *different* pair ever matched images in the *same* pair. The left/right location of the same and different pairs was pseudorandomized according to a Gellermann schedule (Gellermann 1933). To ensure that monkeys attended to each image, we required them to respond to each image by touching it on the screen, which led to the appearance of a red outline around that image. The four images could be touched in any order. Once the monkey had touched all four images, one red response dot appeared between each image pair at the center of the dark gray square. The correct response was to respond to the red dot located between the *same* images, and the incorrect response was to respond to the red dot located between the *different* images. As in Pretraining, correct responses were reinforced with food and were followed by a five-second intertrial interval, whereas incorrect responses were followed by a ten-second intertrial interval.

For the full task, sessions contained 96 trials each. Within each session, there were four trial types: normal, size, rotation, and view. In normal trials, images from the set of color photographs were seen in their original state (Fig. 1a). In size trials, the images were either large, medium, or small (110, 85, or 55 pixels squared, respectively; Fig. 1b, left). In rotation trials, the images were

either rotated left, upright, or rotated right ($-90°$, $0°$, or $90°$, respectively; Fig. 1b, center). In view trials, black and white images from the set of separately photographed objects were presented from the leftmost view, straight-on view, or rightmost view ($-45°$, $0°$, or $45°$, respectively; Fig. 1b, right). For the size, rotation, and view trial types, the two images that made up each pair were always from two different levels of the manipulation (e.g., on size trials the two images in each pair would always differ in size). Trials of each type appeared equally often in a pseudo-random order with the constraint that no more than two trials of each type could appear consecutively. We used a novel set of 256 stimuli for training (hereafter: Set A): 64 black and white objects for the view trials, and 192 color photographs randomly assigned to size, rotation, or normal trials. For stimuli used in the size, rotation, and view trials, we produced three versions of each stimulus, one at each level of perceptual manipulation (e.g., left view, center view, right view). The assignments were fixed for the duration of the experiment (e.g., images used in the size trial type occurred only on size trials). Stimuli could appear in either *same* or *different* conditions.

For the images used in each trial type, half were presented in a fixed set of pairs, such that a particular image in the *same* pair was repeatedly seen together with the same two *different* images (e.g., AA would always be seen together with BC). The other half were random, such that any image in the *same* pair could be seen with any two images in the *different* pair. This was done to encourage the monkeys to learn the same/different rule more quickly.

For monkeys whose accuracy failed to improve within a reasonable period of time, we instituted a correction procedure. Following an incorrect trial, the trial was repeated following the intertrial interval. Trials were repeated a maximum of three times. If the monkey chose the incorrect image pair on two consecutive retries, we ran a forced correction trial during which the trial was repeated one more time with the red response dot absent for the *different* pair. Thus, a monkey could experience a trial a total of four times, once as the original trial, twice as identical repetitions, and once more as a forced-choice correction. For data analysis, only the first attempt at each trial was analyzed.

Once monkeys had learned the same/different discrimination to an accuracy of at least 85 % correct overall and at least 70 % correct for each trial type, we assessed their ability to apply the same/different discrimination rule with novel images.

Transfer test 1

We tested transfer of the same/different rule to a new set of 256 stimuli (hereafter: Set B), which was constructed

similarly to Set A. Testing progressed as in the main same/different task with two exceptions. First, we eliminated fixed trials on which pairs of *same* and *different* images always appeared together. For the images used in each trial type, any image in the *same* pair could appear with any two images in the *different* pair. Second, instead of an accuracy criterion, we tested monkeys for a minimum of 15 sessions. For select monkeys whose accuracy was still low after 15 sessions, we allowed up to 20 more sessions of experience (median total sessions = 21.7). However, for the assessment of transfer, we only considered accuracy from the first session with Set B.

Transfer test 2

We tested transfer to a new set of 256 stimuli (hereafter: Set C), which was constructed similarly to Sets A and B. Testing progressed as in Transfer test 1.

