

1 Title: Curvilinear features are important for animate/inanimate categorization in macaques

2

3 Authors: Marissa Yetter¹, Sophia Robert¹, Grace Mammarella², Barry Richmond², Mark A. G.

4 Eldridge², and Leslie G. Ungerleider¹, Xiaomin Yue¹

5

6 ¹Laboratory of Brain and Cognition, NIMH/NIH, Bethesda, MD 20892

7 ²Laboratory of Neuropsychology, NIMH/NIH, Bethesda, MD 20892

8

9 Corresponding author:

10 Xiaomin Yue PhD

11 Laboratory of Brain and Cognition, NIMH/NIH

12 Building 49, Room 6A68

13 49 Convent Drive

14 Bethesda, MD 20892

15 Tel: 301-443-8417

16 Email: xiaominyue@gmail.com

17

18 Number of figures: 7

19

20

21

22

23

24

25 **Author contributions:** Ms. Sophia Robert and Ms. Marissa Yetter contributed to the work
26 equally.

27

28

29

30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Abstract

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 and 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.

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

48

Introduction

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

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. Though 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 (Kalénine et al., 2009, 2014; Blake et al., 2007), while 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 (Nairne et al., 2017; Meyerhoff et al., 2014; Calvillo et al., 2016; Long et al., 2019), even as early as 3 months old (Heron-Delaney et al., 2011; Opfer & Gelman, 2011; Rakison, 2003). 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 (Livingstone et al., 2017; Srihasam et al., 2014). Yet another line of research has shown that human infants might develop concepts of animacy based on differences between biological and non-biological motion (Simion et al. 2008; Mandler, 1992).

79 That the animate-inanimate distinction might be special to our visual system, and that
80 these two categories differentially covary with a number of image features, suggests a plausible
81 mechanism by which the primate visual system evolved to exploit image feature covariances,
82 such as those listed above, to make animate-inanimate categorization judgments. One such
83 feature is curvilinearity, or the extent to which the image of an object is composed of curved
84 lines and textures. Animate objects tend to be more curvilinear than inanimate objects (Kurbat,
85 1997; Levin et al., 2001). A recent study by Zachariou et al. (2018) demonstrated that, when
86 deprived of global shape cues, humans were able to categorize animate and inanimate objects
87 using just curvilinear information. Further, curvilinear information was positively correlated with
88 performance on images of animate objects and negatively correlated with performance on
89 inanimate objects. Given the lack of object shape information in the stimuli used and the lack of
90 relationship between subjects' confidence ratings and their accuracy, it appears that this
91 categorization ability is driven by an implicit, primarily bottom-up process.

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

100

101

102

Materials and Methods

103 Subjects:

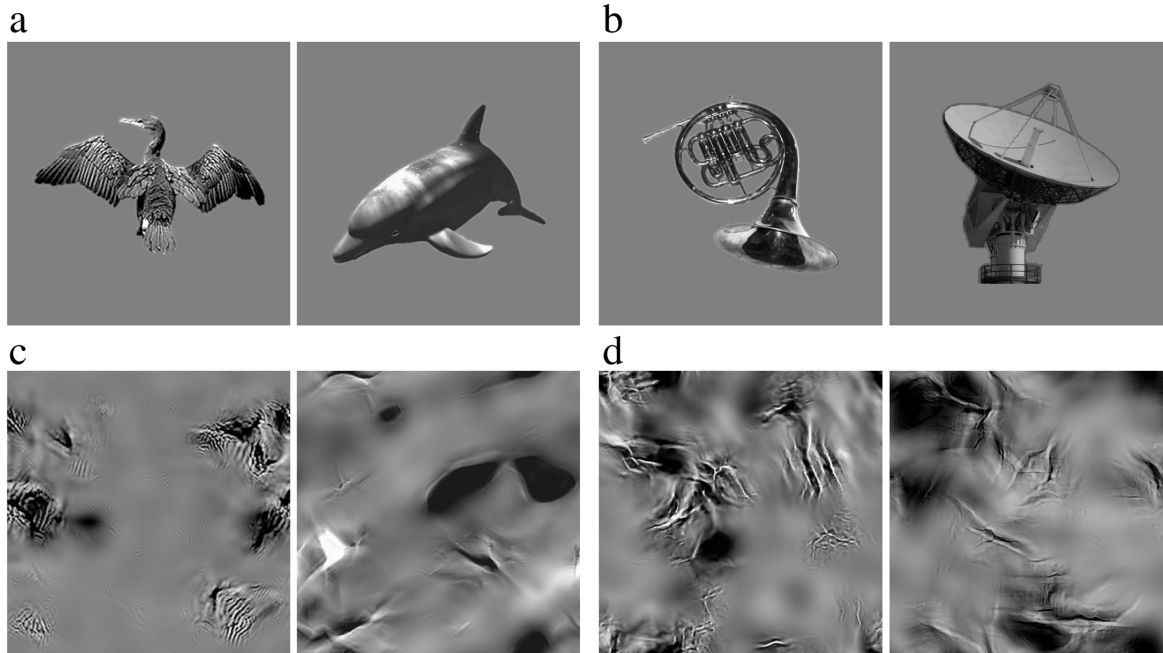
104 Three male rhesus macaques (5 - 8 kg) were used in two behavioral experiments. All
105 experimental procedures were approved by the National Institute of Mental Health Animal Care
106 and Use Committee.

