Curvilinear features are important for animate/inanimate categorization in macaques

Marissa Yetter*

Laboratory of Brain and Cognition, NIMH/NIH, Bethesda, MD, USA ✉

Sophia Robert*

Laboratory of Brain and Cognition, NIMH/NIH, Bethesda, MD, USA ✉

Grace Mammarella

Laboratory of Neuropsychology, NIMH/NIH, Bethesda, MD, USA ✉

Barry Richmond

Laboratory of Neuropsychology, NIMH/NIH, Bethesda, MD, USA ✉

Mark A. G. Eldridge

Laboratory of Neuropsychology, NIMH/NIH, Bethesda, MD, USA ✉

Leslie G. Ungerleider

Laboratory of Brain and Cognition, NIMH/NIH, Bethesda, MD, USA ✉

Xiaomin Yue

Laboratory of Brain and Cognition, NIMH/NIH, Bethesda, MD, USA ✉

The current experiment investigated the extent to which perceptual categorization of animacy (i.e., the ability to discriminate animate and inanimate objects) is facilitated by image-based features that distinguish the two object categories. We show that, with nominal training, naïve macaques could classify a trial-unique set of 1000 novel images with high accuracy. To test whether image-based features that naturally differ between animate and inanimate objects, such as curvilinear and rectilinear information, contribute to the monkeys' accuracy, we created synthetic images using an algorithm that distorted the global shape of the original animate/inanimate images while maintaining their intermediate features (Portilla & Simoncelli, 2000). Performance on the synthesized images was significantly above chance and was predicted by the amount of curvilinear information in the images. Our results demonstrate that, without training, macaques can use an intermediate image feature, curvilinearity, to facilitate their categorization of animate and inanimate objects.

Introduction

Primates can recognize objects with remarkable speed and accuracy—an ability that is crucial for avoiding predators, identifying food sources, and otherwise surviving in their natural habitat. Although seemingly effortless, decades of research in visual neuroscience and computer vision have shown that the ability to extract an object from a visual scene and categorize it is far from trivial (e.g., Pinto et al., 2008). The primate brain is equipped to deal with this computational problem by exploiting a vast array of features to classify objects into categories. Some distinctions are made based on knowledge or experience with the object, such as how it can be used (Bovet & Vauclair, 1998; Träuble & Pauen, 2007), whether it is threatening (Lipp, 2006; LoBue & DeLoache, 2011), or what contexts it is often found in (Blake et al., 2007; Kalénine et al., 2009; Kalénine et al., 2014), whereas others are determined based on the appearance of the object alone, by using its visual features such as color, size, global shape, and texture, etc.

The relative contribution of knowledge- and image-based information to object categorization varies across situations due to a number of factors. A crucial factor is the extent to which image-based features are predictive of a meaningful category or object class—a reasonable prerequisite for a visual system to rely on visual cues for object classification. Furthermore, the category or object class itself might influence the relative contribution of image information and prior experience needed to perform categorization. A long-standing line of research in evolutionary psychology has suggested that the primate visual system is highly tuned for the detection and recognition of animacy (Meyerhoff et al., 2014; Calvillo & Hawkins, 2016; Nairne et al., 2017; Long et al., 2019), even as early as 3 months old (Rakison, 2003; Heron-Delaney et al., 2011; Opfer & Gelman, 2011). A number of biological processes and key image feature differences have been proposed to explain how this discriminative ability might emerge so early in development. For example, some researchers have argued that innate processing biases interact with crude image-based biological templates to produce a sensitivity to faces from birth (Chiara et al., 2008; Sugita, 2008). Others have argued for a greater emphasis on the role of experience, through which persistent social exposure to faces early in life leads to a preference for face stimuli via more domain-general neural mechanisms (Srihasam et al., 2014; Livingstone et al., 2017). Yet another line of research has shown that human infants might develop concepts of animacy based on differences between biological and nonbiological motion (Mandler, 1992; Simion et al., 2008).

That the animate-inanimate distinction might be special to our visual system, and that these two categories differentially covary with a number of image features, suggests a plausible mechanism by which the primate visual system evolved to exploit image feature covariances to make animate-inanimate categorization judgments. One such feature is curvilinearity, or the extent to which the image of an object is composed of curved lines and textures. Animate objects tend to be more curvilinear than inanimate objects (Kurbat, 1997; Levin et al., 2001).

A recent study by Zachariou et al. (2018) demonstrated that, when deprived of global shape cues, humans were able to categorize animate and inanimate objects using just curvilinear information. Further, curvilinear information was positively correlated with performance on images of animate objects and negatively correlated with performance on inanimate objects. Given the lack of object shape information in the stimuli used and the lack of relationship between subjects' confidence ratings and their accuracy, it appears that this categorization ability is driven by an implicit, primarily bottom-up process.

If the human visual system can implicitly rely on curvilinear information to perform animate-inanimate categorization, it is possible that this may be a property of the primate visual system more broadly. To test this hypothesis, the current study sought to establish the contribution of image-based information to animate-inanimate categorization in a nonhuman primate, the rhesus macaque, by (1) testing the ability of macaques to categorize a large trial-unique set of animate and inanimate intact images that were unfamiliar to them, and (2) testing whether the macaques could use curvilinearity, without training, to categorize the objects when global shape information was removed.

Materials and methods

Subjects

Three male rhesus macaques (5–8 kg) were used in two behavioral experiments. All experimental procedures were approved by the National Institute of Mental Health Animal Care and Use Committee.

Visual stimuli

The first experiment included 500 images of animate objects and 500 images of inanimate objects, which were downloaded from open-source repositories on the internet. The animate images were comprised of mammals, birds, fish, reptiles, and insects (Figure 1a). The inanimate images included human-made objects, such as tools, vehicles, buildings, various household items, and natural objects, such as rocks and flowers (Figure 1b). All object images were digitally processed (see Supplementary Materials for a detailed description of this process) to match size, background, mean luminance, and root-mean-square (RMS) contrast. All images were resized to 200×200 pixels.

For the second experiment, we used an algorithm, described in detail in Portilla and Simoncelli (2000), to generate synthesized images of animate and inanimate objects (Figures 1c, 1d) that abolished the global shape of the original images but maintained their intermediate visual features (see Supplementary Materials). One thousand synthesized images were generated using the testing set of 500 animate and 500 inanimate intact images used in experiment 1.

Experimental procedures

The monkeys sat in a primate chair inside a darkened, sound-attenuated testing chamber. They
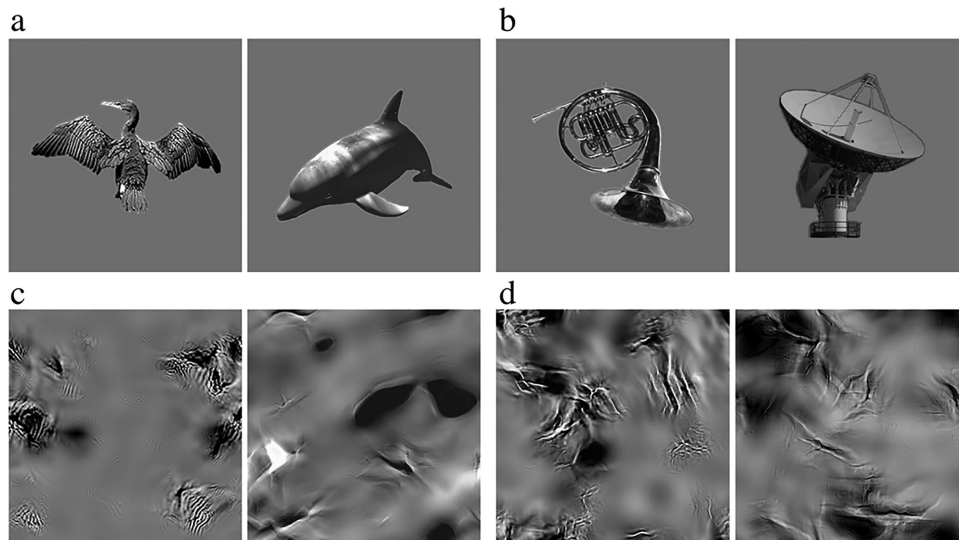
Figure 1. Examples of stimuli: (**a**) animate images; (**b**) inanimate images; (**c**) synthesized animate images; and (**d**) synthesized inanimate images.

were positioned 57 cm from a computer monitor (Samsung 2233RZ; Wang & Nikolic, 2011) subtending 40 degrees × 30 degrees of visual angle. The design and control of task timing and visual stimulus presentation were executed with networked computers running custom written (Real-time Experimentation and Control, REX; Hays et al., 1982) and commercially available (Presentation, Neurobehavioral Systems) software.