Data analysis

Accuracy was compared to chance using one-sample t tests and to accuracy in different transfer tests using two-sample repeated-measures t tests. How accuracy varied across all trial types or across all manipulation levels within a trial type was assessed with repeated-measures ANOVA and followed up when appropriate with Bonferroni-corrected pairwise comparisons. Alpha level for all tests was set at .05. Because proportion correct is a measure that suffers from a restricted range, we arcsine transformed all proportions prior to statistical analysis (Howell 1997). Due to a computer error, trial-by-trial data were not available for one monkey performing the transfer to image Set C. Thus, for image Set C, data on overall accuracy come from seven monkeys, whereas data as a function of stimulus type come from six monkeys. To guide our interpretation of the statistical results, we used the criteria suggested by Katz et al. (2007). We interpreted performance as full transfer if accuracy on the first session with new images did not differ from that on the last session with old images. We interpreted performance as partial transfer if accuracy on the first session with new images was significantly lower than that on the last session with old images, but was still significantly above chance. We interpreted performance as a failure of transfer if accuracy on the first session with new images was not significantly different from chance.

Results

Initial accuracy on the same/different task with Set A did not differ from chance (Fig. 2; $t_6 = -1.01$, $P = .35$), but significantly improved over training (Fig. 2; $t_6 = 15.98$,

$P < .001$). By design, accuracy was also significantly above chance in the last session before the first transfer test (Fig. 2; $t_6 = 16.70$, $P < .001$).

In the first session with images from Set B, transfer accuracy was significantly above chance (Fig. 2; $t_6 = 5.54$, $P = .001$), though was significantly worse than during the last session with Set A (Fig. 2; $t_6 = -7.60$, $P < .001$). Analysis of individual trial types revealed that monkeys performed at levels significantly above chance for the normal, size, and rotation manipulations (normal: mean = .71, $t_6 = 6.74$, $P = .001$; size: mean = .62, $t_6 = 5.48$, $P = .002$; rotation: mean = .68, $t_6 = 6.68$, $P = .001$). Monkeys failed to transfer performance with view images, as accuracy was not significantly different than chance (mean = .47, $t_6 = -.46$, $P = .66$). Thus, training with Set A led to a partial transfer of the rule to Set B.

In the first session with images from Set C, transfer accuracy was significantly above chance (Fig. 2; $t_6 = 10.99$, $P < .001$) and not significantly different from that during the last session with Set B (Fig. 2; $t_6 = -.98$, $P = .36$). Accuracy across the four trial types did not differ significantly, but did show a nonsignificant trend (mean proportion correct: normal = .80, size = .79, rotation = .84, view = .70; $F_{(3,15)} = 2.90$, $P = .07$). Visual inspection of the accuracy scores suggests that this trend was likely caused by lower performance on the view trials than on the other three trial types. Considered independently, accuracy did not differ from the last session with Set B to the first session of Set C for any of the four trial types (normal: $t_5 = -1.22$, $P = .28$; size: $t_5 = .29$, $P = .78$; rotation: $t_5 = .46$, $P = .67$; view: $t_5 = -.74$, $P = .49$), indicating full transfer for all trial types.
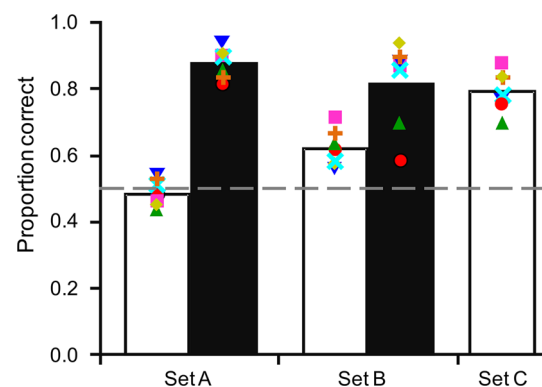

Fig. 2 Transfer of same/different discrimination rule across image sets. Proportion correct is depicted for the first and last sessions with the first two image sets, and the first session with the third image set. *Colored symbols* represent individual monkeys, *bars* represent group means for the first (*white*) and last (*black*) sessions with each set, and the *dashed gray line* represents chance

To evaluate the degree to which monkeys' performance was controlled by the variance of perceptual features, we analyzed accuracy for the first session with Set C, on the size, rotation, and view images as a function of perceptual disparity between the two images that made up the *same* option. If performance was based on one of these unidimensional perceptual features, accuracy should be higher when the manipulation of the two *same* images was more similar than when it was more disparate. For example, large and medium images should be more likely to be judged as the same than large and small images. This was not the case. Accuracy did not vary as a function of perceptual disparity for size or rotation manipulations (Fig. 3; size: $F_{(2,10)} = .62$, $P = .56$; rotation: $F_{(2,10)} = .55$, $P = .60$).