107

108 Visual stimuli:

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

117 For the second experiment, we used an algorithm, described in detail in Portilla and
118 Simoncelli (2000), to generate synthesized images of animate and inanimate objects (Figure 1c
119 and 1d) that abolished the global shape of the original images but maintained their intermediate
120 visual features (see Supplementary Materials). 1000 synthesized images were generated using
121 the testing set of 500 animate and 500 inanimate intact images used in Experiment 1.



122
123 Figure 1: Examples of stimuli: (a) animate images; (b) inanimate images; (c) synthesized
124 animate images; (d) synthesized inanimate images.

125

126 Experimental procedures:

127 The monkeys sat in a primate chair inside a darkened, sound-attenuated testing chamber.
128 They were positioned 57 cm from a computer monitor (Samsung 2233RZ, Wang and Nikolic
129 2011)) subtending $40^\circ \times 30^\circ$ of visual angle. The design and control of task timing and visual
130 stimulus presentation were executed with networked computers running custom written (Real-
131 time Experimentation and Control, REX (Hays, Richmond et al. 1982)) and commercially
132 available (Presentation, Neurobehavioral Systems) software.

133 Training for Experiment 1:

134 Monkeys were initially trained to grasp and release a touch sensitive bar to earn water
135 rewards. After this initial shaping, a red/green color discrimination task was introduced.
136 Red/green trials began with a bar press, and 100 ms later a small red target square (0.5°) was
137 presented at the center of the display (over-laying a white noise background). Animals were
138 required to continue grasping the touch bar until the color of the target square changed from red
139 to green, this occurred randomly between 500–1,500 ms after bar touch. Rewards were delivered
140 if the bar was released between 200–1,000 ms after the color change; releases occurring either

141 before or after this epoch were counted as errors. All correct responses were followed by visual
142 feedback (target square color changed to blue) after bar release and reward was delivered
143 between 200–400 ms after visual feedback. There was a 2 second inter-trial interval (ITI),
144 regardless of the outcome of the previous trial.

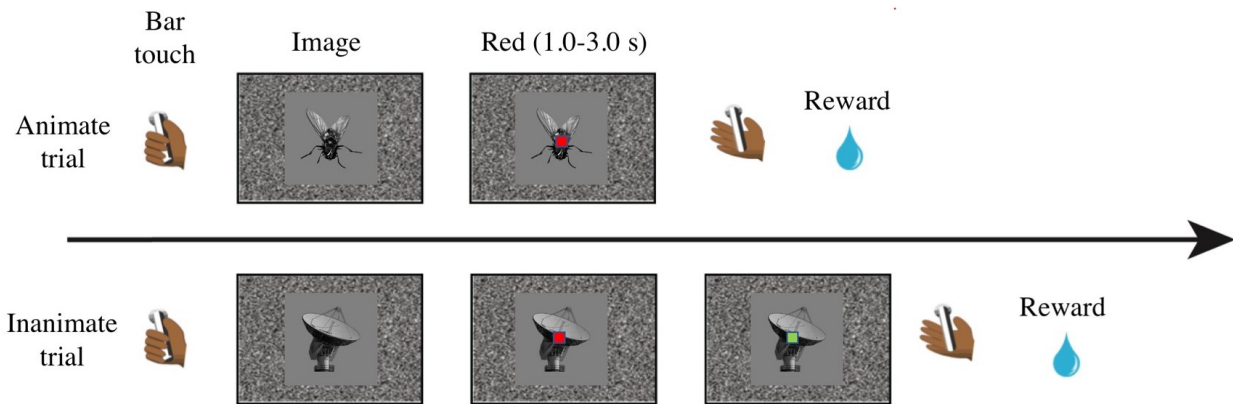
145 After each monkey reached criterion in the red/green task (two consecutive days with
146 >85% correct performance) a visual categorization task was introduced. Each trial began when
147 the animal grasped the touch bar. Next, an image (14° x 14°) appeared at the center of the
148 screen, followed by a red cue over the center of the image. When the image presented was
149 animate, the monkey had to release the bar before the red cue turned green to receive a liquid
150 reward. When it was an inanimate trial, the monkey had to continue to hold the bar until the red
151 cue turned green and then release the bar to receive a liquid reward (Figure 2). The red cue was
152 displayed on the screen for 1-3 seconds before turning green in inanimate trials. If the monkey
153 released during the red target and an inanimate image was presented, no reward was delivered,
154 and the image was displayed on the screen for a 4–6 second time-out. If the monkey did not
155 release during the inanimate image presentation within 1000 ms after the red target turned green,
156 no reward was delivered and there was a 3 second time-out.