Training for experiment 1

Monkeys were initially trained to grasp and release a touch sensitive bar to earn water rewards. After this initial shaping, a red/green color discrimination task was introduced. Red/green trials began with a bar press, and 100 ms later a small red target square (0.5 degrees) was presented at the center of the display (over-laying a white noise background). Animals were required to continue grasping the touch bar until the color of the target square changed from red to green, this occurred randomly between 500 and 1500 ms after bar touch. Rewards were delivered if the bar was released between 200 and 1000 ms after the color change; releases occurring either before or after this epoch were counted as errors. All correct responses were followed by visual feedback (the target square color changed to blue) after bar release and reward was delivered between 200 and 400 ms after visual feedback. There was a 2 second inter-trial interval (ITI), regardless of the outcome of the previous trial.

After each monkey reached criterion in the red/green task (2 consecutive days with >85% correct performance), a visual categorization task was introduced. Each trial began when the animal grasped the touch bar. Next, an image (14 degrees × 14 degrees) appeared at the center of the screen, followed by a red cue over the center of the image. When the image presented was animate, the monkey had to release the bar before the red cue turned green to receive a liquid reward. When it was an inanimate trial, the monkey had to continue to hold the bar until the red cue turned green and then release the bar to receive a liquid reward (Figure 2). The red cue was displayed on the screen for 1 to 3 seconds before turning green in inanimate trials. If the monkey released the bar during the red target and an inanimate image was presented, no reward was delivered, and the image was displayed on the screen for a 4 to 6 second time out. If the monkey did not release the bar during the inanimate image presentation within 1000 ms after the red target turned green, no reward was delivered and there was a 3 second time out.

If an equal drop size was used as reward for both conditions, monkeys would tend to favor a release on red because of the delay discounting effect when waiting for green. Therefore, the number of reward drops delivered for correct responses to red or green was adjusted during the training phase to reduce the bias in responding to each category for each animal. As such, the drop ratio for correct animate versus correct inanimate trials was 1 to 7 for monkey 1 (M1), 1 to 6 for monkey 2 (M2), and 1 to 9 for monkey 3 (M3). Each monkey was trained on a repeated set of 20 animate and 20 inanimate images for several days until their choice accuracy reached above 85% accuracy for 2 consecutive days. The categorization accuracy in the last training day was 98% for M1, 96% for M2, and 88% for M3.
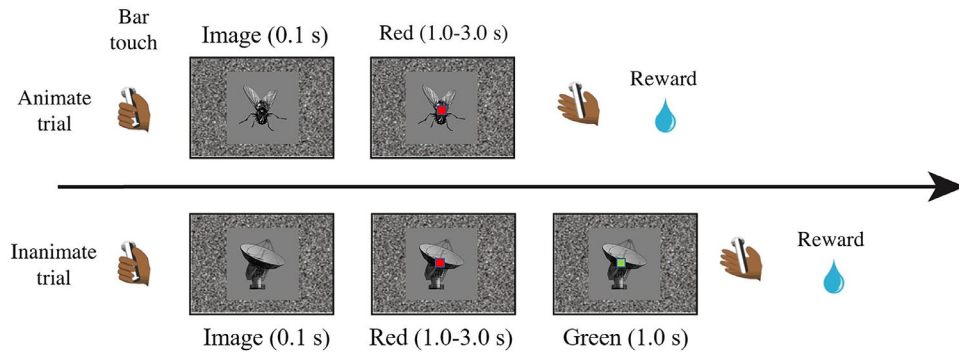
Figure 2. Experimental procedure. Each trial began when the animal grasped the touch bar. An image appeared at the center of the screen, followed by a red cue over the center of the image. When the image presented was animate, the monkey had to release the bar within 1 to 3 seconds of the appearance of the red cue to receive a liquid reward. When it was an inanimate trial, the monkey had to continue to hold the bar until the red cue turned to green (between 1 and 3 seconds after red cue onset) and then release the bar during the green cue to receive a liquid reward.

Testing for experiments 1 and 2

During the testing phase of experiment 1, monkeys were tested on trial-unique sets of 100 novel animate and 100 novel inanimate intact images for 3 (M1) or 5 days (M2 and M3). After the third testing day on classifying intact images into animate and inanimate categories, M1 reached an accuracy of 91%. Due to this clear demonstration of high performance categorizing intact images, we stopped testing M1 on intact images and moved onto testing classification of synthesized images. Crucially, the training images were never shown in the testing sets, and on each testing day, monkeys were presented with a new set of unfamiliar images. Immediately after experiment 1, monkeys were moved to experiment 2, in which they were tested on trial-unique sets of 100 synthesized animate and 100 synthesized inanimate images (see Figures 1c, 1d) for 5 days (M1, M2, and M3).

Classification analyses

The statistical significance of classification accuracy was evaluated for each monkey individually using a permutation test. For each monkey, we created a vector comprised of his responses on each trial (animate or inanimate), which we labeled as Vr, and an additional vector comprised of values representing the actual category of a trial (animate or inanimate), which we labeled as Vc. We then shuffled both the order of Vr and Vc. Then, for each row of the vectors, if the value in Vr matched that of Vc, we labeled that trial as correct and if not, as incorrect. Using this method, we calculated the overall accuracy (percentage correct irrespective of category), the accuracy for the animate category (percentage of animate trials correctly classified), and

the accuracy for the inanimate category (percentage of inanimate trials correctly classified). The shuffling procedure was repeated 10,000 times for each monkey and for each permutation, we recorded these 3 accuracy values. At the end of the 10,000 permutations, each monkey had its own chance distributions (with 10,000 data points each), representing overall accuracy. Using these chance distributions, we evaluated the significance of each monkey's actual mean classification accuracy. The permutation test was run for each monkey for each experiment separately.

Reaction time

Because the experiments used an asymmetric design, monkeys had more time to make a decision on inanimate trials, and less time on animate trials. As such, analysis of reaction time would not yield useful information on how monkeys performed the task. Therefore, reaction time was not analyzed and presented here.

Quantifying the amount of curvilinear and rectilinear information of the stimuli

We calculated the amount of curvilinear and rectilinear information present in each image using a method presented previously in Yue et al. (2014), Zachariou et al. (2018), and Yue et al. (2020).

After normalizing the mean luminance and RMS contrast of the stimuli, we calculated the amount of curvilinear and rectilinear information present in each image using curved Gabor filters developed by Krüger et al. (2001): these curved Gabor filters are a product of a rotated complex harmonic wave function and a

two-dimensional bent and rotated Gaussian function. It is formulated as follows:

$$B^{\vec{b}}(x, y) = r^{\vec{b}} * G^{\vec{b}}(x, y) * \left(F^{\vec{b}}(x, y) - DC^{\vec{b}} \right)$$

where $F^{\vec{b}}$ is the rotated complex harmonic wave function, $G^{\vec{b}}$ is the two-dimensional bent and rotated Gaussian function, and a vector $\vec{b}$ includes three variables: frequency, orientation, and level of curvature. The bank of the curved Gabor filters is composed of 120 individual curved Gabor filters, including three spatial scales (frequency), eight orientations, and five levels of curvature. Each stimulus image was resized to 256 × 256 pixels and processed using local input divisive normalization (Pinto et al., 2008). The images were then convolved with the bank of curved Gabor filters, which produced 120 (3 × 8 × 5) curved Gabor coefficients, each presented as an image. Each curved Gabor coefficient image, including both the complex and real components, represented the result from a curved Gabor filter with a unique combination of a spatial scale, an orientation, and a level of curvature. The magnitudes of each curved Gabor coefficient image on each pixel were calculated as the square root of the sum of squared coefficients of complex and real components. Then, the largest magnitude across all 120 curved Gabor coefficient images was extracted for each pixel to create a peak curved Gabor coefficient image. This step eliminated responses produced by nonoptimal curved Gabor filters, so that the peak curved Gabor coefficient image represented the optimal curved Gabor filter response to an image across scales, orientations, and levels of curvature.

The same procedure was repeated using the bank of rectilinear Gabor filters, composed of 3 spatial scales and 8 orientations, to generate 24 rectilinear Gabor coefficient images. Next, the largest magnitude across all 24 rectilinear Gabor coefficient images was extracted for each pixel to generate a peak rectilinear Gabor coefficient image. Then, the magnitude in the peak curved Gabor coefficient image was set to zero at a pixel if its magnitude was smaller than that in the peak rectilinear Gabor coefficient image in that pixel. The procedure went through all pixels to create a curved Gabor coefficient image with no rectilinear features represented, which we called a unique curved Gabor coefficient image. Finally, a curvilinear value of the stimulus image was produced by averaging the unique curved Gabor coefficient image across all pixels. The degree of curvature of our filters was essentially formalized by the result of a second polynomial function in two dimensions. In this framework, the sharpest tight curve would not break up to become a rectilinear corner and the shallowest curve would

approach a straight line, but not become a straight line as long as the coefficient of the second-degree term was non-zero.