For the view manipulation, accuracy did vary with perceptual disparity (Fig. 3; $F_{(2,10)} = 4.79$, $P = .03$, partial $\eta^2 = .49$), but not in a pattern that was reliable or consistent with any theoretical interpretation. Pairwise comparisons, Bonferroni adjusted, revealed that accuracy on trials with the most disparate views, $-45°$ and $45°$, was significantly higher than on trials with views at $0°$ and $-45°$ (Fig. 3, two furthest right bars; $P = .028$). No other pairwise comparisons within the view manipulation reached significance (all $P > .38$). We believe this difference is a type I error for three reasons. First, no theoretical interpretation predicts higher judgments of similarity with more perceptually disparate images. Second, the difference
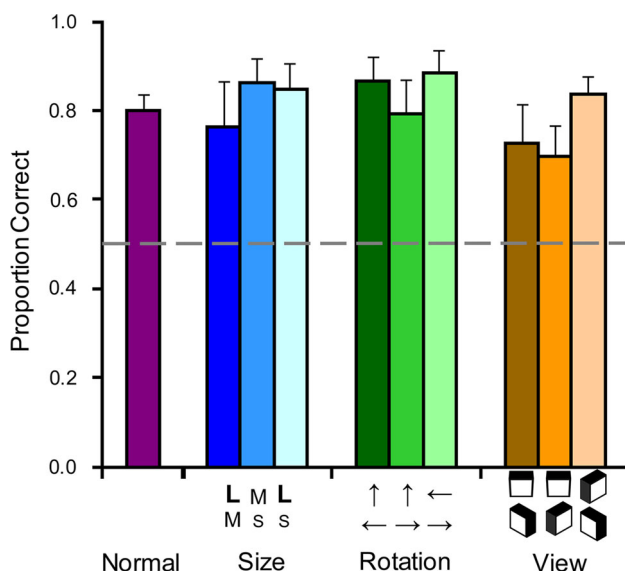

Fig. 3 Transfer as a function of perceptual manipulation. Mean proportion correct (+SEM) during the first session with Set C, broken down by trial type. For size, rotation, and view images, the two symbols below each *bar* represent the pair of same images. L: large; M: medium; S: small; ↑: upright; ←: rotated left; →: rotated right; ◗: leftmost view; ▢: straight-on view; and ◖: rightmost view. The *dashed gray line* represents chance

was not seen when comparing to the 0 and 45° pairs, which should have a perceptual disparity roughly similar to the 0 and $-45°$ pairs. Third, this difference did not persist into the second session ($F_{(2,10)} = -.32$, $P = .73$) or when all 15 sessions were considered together ($F_{(2,10)} = -.35$, $P = .71$).

Next, we analyzed the view images to determine whether performance was controlled by variability in luminance. On normal, rotation, and size trials, the two *same* images always had the same average luminance within the stimulus boundaries, whereas *different* images usually differed in luminance. It is possible that monkeys had learned to choose the pair with the same luminance. Our view manipulation allowed us to evaluate this hypothesis because the same object photographed from different angles often differed in luminance. For example, the left and center views of the view condition in Fig. 1b are more similar in luminance to each other than they are to the right view of the same object. Other objects, such as a stuffed panda or a white iron with a black bottom, differed substantially in luminance depending on view angle. To evaluate whether luminance similarity controlled behavior, we extracted each view image from its background and determined its mean luminance using the histogram function in the GIMP image editing program (www.gimp.org). We then divided trials in the first transfer session with Set C into trials in which the two *same* images were more similar in luminance than the two *different* images, and trials in which the two *same* images were more disparate in luminance than the two *different* images. Accuracy did not differ significantly as a function of whether the *same* images or the *different* images were more similar in luminance (mean proportion correct: same images more similar = .77, different images more similar = .75; $t_5 = .19$, $P = .86$).

Finally, we analyzed whether variation in perceptual disparity affected learning of the same/different rule in Set B prior to the final transfer to Set C. It is possible that more perceptually similar images were learned quicker than more perceptually disparate images (e.g., monkeys might be quicker to reach a high accuracy with large–medium pairs than large–small pairs). However, we found that the degree of perceptual disparity (e.g., large–medium, large–small, medium–small) did not interact with session (1–15) for size, rotation, or view trials (two-factor repeated-measures ANOVA interactions; size: $F_{(28,168)} = .89$, $P = .63$; rotation: $F_{(28,168)} = .83$, $P = .72$; view: $F_{(28,168)} = .85$, $P = .68$). Thus, we found no evidence to support the hypothesis that degree of perceptual variation affected learning. Given that the current study was designed to assess transfer and not learning rates, readers should interpret this null result with caution pending future studies.

Discussion

Our observations on perceptual distance and luminance strongly suggest that monkeys' same/different judgments were size, rotation, view, and luminance invariant. In contrast to previous reports (Flemming et al. 2007; Wasserman et al. 2001), monkeys learned a same/different discrimination with two-item arrays. Their success demonstrates that the use of differential cues or multi-item arrays is not necessary for learning. Together, this presents problems for models that posit judgments based on a continuous spectrum of perceptual variability (e.g., Young and Wasserman 1997; Young et al. 2007). It is possible that monkey's same/different judgments are more categorical than previously thought.

The progression from partial transfer to full transfer can be interpreted in two ways. First, it is possible that same/different learning progresses from concrete to abstract as the size of the stimulus set increased. On this view, after experience with Set A, monkeys had learned a rule that was semi-abstract but still heavily tied to the specific training stimuli. For example, they may have memorized through repeated exposure that certain combinations of stimuli were rewarded (e.g., a boot and a boot are rewarded, whereas a boot and a car are not rewarded). After expanding the training images to include Set B, their rule became more abstract (e.g., two same items are rewarded, whereas two different items are not rewarded). Alternatively, the full transfer might be related to the amount of experience with the abstract rule; in the current study, the increase in set size was confounded with an increase in experience. Yet another possibility is that monkeys had already learned an abstract same/different rule and partial transfer resulted from errors due to other reasons. Errors could result from surprise or distraction to novel stimuli, from confusion about whether the learned rule still applied in the new context, from a reduced ability to discriminate unfamiliar items, or from random chance producing a high proportion of stimuli in Set B that were difficult to discriminate. We cannot determine with certainty which interpretation is correct; however, the first interpretation is consistent with other research suggesting that rule learning progresses from concrete to abstract as one increases the size of the training set of images (e.g., Wright and Katz 2007).

One of the major debates in comparative studies of same/different learning is the degree to which nonhumans' judgments of sameness are rooted in perceptual similarity. Penn et al. (2008) propose a distinction among species, arguing that only humans can apply a concept of sameness to perceptually dissimilar items (e.g., a bullet is like an arrow), whereas nonhumans must rely on perceptual similarity. They suggest that nonhumans may solve same/

different discriminations, like the one presented in this study, by extracting a level of perceptual variability in each display and then applying a discrimination rule to choose whichever level of perceptual variability had previously been rewarded. This is similar to an explanation based on array entropy (Young and Wasserman 1997) or detection of accumulated differences (Young et al. 2007). Performance would transfer to novel stimuli because the cue controlling behavior, perceptual variability, was still present. In the current study, transfer accuracy was independent of disparities in rotation, size, view, and luminance. For the perceptual features we investigated, we found no evidence that the degree of perceptual variability controlled learning or generalization. It remains possible that behavior was controlled by perceptual variability, but that we did not identify the correct stimulus dimension that produced the variability. One fruitful avenue for future research would be to test transfer to novel perceptual variations after similar training (e.g., line drawings, scrambling, or desaturation). Note that we do not suggest that monkeys' judgments were divorced from the perceptual properties of the stimuli. This is clearly a task of perceptual similarity; two views of one object will be more similar than two views of two different objects. However, if monkeys' same/different judgments are based on perceptual variability, that variability is unlikely to be captured by unidimensional stimulus qualities.