Thus, at the end of this process, each image was assigned a curvilinear value. For each condition, a curvilinear value was produced by averaging all eight curvilinear values of the eight images in each condition. This curvilinear value represented the amount of curvature in that condition across scales, orientations, and levels of curvature. Similarly, the rectilinear values of each image were averaged to produce a rectilinear value for each condition, representing the residual intermediate feature information that was not captured by curvature information.

Predicting categorical membership of images using curvilinear and rectilinear information

We performed a logistic regression of curvilinear and rectilinear values on the image category. This logistic regression model allowed us to calculate a likelihood of an image belonging to the animate category with a given set of curvilinear and rectilinear values. With this procedure, we created an animacy probability map across curvilinear and rectilinear space to establish the relationship between image category membership and curvilinear and rectilinear values and plotted the raw values for all of the images as black circles (animate) and white crosses (inanimate) on top of the probability map.

To further illustrate the consistency in the relationship of curvilinear and rectilinear values with image categories between the raw data and logistic regression data, we created a map from the raw data using the following linear interpolation method procedure. Each image was represented as a point in three-dimensional space: category membership (1 or 0), curvilinear values, and rectilinear values. The categorical membership values between any two adjacent points in curvilinear and rectilinear space were linearly interpolated to fill the empty space between those two points and create a smooth surface plot. This process was repeated for all points. The categorical membership values between the raw data points are not meaningful but help illustrate the distribution of animate and inanimate images within the curvilinear and rectilinear space and its similarity to the logistic regression model results.

Logistic regression of monkeys' performance with trial numbers

As the monkeys were rewarded when they correctly performed the categorization in the testing phase of

experiments 1 and 2, their averaged performance likely resulted from both the use of features they learned from the training images to categorize animate and inanimate images and continuous learning during the testing phase. To determine the contribution of these two factors to the overall performance, we conducted a logistic regression on each monkey's performance using trial number as a regressor. Specifically, we regressed the monkey's response for each trial (either right or wrong) with the trial number, in which the trial number was treated as a continuous variable. The trials in which monkeys failed to respond were excluded from the analysis. In this model, a significantly positive non-zero intercept means that the ratio of performing right over wrong is substantially larger than 1, indicating that a monkey performed the task significantly above the chance at the beginning of the experiment. A significantly larger than zero slope means their performance continuously improved as the experiment proceeded.

Performance consistency analysis across monkeys

To assess the similarity of performance across the three monkeys, we assessed performance consistency for both experiment 1 and 2 with a Cronbach alpha test. First, all images sorted in ascending order of curvilinear values were grouped into 10 bins with an approximately equal number of images in each bin. Then, the categorization accuracy was calculated for each bin, which generated 10 categorization accuracies for a monkey. The same procedure was repeated for each monkey separately. We then used those categorization accuracies per monkey to compute the Cronbach alpha to evaluate the performance consistency across monkeys. Ten bins were chosen to contain enough trials within each bin to get reliable accuracy.

For experiment 1, the data analyzed were collected on day 1 to day 3 for M1, day 2 to day 5 for M2, and day 1 to day 5 for M3. To compare the performance consistency across monkeys, we used the data collected only on day 2 and day 3 in which the same sets of visual stimuli were presented to all monkeys. For experiment 2, we used data collected from all 5 days because all monkeys were examined over the whole testing phase.

Logistic regression of monkeys' performance with curvilinear and rectilinear values of visual stimuli

To determine whether and the extent to which the amount of intermediate image features (such as curvilinearity and rectilinearity) presented in experiments 1 and 2 contribute to monkeys'

performance, we conducted a logistic regression of monkeys' performance (right or wrong) with the curvilinear and rectilinear values of our visual stimuli (Yue et al., 2014; Zachariou et al., 2018). The trials in which monkeys failed to respond were excluded from the analysis.

The analysis was conducted at the group level to increase the signal-to-noise ratio using MATLAB (MathWorks, Inc.) with the following procedure. First, the performance from the three monkeys was concatenated to create a group response. Then curvilinear and rectilinear values for each stimulus were entered into the logistic regression model as two independent regressors. We included stimulus type (animate or inanimate) as a categorical variable in the logistic regression model to examine the interaction between the amount of intermediate image features and stimulus type on the monkeys' performance. As raw responses from each monkey were used, curvilinear and rectilinear values of a stimulus that more than one monkey responded to appeared more than once in the regression model.

To determine the contribution of the amount of intermediate visual features to the monkeys' performance, we fitted a logistic regression to raw responses, instead of average response accuracies per stimulus in a linear regression, for two reasons: (1) to avoid overestimating the influence of stimuli that only one monkey responded to, and (2) to avoid creating artificially continuous responses with averaging, because responses were discrete.

Deep convolutional neural network training and correlation analysis

The deep convolutional neural network (DCNN), AlexNet (Krizhevsky et al., 2012), was imported into MATLAB, and pretrained on the ImageNet database (Deng et al., 2009). All pretrained weights in the first 22 layers were kept the same, whereas the last 3 layers—fully connected layer, SoftMax layer, and classification layer—were trained to classify each intact image into animate or inanimate categories. The training was conducted on the 500 intact animate and 500 intact inanimate images used in experiment 1, using the stochastic gradient descent with momentum optimizer, minimum batch size 64, maximum epochs 20, and an initial learning rate of 10^{-4}. After 300 iterations, the neural network performance converged on an accuracy of 99.9%. Then, the trained neural network was tested to classify the same 1000 synthesized images used in experiment 2 into either the animate or inanimate category.

As the images that the monkeys skipped varied across individuals, it was not possible to conduct a correlation analysis of the monkeys' group performance

with DCNN performance at the individual image level. Thus, to compute the correlation of the DCNN classification accuracies and the monkeys' response accuracies to the synthesized images in experiment 2, we arranged the responses of the DCNN and each monkey according to the ascending order of curvilinear values of the synthesized images presented in each trial. The ordered responses were then grouped into 40 bins. The monkeys' accuracies used for the correlation analysis were averaged across all three animals. Next, the response accuracy for each bin was calculated for the DCNN and monkeys, resulting in 2 sets of 40 data points. The significance of the correlation was assessed by a permutation test (10,000 iterations).

Results

Experiment 1: Intact images

1) Overall classification accuracy for individual monkeys

During the testing phase of experiment 1, in which novel intact images were used for the categorization task, each image was presented only once regardless of the monkeys' responses. This eliminated the option of memorizing test images to perform the task. Across 5 days of testing, all monkeys performed the task significantly above chance (overall accuracy for M1 = 80.88%, $p < 0.0001$; M2 = 78.38%, $p < 0.0001$; and M3 = 76.95%, $p < 0.0001$). The statistical significance was determined by permutation test (see Methods). The overall response rate was 99.64% for M1, 73.43% for M2, and 98.86% for M3.

M2 performed significantly above chance level (50%) on the last day of training, with an accuracy of 96% ($p < 0.001$, permutation test). Meanwhile, his performance on the first half of day 1 testing was below chance (accuracy: 38.75%, $p < 0.0001$, permutation test) as shown in Supplementary Figure S1. This large drift in performance from training to the first day's testing, which was not observed in the other two monkeys, might be explained by a strategy in which M2 memorized the small number of the training images instead of learning a rule to perform the task during the training phase. Thus, in the first day of testing, M2 was learning the categorization task. After eliminating data from this day, overall performance was 85.64% ($p < 0.001$), and overall response rate was 73.3%. Unless stated otherwise, subsequent analyses used M2's testing data from day 2 to day 5 only. Data from all 5 days of testing are included in Supplementary Figure S2.

The data show that monkeys were able to successfully classify intact images that they had no previous experience with into animate and inanimate object categories, suggesting that image-based features distinguishing the two categories played a significant role in the monkeys' categorization performance.

2) Generalization and learning effect for individual monkeys

Because monkeys were given a liquid reward whenever they categorized images correctly in the testing phase, their overall performance could have resulted from continuously learning to categorize testing images as animate and inanimate due to reward feedback. In other words, significantly above-chance performance in the testing phase may not have captured the full picture of the monkeys' complex processing. Their performance could have more to do with this continuous feedback than with generalizing visual features learned during the training set to categorize the testing images. To separate the effect of generalization from the effect of learning during the testing phase, we performed a logistic regression (see Methods) on a single-trial basis to quantify the generalization as the intercept and learning as the slope of the regression model. We anticipated that, if there were a generalization effect, then the intercept of the logistic regression model would be significantly greater than zero, and if there were a learning effect, then the slope of the regression model would be significantly greater than zero.

Monkeys were able to use the information they learned during training to perform the categorization task on unfamiliar images at the onset of the testing phase, as shown in Table 1, where the intercept of the logistic regression is shown to be significantly above chance for all three monkeys. The slope of the logistic regression was positive and significantly different from zero in all monkeys, indicating that performance improved as testing progressed. All three monkeys' performances were significantly associated with trial number, as shown in Figure 3 and Table 1 (for M1: χ^2 [595] = 58.545, $p = 1.98 \times 10^{-14}$; M2: χ^2 [584] = 18.361, $p = 1.828 \times 10^{-5}$; and M3: χ^2 [986] = 13.252, $p = 2.72 \times 10^{-4}$), further indicating that monkeys continued to learn during the testing phase, improving their performance even though each image was presented only once.

Taken together, the significantly above-chance performance and significant generalization effect in categorizing the intact novel images suggests that all three monkeys learned to distinguish between the two categories during the training phase (M1 and M3) or after the first day of testing (M2), by generalizing the features learned from the small set of training images to the unfamiliar images in the larger testing set.

As the experiment used an asymmetric design, the performance improvement across trials could be driven by learning that occurred for either the animate

	Logistic regression	
Monkeys	Intercept	Slope
M1	$0.359 \ (p = 2.4 \times 10^{-2})$	$2.951 \times 10^{-2} \ (p = 5.241 \times 10^{-13})$
M2	$1.086 \ (p = 2.921 \times 10^{-9})$	$1.94 \times 10^{-2} \ (p = 2.746 \times 10^{-5})$
M3	$0.809 \ (p = 1.211 \times 10^{-10})$	$6.368 \times 10^{-3} \ (p = 2.969 \times 10^{-3})$

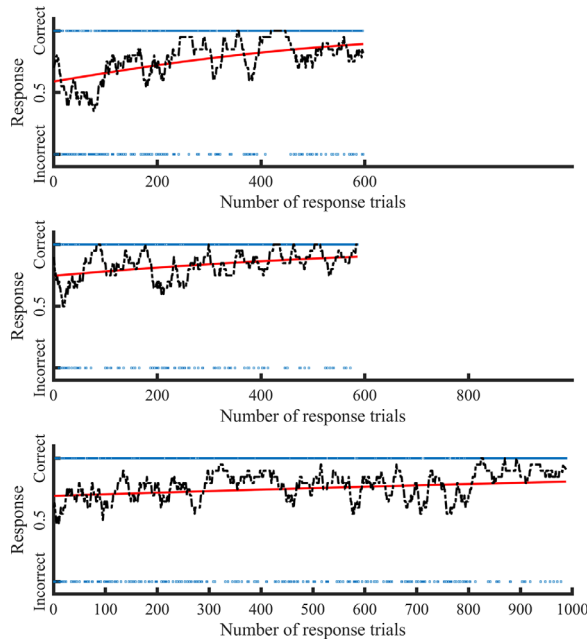Table 1. Logistic regression results from experiment 1.

Figure 3. The logistic regression results of experiment 1 for M1 (top), M2 (middle), and M3 (bottom). The *x*-axis represents the number of response trials (trials without responses were removed), and the *y*-axis represents the monkey's response accuracy. As M2's response rate was 73%, only 584 trials remained. The monkeys' responses for each trial are shown as blue dots, which appears as a blue line because of the large number of trials. The red line represents the predicted response probability produced from the logistic regression analysis. The black dotted line represents the response accuracy of a moving average of 20 trials, which is for illustration purposes only and not used for calculating logistic regression. The intercepts of the regression lines for all 3 monkeys were larger than 0.5, indicating that all 3 monkeys were able to generalize from the training set to the testing set. The regression line increased along with the trial number, suggesting that monkeys continued to learn during the testing phase to improve their performance. M1 was tested only for 3 days; therefore, it has only 600 trials. M2 was tested for 5 days, but data from the first day were removed from the logistic regression due to significantly below chance categorization that likely resulted from a memorization strategy used during the training period.

or inanimate category, or both. By investigating the underlying cause of the performance improvement across categories during the testing phase, we can better understand the learning dynamics between categories across trials. Therefore, we re-ran the above logistic regression by adding stimulus category as an additional categorical regressor to examine the possible interaction between category and trial number. As before, we found a significant intercept of the logistic regression model for all monkeys (M1: beta = 0.737, $p = 5.832 \times 10^{-4}$; M2: beta = 0.813, $p = 1.134 \times 10^{-3}$; and M3: beta = 0.376, $p = 0.0247$), as well as significant coefficients for the trial number (M1: beta = 0.0042, $p = 1.091 \times 10^{-14}$; M2: beta = 0.00123, $p = 0.0378$; and M3: beta = 0.00103, $p = 1.887 \times 10^{-5}$), which suggests significant generalization and learning effects. There was also a significant interaction between trial number and category for M1 (beta = -0.00327, $p = 5.463 \times 10^{-4}$) and M3 (beta = -0.00121, $p = 7.358 \times 10^{-4}$). These results suggest that the difference in categorization accuracy between animate and inanimate categories decreased significantly as testing progressed, implying that a greater learning effect occurred for animate than for inanimate trials. However, we did not observe a significant interaction between category and trial number for M2 (beta = 0.00142, $p = 0.1334$).

3) Contribution of curvilinear and rectilinear features to monkeys' performance at the group level

We aimed to understand the extent to which the amount of intermediate image features, specifically curvilinear and rectilinear features (see Methods), present in the images in experiment 1 contributed to the monkeys' performance on the categorization task. To answer this question, we conducted a logistic regression analysis of curvilinear and rectilinear values with the monkeys' performance, which was performed at the group level to increase the signal-to-noise ratio (see Methods).

We found that the amount of intermediate image features in the intact images significantly predicted the monkeys' performance (main effect: $\chi^2 \ [2159] = 107.4$, $p = 1.450 \times 10^{-21}$), suggesting that the amount of intermediate image features might assist them in categorizing intact images into animate and inanimate groups. Furthermore, we found that curvilinear values
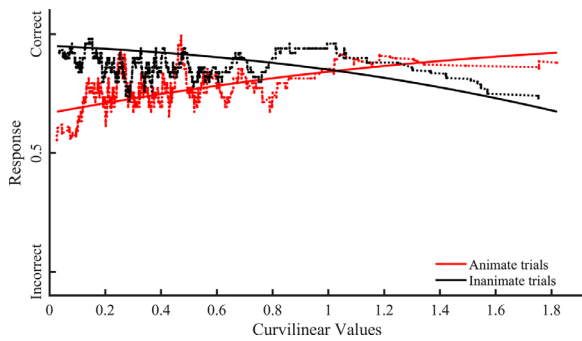
Figure 4. Functional relationship between the amount of curvilinear information present in visual stimuli and monkeys' performance across stimulus category in experiment 1. The x-axis represents the curvilinear values of the stimuli. The *y*-axis represents the response probability of the monkeys' performance. The solid lines represent the response probability to visual stimuli calculated with the logistic regression model that was created using the monkeys' group raw response. The dotted lines represent a moving average of 60 trials, which is for illustration purposes only and was not used for fitting the logistic regression model. The red line represents the response probability resulting from the logistics regression fit for the animate trials. The black line represents the response probability resulting from the logistics regression fit for the inanimate trials.

of intact images significantly predicted the monkeys' performance (beta = 0.974, $p = 0.031$), but rectilinear values did not (beta = -0.4817, $p = 0.272$). There was a significant interaction between the curvilinear values and the stimulus category (beta = -2.21, $p = 1.118 \times 10^{-4}$), indicating that curvilinear values predicted the monkeys' performance in animate trials differently than on inanimate trials. Figure 4 shows the functional relationship between curvilinear values and the monkeys' performance across animate and inanimate trials, which was produced from the logistic regression model. As the amount of curvilinear information in an image increased, the monkeys' performance increased for animate images and decreased for inanimate images. The relationship with rectilinear values is shown in Supplementary Figure S3.

These results suggest that, in addition to recognizing local or global features that they had learned during daily training, the monkeys may have used the amount of curvilinear image features present in the stimuli to categorize objects into animate and inanimate groups.