It is possible that we did not observe an effect of perceptual variability because monkeys' same/different learning is more categorical than previously thought. If judgments are categorical, then the effect of perceptual variability will be minimal or nil if that variability does not cross category boundaries. Flemming et al. (2013) also suggest that monkeys' sameness judgments, while influenced by array entropy, do show evidence of a categorical boundary. The theory that nonhumans compute sameness along a continuous scale is based primarily on findings from laboratory tasks that used multi-item arrays. The use of multi-item arrays has been criticized as not actually testing same/different learning because "any species that is purported to be able to learn a S/D concept should be able to do it with two items" (Katz et al. 2007). It is possible that evidence for a continuous scale is the result of training with multi-item arrays, which might promote discrimination based on entropy or accumulated differences.

Nonhumans' seemingly inconsistent history of success in two-item same/different tasks may depend in part on deceptively minor methodological differences between studies. For example, although monkeys initially failed to learn a same/different rule when responding directly to one of the two pairs of items (Flemming et al. 2007), others have found that monkeys learn very well when they have to

make one of the two responses in the presence of a single pair of items (Katz et al. 2002). One proposed explanation is that monkeys experienced the two images successively in the latter method (Wasserman and Young 2010), as the second item only came up after responding to the first item. It is unclear why this would facilitate performance. Regardless, it would not explain success in the current study, as all items appeared at the same time. Another explanation is that vertical or horizontal presentation of the two images creates symmetry when the two items are the same (Wasserman and Young 2010) and that monkeys may find it easier to learn symmetry discriminations than same/different discriminations. Horizontal or vertical symmetry is unlikely to explain success in the current study, as image pairs were presented in a diagonal configuration as often as they were presented in vertical or horizontal configurations. Additionally, different views of the same object would likely influence a discrimination based purely on symmetry, but we observed no such effect. One notable difference is that both our methods and those of Katz et al. (2002) require monkeys to respond directly to the stimuli before making a same/different choice. In contrast, in the methods of Flemming et al. (2007) and Wasserman et al. (2001), monkeys were not required to respond to the stimuli before making their same–different choice. Requiring an observing response may make the relation between stimuli more apparent. Consistent with this, Katz et al. (2002) found that omitting the observing response hindered same–different learning. Another factor that might substantially affect the learned strategy is the presence of perceptual manipulations during training. It is possible that the inclusion of stimuli that were rotated, re-sized, or viewed from a different angle encouraged a discrimination that was based more on stimulus identity and less on the perceptual variability within the stimulus array. Future studies should directly examine how including such variability during training affects the strategy subjects learn.

It is also possible that inconsistent findings from pigeons and monkeys are due to species differences in the visual processing of stimuli used in same/different tasks. The last common ancestor of birds and mammals—an amniote—had no neocortex (Jarvis et al. 2005), and the visual systems of birds and mammals have evolved independently for over 300 million years (Benton et al. 2009). As a result, the pigeon and primate visual systems have different neural machinery to handle visual sensory processing (reviewed in Shimizu and Watanabe 2012). Although pigeon and primate vision share many basic properties, and pigeons process stimuli as objects in many, if not all, instances (Soto and Wasserman 2012), pigeons nevertheless sometimes show striking differences from primates. For example, pigeons routinely fail to mentally complete perceptually occluded objects, and they generally preferentially process stimuli

based on local rather than global cues (Fujita 2004; Qadri and Cook 2015). It is possible that these differences bias pigeons away from processing stimuli in same–different tasks as objects that are categorically same or different. How differences in visual systems affect an organism's computation of sameness requires more investigation.