It is possible that the significance of the contribution of intermediate image features to categorization accuracy observed in the group analysis is mainly caused by the performance of one monkey. We re-ran the above logistic regression by incorporating animals in the model as an additional categorical regressor to address this concern. In the model, M1 was encoded as the reference category compared with M2 and M3.

We found (1) a significant contribution of intermediate image features to the monkeys' performance (χ^2 [2157] = 125.0, $p = 8.150 \times 10^{-24}$); (2) that curvilinear values of intact images significantly predicted the monkeys' performance (beta = 1.029, $p = 0.024$), but rectilinear values did not (beta = -0.508, $p = 0.251$); and (3) a significant interaction between the curvilinear values and the stimulus category (beta = -2.256, $p = 8.960 \times 10^{-5}$). Furthermore, categorical coefficients of M2 and M3 in the logistic regression model were not significant (for M2: beta = 0.250, $p = 0.067$; for M3: beta = -0.242, $p = 0.060$), suggesting that the effects of intermediate visual features on categorization were consistent across animals.

We computed the Cronbach's alpha (see Methods) to further evaluate response consistency across monkeys. The Cronbach's alpha was 0.809, which suggests that categorization performance was reliable across monkeys in experiment 1.

4) Predicting stimulus category with curvilinear and rectilinear values

It is of great interest to understand whether categorical membership of an image can be predicted by the images' curvilinear and/or rectilinear values calculated with the current method. As such, we ran a logistic regression of curvilinear and rectilinear values of intact images on the stimulus category to examine whether and the extent to which those measures could be used to determine the stimulus category. We found that both curvilinear and rectilinear values significantly predicted stimulus category (main effect: χ^2 [997] = 104, $p = 3.3 \times 10^{-23}$; for curvilinear values: beta = 2.893, $p = 6.538 \times 10^{-14}$; for rectilinear values: beta = -3.204, $p = 3.501 \times 10^{-18}$). This result suggests that the likelihood of an image being animate increases significantly as curvilinear values increase, and the likelihood of images being inanimate increases with increased rectilinear values, as clearly demonstrated in Figure 5. In the curvilinear and rectilinear space (Figure 5), an image's categorical membership could be readily determined along the diagonal, especially when the values are less than 1 in the raw data.

We found that animate objects had larger curvilinear values than inanimate objects, on average, in our image set (t(499) = 3.659, $p = 2.721 \times 10^{-4}$), which is consistent with results from previous studies (Kurbat, 1997; Levin et al., 2001) showing that animate objects have more curved features. By contrast, animate objects have smaller rectilinear values than inanimate objects (t(499) = -6.168, $p = 1.177 \times 10^{-9}$).

Experiment 2: Synthesized images

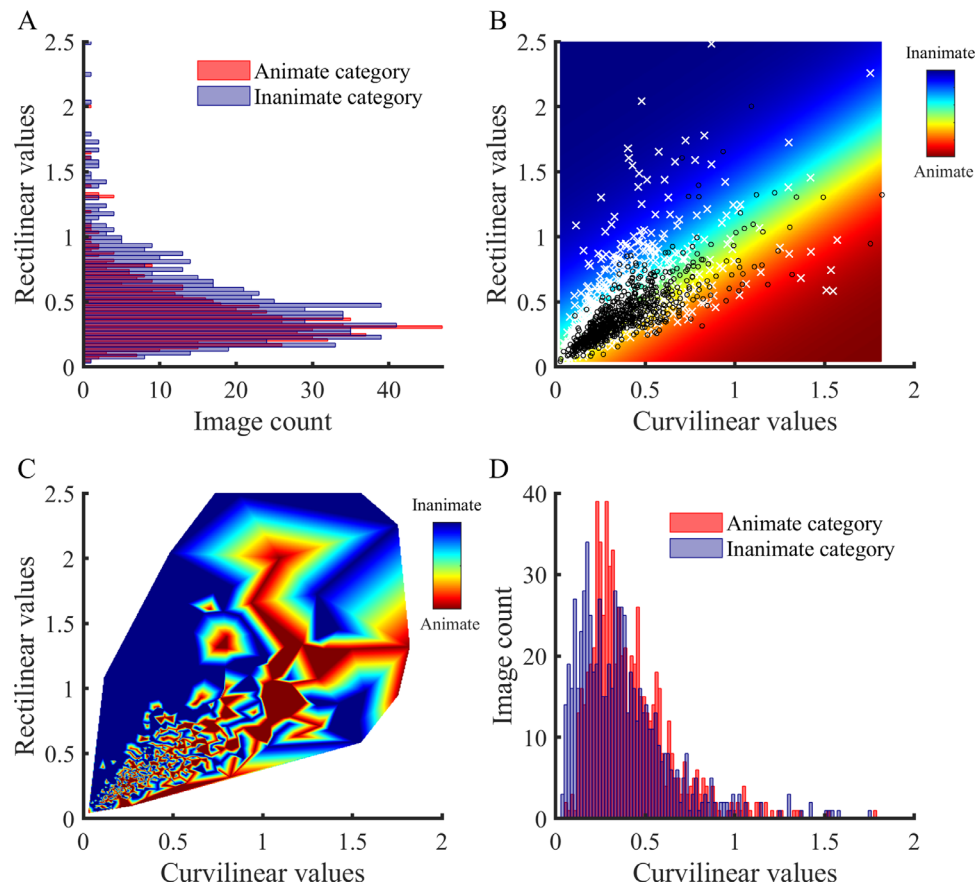1) Overall classification accuracy for individual monkeys

Figure 5. Distribution of curvilinear and rectilinear values of visual stimuli and relationship with stimulus category in experiment 1. (**A**) The distribution of rectilinear values of images for animate (red) and inanimate category (blue). (**B**) The relationship between the raw curvilinear and rectilinear values of the stimuli and their category captured by the logistic regression. Each data point represents an image and its corresponding raw curvilinear and rectilinear values, with black circles denoting the animate images and white crosses representing the inanimate images. All 500 images of each category were plotted. Not all data points are easily visible due to a high degree of overlap of the images with curvilinear and rectilinear values between 0 and 1. The colors of the plot represent the probability of the images belonging to the animate (warm colors) or inanimate categories (cool colors) as predicted by logistic regression. Both curvilinear and rectilinear values significantly predicted stimulus category (main effect: χ^2 [997] = 104, $p = 3.3 \times 10^{-23}$; for curvilinear values: beta = 2.893, $p = 6.538 \times 10^{-14}$; for rectilinear values: beta = -3.204, $p = 3.501 \times 10^{-18}$). (**C**) Linearly interpolated relationship between the curvilinear and rectilinear values of the stimuli and their category membership (see Methods). The warm colors represent the animate category and cool colors represent the inanimate category. The x-axis and y-axis represent the curvilinear and rectilinear values of stimuli. (**D**) The distribution of curvilinear values of images for animate (red) and inanimate category (blue).

The monkeys were never trained to categorize the synthesized images presented in experiment 2. Furthermore, the synthesized images were each shown only once, regardless of the monkeys' responses. All three monkeys performed the categorization task significantly above chance (overall accuracy for M1, 64.48% [chance: 50.01%], $p < 0.0001$; M2, 59.10% [chance: 49.83%], $p < 0.0001$; and M3, 60.27% [chance: 49.99%], $p < 0.0001$). The overall response rate was 99.6% for M1, 92.7% for M2, and 85.1% for M3. Although the overall classification accuracies were lower than those for the intact images in experiment 1, the significant above-chance performances suggest that the image features distinguishing the two groups

of synthesized images provided sufficient information for monkeys to classify the images into the two categories.

2) Generalization and learning effect for individual monkeys

To provide a parallel analysis to the one performed in experiment 1, we ran a logistic regression to evaluate if the monkeys' overall accuracies for categorizing the synthesized images resulting from generalizing visual features learned from the intact images to the synthesized images and/or continuous learning (Figure 6). We found that the intercept, but not the

	Logistic regression	
Monkeys	Intercept	Slope
M1	0.533 ($p = 2.038 \times 10^{-6}$)	9.095×10^{-5} ($p = 0.521$)
M2	0.313 ($p = 1.480 \times 10^{-2}$)	1.150×10^{-3} ($p = 0.606$)
M3	0.428 ($p = 4.816 \times 10^{-3}$)	-1.930×10^{-5} ($p = 0.919$)

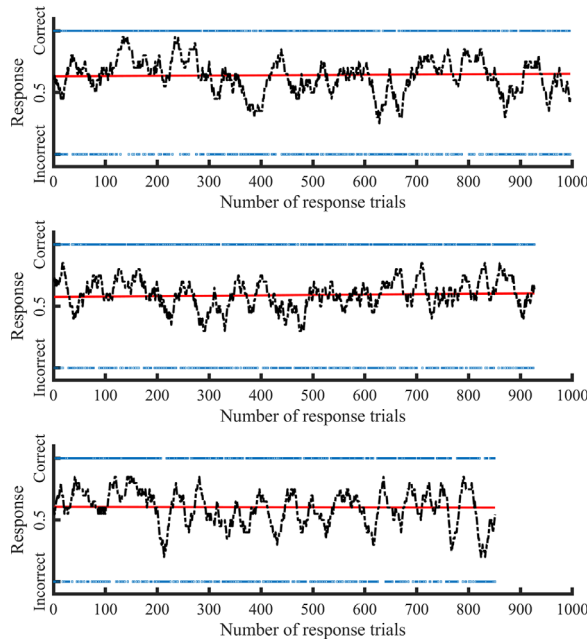Table 2. Logistic regression result of Experiment 2.