One limitation of this study is that the discrimination task was not counterbalanced so that half the subjects were rewarded for choosing the *different* pair. This raises the possibility that the monkeys may have learned a strategy that is idiosyncratic to selecting the *same* pair rather than a true same/different rule. This remains to be tested in a future study; however, the success of Flemming et al. (2007) in training monkeys to conditionally respond to same or different two-item arrays, after multi-item array training, suggests that monkeys would likely be able to flexibly use either rule.

Another concern is that six of our initial thirteen monkeys were removed from this study prior to the transfer tests. One of the six monkeys was removed during Pre-training Phase 5 for failure to learn the S+/S− discriminations and thus was never assessed on the same–different test. Three of the six monkeys were reassigned to other studies early in training on the main test (mean completed sessions = 49.6) for reasons unrelated to the current study. The seven monkeys that completed the main test did so in a mean of 115.7 sessions, so the reassigned monkeys may eventually have learned the main test given a similar amount of training. The remaining two monkeys were removed after 190 and 130 sessions of the main task because their performance on the view trials was still at chance. These monkeys may have eventually learned with additional remedial training; however, view trials were the most difficult trials for all monkeys, and the failure of these two monkeys may indicate that these discriminations are near the limit of rhesus monkeys' capability.

If monkeys do indeed have the abstract concepts of "same" and "different," we might expect that those concepts could be used to guide their behavior. For example, just as monkeys can flexibly adjust their behavior when an object–reward association is reversed (Izquierdo and Murray 2007; Rudebeck and Murray 2008), they might also be able to respond flexibly to a reversal of concept–reward association. Similarly, they might also be able to form associations between concepts and behavioral responses or subjective values. The finding that monkeys could perform a conditional discrimination, in which the required response to *same* or *different* displays depended on the background color of the computer screen (Flemming et al. 2007), suggests that this may be the case.

In conclusion, rhesus monkeys can learn a simultaneous two-item same/different discrimination without prior training with multi-item arrays. Their same/different

judgments are independent of perceptual variances in size, rotation, view, or luminance. This stands in contrast to previous work with nonhumans, especially pigeons, and suggests a computation of sameness that is more categorical, and less continuous, than previously thought.

References

Benton MJ, Donoghue PCJ, Asher RJ (2009) Calibrating and constraining molecular clocks. In: Hedges SB, Kumar S (eds) The timetree of life. Oxford Univ Press, Oxford, pp 35–86

Blaisdell AP, Cook RG (2005) Two-item same-different concept learning in pigeons. Anim Learn Behav 33(1):67–77. doi:10.3758/BF03196051

Brown MF, Sayde JM (2013) Same/different discrimination by bumblebee colonies. Anim Cognit 16(1):117–125. doi:10.1007/s10071-012-0557-z

Castro L, Kennedy PL, Wasserman EA (2010) Conditional same-different discrimination by pigeons: acquisition and generalization to novel and few-item displays. J Exp Psychol Anim Behav Process 36(1):23–38. doi:10.1037/a0016326

Flemming TM, Beran MJ, Washburn DA (2007) Disconnect in concept learning by rhesus monkeys (*Macaca mulatta*): judgment of relations and relations-between-relations. J Exp Psych Anim Behav Process 33(1):55–63. doi:10.1037/0097-7403.33.1.55

Flemming TM, Thompson RKR, Fagot J (2013) Baboons, like humans, solve analogy by categorical abstraction of relations. Anim Cognit 16(3):519–524. doi:10.1007/s10071-013-0596-0

Fujita K (2004) How do nonhuman animals perceptually integrate figural fragments? Jpn Psychol Res 46(3):154–169. doi:10.1111/j.1468-5584.2004.00248.x

Gellermann LW (1933) Chance orders of alternating stimuli in visual discrimination experiments. Pedagog Semin J Genet Psychol 42(1):206–208. doi:10.1080/08856559.1933.10534237

Giurfa M, Zhang S, Jenett A, Menzel R, Srinivasan MV (2001) The concepts of 'sameness' and 'difference' in an insect. Nature 410(6831):930–933. doi:10.1038/35073582

Howell DC (1997) Statistical methods for psychology, 4th edn. Duxbury Press, Belmont