Figure 6. The logistic regression results of experiment 2 for M1 (top), M2 (middle), and M3 (bottom). Axes are the same as those used in Figure 3. As shown in Table 2, all three monkeys showed significant generalization but no learning effects. These results suggest that the monkeys used some image features distinguishing intact animate images from intact inanimate images to categorize the synthesized images as animate or inanimate.

slope, of the logistic regression model was significant for all three monkeys, as shown in Table 2. Performance was not significantly determined by test trial number for any monkeys (for M1: χ^2 [994] = 0.365, $p = 0.546$; M2: χ^2 [925] = 0.340, $p = 0.560$; and M3: χ^2 [849] = 0.032, $p = 0.859$), indicating that the monkeys' performance did not improve as testing progressed. These results reveal that, at the onset of experiment 2, all three monkeys used information they learned on the categorization task in experiment 1 to classify the synthesized images as animate and inanimate objects.

3) Contribution of curvilinear and rectilinear features to monkeys' performance at the group level

To examine the extent to which the amount of intermediate visual features contributed to the monkeys'

performance in experiment 2, we used the same testing procedure as experiment 1 but with synthesized images.

We found a significant main effect of the amount of curvilinear and rectilinear image features on the monkeys' performance (χ^2 [2768] = 177.160, $p = 2.160 \times 10^{-36}$). Furthermore, both curvilinear and rectilinear values of synthesized images significantly predicted the monkeys' performance (curvilinear: beta = 1.617, $p = 2.615 \times 10^{-7}$; and rectilinear: beta = -1.257, $p = 5.865 \times 10^{-4}$). However, the data suggested that the amount of curvilinear image features present in the synthesized images played a more dominant role than the amount of rectilinear image features. To test this hypothesis, we performed a regression Wald test to examine whether the curvilinear coefficient was significantly different from the rectilinear coefficient. The curvilinear coefficient was significantly larger than the rectilinear coefficient (Wald test: χ^2 [1] = 19.938, $p = 7.994 \times 10^{-6}$), indicating that the amount of curvilinear image features present in the synthesized images was more informative for the categorization task than the amount of rectilinear image features. As such, the following analysis of interaction between the amount of intermediate image features and stimulus category was focused on the contribution of the amount of curvilinear image features on the monkeys' performance across stimulus categories. Results from the analysis of the interaction effect between the amount of rectilinear image features with stimulus category are shown in Supplementary Figure S4.

We observed a significant interaction between the curvilinear values of stimuli and stimulus category (beta = -4.040, $p = 1.672 \times 10^{-20}$). The monkeys' performance on synthesized images increased when curvilinear values increased in the animate trials but decreased in the inanimate trials (Figure 7); similar to what we observed in experiment 1 (Figure 4). These data indicate that the more curvilinear information present in an animate image, the more likely it was to be categorized correctly, whereas the opposite is true for inanimate images.

We conducted the second logistic regression by incorporating animals in the model as an additional categorical regressor to determine whether the contribution of intermediate image features to categorization accuracy observed in the group analysis (Figure 7) was biased by one of the monkeys'
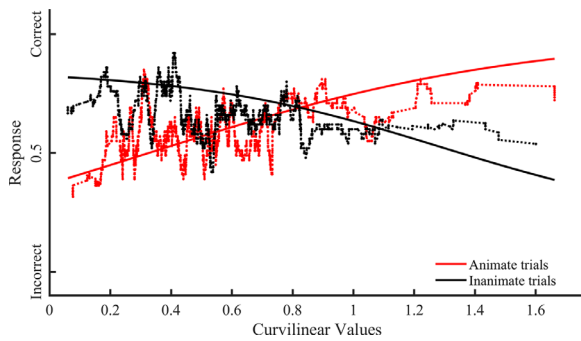
Figure 7. Functional relationship between amount of curvilinear information present in the visual stimuli and monkey's group performance across stimulus category in experiment 2. The *x*-axis represents the curvilinear values of visual stimuli. The *y*-axis represents the response probability of the monkeys' performance. The solid lines represent the response probability to visual stimuli calculated with the logistic regression model that was created using the monkeys' group raw response. The dotted lines represent a moving average of 60 trials, which is for illustration purposes only. The red line represents the response probability resulting from the logistics regression fit for the animate trials. The black line represents the response probability resulting from the logistics regression fit for the inanimate trials.

performance. We found (1) a significant contribution of intermediate image features to the monkeys' performance (χ^2 [2766] = 145.00, $p = 3.860 \times 10^{-28}$); (2) that curvilinear values of texturized images significantly predicted the monkeys' performance (beta = 2.149, $p = 1.069 \times 10^{-9}$), as did rectilinear values (beta = -1.036, $p = 6.912 \times 10^{-3}$); (3) a significant interaction between the curvilinear values and the stimulus category (beta = -4.737, $p = 6.875 \times 10^{-24}$), as well as rectilinear values and stimulus category (beta = 3.384, $p = 1.529 \times 10^{-11}$); and (4) that the categorical coefficients of M2 and M3 in the logistic regression model were insignificant (for M2: beta = -0.0086, $p = 0.931$; for M3: beta = -0.142, $p = 0.116$), suggesting that the effects of intermediate visual features on categorization were consistent across animals. The Cronbach's alpha (see Methods) was 0.737, which indicates that categorization performance was reliable across monkeys in experiment 2.

4) Predicting stimulus category with curvilinear and rectilinear values

The algorithm used to create synthesized images removed the global shape information. It was unclear whether (1) synthesized images' curvilinear and rectilinear values significantly predict their categorical membership, and (2) if these values were significantly different across categories as we observed for the intact images used in experiment 1. To address the first

question, we ran a logistic regression of curvilinear and rectilinear values of synthesized images on the stimulus category. Consistent with what we found in experiment 1, both curvilinear and rectilinear values significantly predicted stimulus category (main effect: χ^2 [997] = 77.1, $p = 1.77 \times 10^{-17}$; for curvilinear values: beta = 2.438, $p = 3.772 \times 10^{-10}$; for rectilinear values: beta = -3.568, $p = 2.139 \times 10^{-14}$). The relationship between curvilinear and rectilinear values of synthesized images with the stimulus category in experiment 2 (Figure 8) is similar to what we observed in experiment 1 (Figure 5), suggesting that the likelihood of an image being animate increases significantly as the increase of curvilinear values.

To address the second question, we ran paired independent *t*-tests. Animate objects had larger curvilinear values (t(499) = 2.285, $p = 0.026$), and smaller rectilinear values than inanimate objects (t(499) = -8.794, $p = 1.172 \times 10^{-17}$) in our set of synthesized images. The amount of intermediate image features between categories is significantly different in both experiments 1 and 2, which confirms that the algorithm preserved some intermediate image features while eliminating the global shape information of intact images.

5) Correlation of the monkeys' performance with DCNN performance at the group level

Because monkeys were never trained to classify synthesized images into animate and inanimate categories, the possibility remained that monkeys categorized the images into two groups using differences between synthesized images that were entirely unrelated to the animate and inanimate category but happened to coincide with the two categories in the set of testing images used. As such, we used the DCNN to address this concern (see Methods). The network was trained to classify the 1000 intact images used in experiment 1 into animate and inanimate categories and then tested on the categorization task with the 1000 synthesized images used in experiment 2 (see Methods). We found a significant positive correlation of the DCNN's categorization performance with the monkeys' group performance (r = 0.739, $p = 5.0502 \times 10^{-8}$; Figure 9), suggesting that the monkeys performed the animate versus inanimate categorization in experiment 2, when the global form in the images was distorted beyond recognition. These data provided further evidence that the monkeys used image features distinguishing intact animate and inanimate images to categorize the synthesized images.