Izquierdo A, Murray EA (2007) Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. J Neurosci 27(5):1054–1062. doi:10.1523/jneurosci.3616-06.2007

Jarvis ED, Güntürkün O, Bruce L et al (2005) Avian brains and a new understanding of vertebrate brain evolution. Nat Rev Neurosci 6(2):151–159. doi:10.1038/nrn1606

Katz JS, Wright AA, Bachevalier J (2002) Mechanisms of same/different abstract-concept learning by rhesus monkeys (*Macaca mulatta*). J Exp Psych Anim Behav Process 28(4):358–368

Katz JS, Wright AA, Bodily KD (2007) Issues in the comparative cognition of abstract-concept learning. Comp Cognit Behav Rev 2:79–92

Mishkin M, Delacour J (1975) Analysis of short-term visual memory in monkeys. J Exp Psych Anim Behav Process 1(4):326–334. doi:10.1037/0097-7403.1.4.326

Nemanic S, Alvarado MC, Bachevalier J (2004) The hippocampal/parahippocampal regions and recognition memory: insights from visual paired comparison versus object-delayed nonmatching in monkeys. J Neurosci 24(8):2013–2026

Penn DC, Holyoak KJ, Povinelli DJ (2008) Darwin's mistake: explaining the discontinuity between human and nonhuman minds. Behav Brain Sci 31(02):109–130

Qadri MAJ, Cook RG (2015) Experimental divergences in the visual cognition of birds and mammals. Comp Cogn Behav Rev 10:73–105. doi:10.3819/ccbr.2015.100004

Quine WV (1969) Natural kinds. In: Rescher N (ed) Essays in honor of Carl G. Hempel: a tribute on the occasion of his sixty-fifth birthday. D. Reidel, Dordrecht, pp 5–23

Rudebeck PH, Murray EA (2008) Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. J Neurosci 28(33):8338–8343

Shimizu T, Watanabe S (2012) The avian visual system. In: Lazareva OF, Shimizu T, Wasserman E (eds) How animals see the world: comparative behavior, biology, and evolution of vision. Oxford University Press, New York, pp 473–482

Soto FA, Wasserman EA (2012) Visual object categorization in birds and primates: integrating behavioral, neurobiological, and computational evidence within a "general process" framework. Cog Affect Behav Neurosci 12(1):220–240. doi:10.3758/s13415-011-0070-x

van den Berg R, Vogel M, Josić K, Ma WJ (2012) Optimal inference of sameness. Proc Nat Acad Sci 109(8):3178–3183. doi:10.1073/pnas.1108790109

Wasserman EA, Young ME (2010) Same–different discrimination: the keel and backbone of thought and reasoning. J Exp Psychol Anim Behav Process 36(1):3–22. doi:10.1037/a0016327

Wasserman EA, Young ME, Fagot J (2001) Effects of number of items on the baboon's discrimination of same from different visual displays. Anim Cognit 4(3–4):163–170. doi:10.1007/s100710100095

Wright AA, Katz JS (2007) Generalization hypothesis of abstract-concept learning: learning strategies and related issues in *Macaca mulatta*, *Cebus apella*, and *Columba livia*. J Comp Psych 121(4):387–397

Young ME, Wasserman EA (1997) Entropy detection by pigeons: response to mixed visual displays after same–different discrimination training. J Exp Psychol Anim Behav Process 23(2):157–170. doi:10.1037/0097-7403.23.2.157

Young ME, Wasserman EA (2001) Evidence for a conceptual account of same-different discrimination learning in the pigeon. Psychon Bull Rev 8(4):677–684. doi:10.3758/BF03196203

Young ME, Wasserman EA, Garner KL (1997) Effects of number of items on the pigeon's discrimination of same from different visual displays. J Exp Psych Anim Behav Process 23(4):491–501. doi:10.1037/0097-7403.23.4.491

Young ME, Ellefson MR, Wasserman EA (2003) Toward a theory of variability discrimination: finding differences. Behav Process 62(1–3):145–155

Young ME, Wasserman EA, Ellefson MR (2007) A theory of variability discrimination: finding differences. Psychon Bull Rev 14(5):805–822. doi:10.3758/BF03194106