Furthermore, we completed a χ^2 test for each monkey individually to investigate whether a monkey's responses to stimuli were independent of DCNN responses at the individual image level (nonresponsive trials were excluded from the analysis). We found
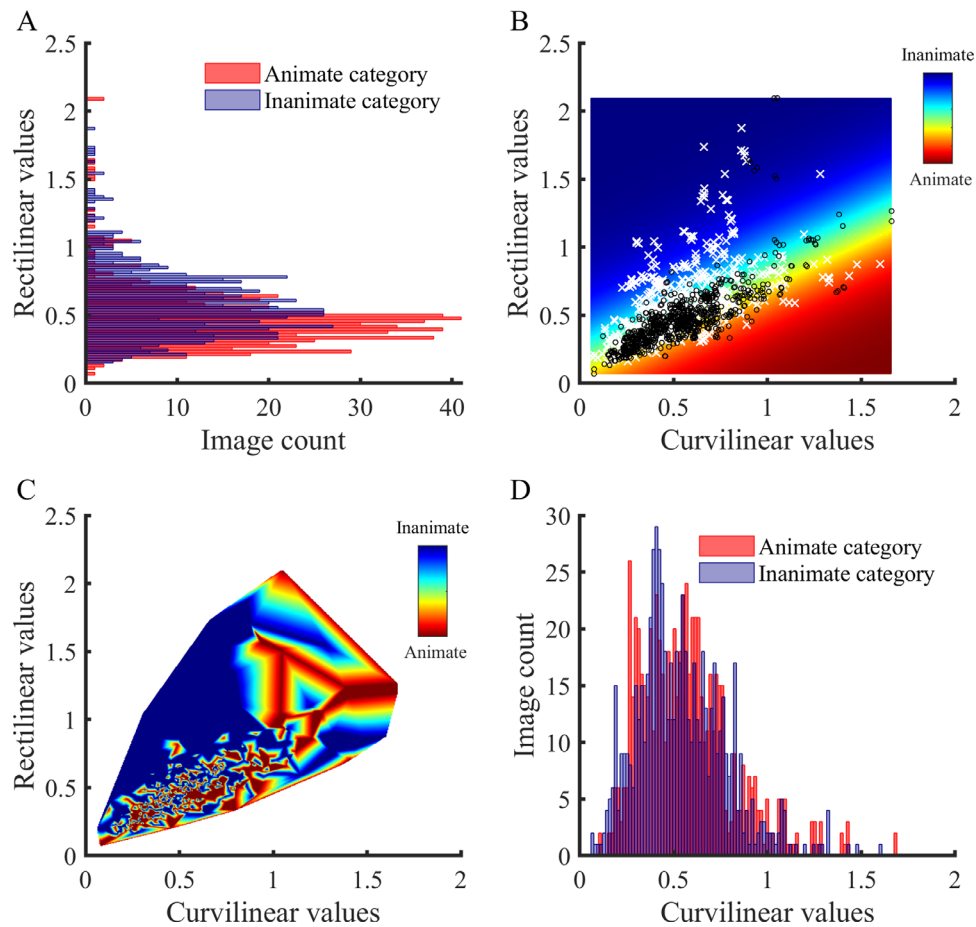
Figure 8. Distribution of curvilinear and rectilinear values of visual stimuli and relationship with stimulus category in experiment 2. (**A**) The distribution of rectilinear values of images for animate (red) and inanimate category (blue). (**B**) The relationship between the raw curvilinear and rectilinear values of the stimuli and their category captured by the logistic regression. Each data point represents an image and its corresponding raw curvilinear and rectilinear values, with black circles denoting the animate images and white crosses representing the inanimate images. All of 500 images of each category were plotted. Not all data points are easily visible due to a high degree of overlap of the images with curvilinear and rectilinear values between 0 and 1. The colors of the plot represent the probability of the images belonging to the animate (warm colors) or inanimate categories (cool colors) as predicted by logistic regression. Both curvilinear and rectilinear values of synthesized images significantly predicted the stimulus category (main effect: χ^2 [997] = 77.1, $p = 1.77 \times 10^{-17}$; for curvilinear values: beta = 2.438, $p = 3.772 \times 10^{-10}$; for rectilinear values: beta = −3.568, $p = 2.139 \times 10^{-14}$). (**C**) Linearly interpolated relationship between the curvilinear and rectilinear values of the stimuli and their category membership (see Methods). The warm colors represent the animate category and cool colors represent the inanimate category. The *x*-axis and *y*-axis represent the curvilinear and rectilinear values of stimuli, respectively. (**D**) The distribution of curvilinear values of images for animate (red) and inanimate category (blue).

significant χ^2 results for all monkeys: $\chi^2 = 24.762$, $p = 6.488 \times 10^{-7}$ for M1, $\chi^2 = 5.697$, $p = 0.017$ for M2, and $\chi^2 = 28.848$, $p = 7.878 \times 10^{-8}$ for M3, indicating that monkeys performed similarly to DCNN at the individual image level.

Discussion

This study investigated the contributions of both training and image-based features to the perceptual categorization of animacy. In experiment 1, we found that naïve monkeys trained to categorize a small set of animate and inanimate images classified a large set of unfamiliar images into animate and inanimate categories with high accuracy. In experiment 2, we tested whether image-based features that differ between the two object categories in the statistics of natural environments (i.e., curvilinear and rectilinear information; Kurbat, 1997; Levin et al., 2001; Perrinet & Bednar, 2015; Long et al., 2017; Zachariou et al., 2018), determined the monkeys' classification accuracy. We created sets of synthetic animate and inanimate images using an algorithm that significantly distorted the global shape of the original images while maintaining the original images' intermediate features (Portilla & Simoncelli, 2000). The monkeys' classification accuracy on these synthesized images was still significantly above
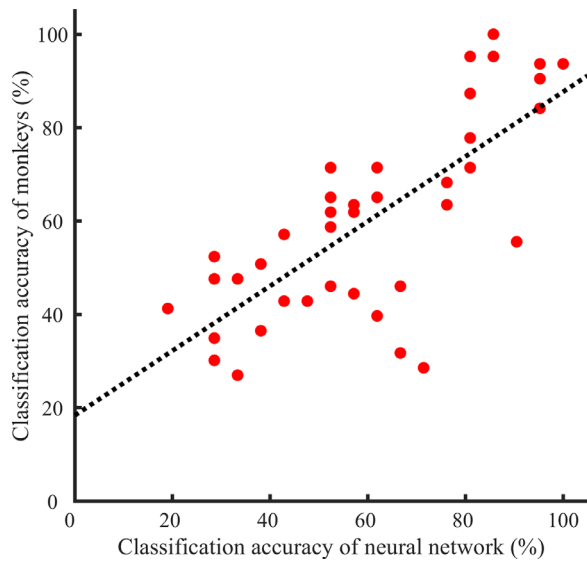
Figure 9. Correlation of monkeys' response accuracies with DCNN classification accuracies. To compute the correlation of the DCNN classification accuracies and monkeys' response accuracies to the synthesized images, we arranged the responses of the DCNN and each monkey according to the ascending order of curvilinear values of the synthesized images. The monkeys' accuracies used for the correlation analysis were averaged across all three animals. The ordered responses were then grouped into 40 bins. Next, the response accuracy for each bin was calculated for the DCNN and monkeys separately, resulting in 2 sets of 40 data points. Each red dot represents the classification accuracy for each bin. We observed a significant correlation between monkeys' response accuracies and DCNN classification accuracies ($r = 0.739$, $p = 5.0502 \times 10^{-8}$), indicating that the monkeys performed the animate versus inanimate categorization.

performance. These findings indicate that image-based features that are predictive of each category provide substantial information that monkeys can use to distinguish the two categories with little training. In other words, experience interacting with objects may not be the only origin of behavioral categorization of animacy in monkeys.

To confirm this, using the synthesized images in experiment 2, we eliminated local features (faces, ears, etc.) that monkeys might have been familiar with and could have used to classify the images into animate and inanimate categories. We found that the monkeys were able to perform the categorization of the synthesized images significantly above chance, which indicates that the image-based features were sufficient for perceptual categorization. It is worth noting that human participants also classified synthesized images similar to those used in this experiment into animate and inanimate categories with significant above-chance accuracy (Long et al., 2017; Zachariou et al., 2018). Although humans and monkeys do not share the collective experience of what and how objects are encountered in daily life, they perform similarly when classifying synthesized images into animate and inanimate categories (Figure 6, Figure 3 in Zachariou et al., 2018), which suggests that image-based feature differences could play a critical role in the emergence of perceptual categorization abilities across species. Together, our findings provide strong evidence in support of the hypothesis that perceptual categorization can emerge from image-based features that are predictive of each category in the natural statistics of the visual environment.

Recent functional magnetic resonance imagery (fMRI) studies (Long et al., 2018; Yue et al., 2020) have shown that visual cortical areas selective for curvilinear features encompass animate-processing visual areas, whereas those selective for rectilinear features encompass inanimate-processing visual areas. These results provide neural evidence to support the current finding that the processing of image-based features, such as curvilinearity, interacts with the representation of animate and inanimate categories.

Overall, monkeys categorized the intact object images with significantly greater accuracy than the synthesized images. However, for synthesized images with high curvilinear values (in the range of 1.4–1.6), the monkeys' classification accuracy for the animate category could reach above 80%, which is comparable to the classification accuracy for intact images (Figure 7). This illustrates that monkeys could achieve high accuracy when synthesized images with extreme curvilinear values were used as stimuli. Thus, the overall difference in classification accuracy between the intact and synthesized images does not argue against the idea that image-based features play a significant role in determining perceptual categorization.

chance and correlated with the amount of curvilinear information present in the stimuli. These data indicate that image-based features, in this case curvilinearity, can be used to distinguish animate from inanimate objects in the absence of global shape information without prior training.

As monkeys raised in the laboratory have limited experiences with objects that humans are otherwise familiar with, they are ideal candidates to study the contribution of experiences and image-based features to the emergence of perceptual categorization (e.g., Arcaro & Livingstone, 2017). Our results show that monkeys performed an animacy categorization task with intact images significantly above chance at the very beginning of the test phase of experiment 1, suggesting that monkeys used what they had learned during training to classify novel images of objects, with which they had no previous experience, into animate and inanimate categories. Further, the curvilinear values of intact images had a significant interaction with stimulus category, and significantly predicted the monkeys'

The primate visual system takes significant time to fully mature postnatally (Ellemberg et al., 1999; Kovacs et al., 1999; Gilmore et al., 2018). During development, young infants view the world as consisting not of coherent objects but instead visual pieces that move in unpredicted ways (Hyvärinen et al., 2014). In such a fragmented visual world, differentiating animate from inanimate objects would be challenging. Infants who can differentiate animate from inanimate objects would have a better chance to avoid being harmed by animals to survive than those who cannot. Through natural selection, our brains may have evolved the capacity to differentiate animate and inanimate objects quite quickly, first based on sensory information that represents visual statistics of the natural environment. Experience with objects would play a significant role in later life to further differentiate categories. Our data provide evidence to support this hypothesis by showing that monkeys (as well as humans; Zachariou et al., 2018) are able to classify into animate and inanimate categories synthesized images that: (1) neither species had experience with, and (2) have similar statistics as the natural original images, with significantly above chance accuracy by using the degree of curvilinearity in the images. This hypothesis raises many interesting questions. For what other object categories and with which image features is the primate brain biased to use image-based differences for perceptual categorization, and under what conditions? The answers to such questions are critical to understanding the functional and anatomic organization of the primate visual system.

Keywords: categorization, animate, curvilinearity, animacy, curvature patches

Acknowledgments

Corresponding author: Xiaomin Yue.
Email: xiaominyue@gmail.com.
Address: Laboratory of Brain and Cognition, NIMH/NIH, Building 49, Room 6A68, 49 Convent Drive, Bethesda, MD 20892, USA.

*MY and SR contributed to the work equally.

References

Arcaro, M. J., & Livingstone, M. S. (2017). A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife, 6*, e26196.

Blake, C. E., Bisogni, C. A., Sobal, J., Devine, C. M., & Jastran, M. (2007). Classifying foods in contexts: how adults categorize foods for different eating settings. *Appetite, 49*(2), 500–510.

Bovet, D., & Vauclair, J. (1998). Functional categorization of objects and of their pictures in baboons (Papioanubis). *Learning Motivation, 29*(3), 309–322.

Calvillo, D. P., & Hawkins, W. C. (2016). Animate objects are detected more frequently than inanimate objects in inattentional blindness tasks independently of threat. *Journal of General Psychology, 143*(2), 101–115.

Chiara, T., Bulf, H., & Simion, F. (2008). Newborns' face recognition over changes in viewpoint. *Cognition, 106*, 1300–1321.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255).

Ellemberg, D., Lewis, T. L., Liu, C. H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research, 39*(14), 2325–2333.

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Review Neuroscience, 19*(3), 123–137.

Hays, A. V., Richmond, B. J., & Optican, L. M. (1982). A UNIX-based multiple-process system for real-time data acquisition and control. *WESCON Conference Proceedings*, 2, 1–10.

Heron-Delaney, M., Wirth, S., & Pascalis, O. (2011). Infants' knowledge of their own species. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 366*(1571), 1753–1763.

Hyvärinen, L., Walthes, R., Jacob, N., Chaplin, K. N., & Leonhardt, M. (2014). Current understanding of what infants see. *Current Ophthalmology Reports, 2*(4), 142–149.

Kalénine, S., Bonthoux, F., & Borghi, A. M. (2009). How action and context priming influence categorization: a developmental study. *The Brit Journal of Devopmental Psychology, 27*(3), 717–730.

Kalénine, S., Shapiro, A. D., Flumini, A., Borghi, A. M., & Buxbaum, L. J. (2014). Visual context modulates potentiation of grasp types during semantic object categorization. *Psychonomic Bulletin Review, 21*(3), 645–651.

Kovacs, I, Kozma, P, Fehér, A., & Benedek, G. (1999). Late maturation of visual spatial integration in humans. *Proceedings of the National Academy of Sciences of the United States of America, 96*(21), 12204–12209.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 60*(6), 1097–1105.

Krüger, N. (2001). Learning object representations using a priori constraints within ORASSYLL. *Neural Computation, 13*(2), 389–410.

Kurbat, M. A. (1997). Can the recognition of living things really be selectively impaired? *Neuropsychologia, 35*(6), 813–827.

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Attention, Perception Psychophysics, 63*(4), 676–697.

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision, 17*, 20–20.

Long, B., Yu, C. P., & Konkle, T. (2018).Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America, 115*, E9015–9024.

LoBue, V., & DeLoache, J. S. (2011). What's so special about slithering serpents? Children and adults rapidly detect snakes based on their simple features. *Visual Cognition, 19*, 129–143.

Long, B., Moher, M., Carey, S. E., & Konkle, T. (2019). Animacy and object size are reflected in perceptual similarity computations by the preschool years. *Visual Cognition, 27*(5–8), 435–451.

Lipp, O. V. (2006). Of snakes and flowers: Does preferential detection of pictures of fear-relevant animals in visual search reflect on fear-relevance? *Emotion, 6*, 296–308.

Livingstone, M. S., Vincent, J. L., Arcaro, M. J., Srihasam, K., Schade, P. F., & Savage, T. (2017). Development of the macaque face-patch system, *Nature Communications, 8*, 14897.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99*(4), 587.

Meyerhoff, H. S., Schwan, S., & Huff, M. (2014). Perceptual animacy: Visual search for chasing objects among distractors. *Journal of Experimental Psychology: Human, 40*(2), 702–717.

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science, 26*(1), 22–27.

Opfer, J. E., & Gelman, S. A. (2011). Development of the animate-inanimate distinction. *The Wiley-Blackwell Handbook of Childhood Cognitive Development, 2*, 213–238.

Perrinet, L. U., & Bednar, J. A. (2015). Edge co-occurrences can account for rapid categorization of natural versus animal images. *Scientific Reports, 5*, 11400.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40*(1), 49–70.

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology, 4*(1), e27.

Rakison, D. H. (2003). Parts, motion, and the development of the animate–inanimate distinction in infancy. In D. H. Rakison, & L. M. Oakes (Eds.), *Early Category and Concept Development* (pp. 159–192). Oxford University Press.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences of the United States of America, 105*(2), 809–813.

Srihasam, K., Vincent, J.L., & Livingstone, M.S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature Neuroscience, 17*, 1776–1783.

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America, 105*(1), 394–398.

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition, 105*(2), 362–379.

Wang, P., & Nikolic, D. (2011). An LCD monitor with sufficiently precise timing for research in vision. *Frontiers in Human Neuroscience, 5*, 85.

Yue, X., Pourladian, I. S., Tootell, R. B. H., & Ungerleider, L.G. (2014). Curvature-processing network in macaque visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 111*, E3467–E3475.

Yue, X., Robert, S., & Ungerleider, L. G. (2020). Curvature processing in human visual cortex. *NeuroImage, 222*, 117295.

Zachariou, V., Del Giacco, A. C., Ungerleider, L. G., & Yue, X. (2018). Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *Journal of Vision, 18*, 388–398.