157 If an equal drop size was used as reward for both conditions, monkeys would tend to
158 favor a release on red because of the delay discounting effect when waiting for green. Therefore,
159 the number of reward drops delivered for correct responses to red or green was adjusted during
160 the training phase to reduce the bias in responding to each category for each animal. As such, the
161 drop ratio for correct animate vs. correct inanimate trials was 1: 7 for monkey 1(M1), 1: 6 for
162 monkey 2 (M2), and 1: 9 for monkey 3 (M3). Each monkey was trained on a repeated set of 20
163 animate and 20 inanimate images for several days until their choice accuracy reached above 85%
164 accuracy for two consecutive days.

165 Testing for Experiments 1 and 2:

166 During the testing phase of Experiment 1, monkeys were tested on trial-unique sets of
167 100 novel animate and 100 novel inanimate intact images for 3 (M1) or 5 days (M2 and M3).
168 After the third testing day on classifying intact images into animate and inanimate categories,
169 M1 reached an accuracy of 91%. Due to this clear demonstration of high performance
170 categorizing intact images, we stopped testing M1 on intact images and moved onto testing
171 classification of synthesized images. Crucially, the training images were never shown in the

172 testing sets, and on each testing day, monkeys were presented with a new set of unfamiliar
173 images. Immediately after Experiment 1, monkeys were moved to Experiment 2, in which they
174 were tested on trial-unique sets of 100 synthesized animate and 100 synthesized inanimate
175 images (Figure 1c and 1d) for 5 days (M1, M2, M3).



176
177 Figure 2: Experimental procedure. Each trial began when the animal grasped the touch bar. An
178 image appeared at the center of the screen, followed by a red cue over the center of the image.
179 When the image presented was animate, the monkey had to release the bar within 3 seconds of
180 the appearance of the red cue to receive a liquid reward. When it was an inanimate trial, the
181 monkey had to continue to hold the bar until the red cue turned green to and then release the bar
182 to receive a liquid reward. The red cue was displayed on the screen for 1-3 seconds before
183 turning green in inanimate trials.

184

185 Classification analyses:

186 The statistical significance of classification accuracy was evaluated for each monkey
187 individually using a permutation test. For each monkey, we created a vector comprised of his
188 responses on each trial (animate or inanimate), which we labeled as V_r , and an additional vector
189 comprised of values representing the actual category of a trial (animate or inanimate), which we
190 labeled as V_c . We then shuffled both the order of V_r and V_c . Then, for each row of the vectors,
191 if the value in V_r matched that of V_c , we labeled that trial as correct and if not, as incorrect.
192 Using this method, we calculated the overall accuracy (% correct irrespective of category), the
193 accuracy for the animate category (% of animate trials correctly classified) and the accuracy for
194 the inanimate category (% of inanimate trials correctly classified). The shuffling procedure was
195 repeated 10,000 times for each monkey and for each permutation, we recorded these three

196 accuracy values. At the end of the 10,000 permutations, each monkey had his own chance
197 distributions (with 10,000 data points each), representing overall accuracy. Using these chance
198 distributions, we evaluated the significance of each monkey's actual mean classification
199 accuracy.

200

201 Reaction time:

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

206

207 Quantifying the amount of curvilinear and rectilinear information of the synthesized stimuli:

208 After matching the stimuli on size, background, mean luminance and contrast, we
209 calculated the amount of curvilinear and rectilinear information present in each image using a
210 method presented previously in Zachariou et al. (2018) and Yue et al. (2014, 2020) (see
211 Supplementary Materials for a detailed description).

212

213 Logistic regression of monkeys' performance with trial numbers:

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

226

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

229 To determine whether and the extent to which the amount of intermediate image features
230 (such as curvilinear and rectilinear image features) presented in Experiments 1 and 2 contribute
231 to monkeys' performance, we conducted a logistic regression of monkeys' performance (right or
232 wrong) with the curvilinear and rectilinear values of our visual stimuli (Yue et al. 2014;
233 Zachariou et al., 2018). The trials in which monkeys failed to respond were excluded from the
234 analysis.

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

244 To determine the contribution of the amount of intermediate visual features to the
245 monkeys' performance, we used raw responses in a logistic regression instead of average
246 response accuracies per stimulus in a linear regression for two reasons. First, to avoid over-
247 estimating the influence of stimuli that only one monkey responded to, and second, to avoid
248 creating artificially continuous responses with averaging because responses were discrete.

249

250 Deep convolutional neural network (DCNN) training and correlation analysis:

251 The DCNN, AlexNet (Krizhevsky et al, 2012), was imported into MATLAB, and pre-
252 trained on the ImageNet database (Deng et al., 2009). All pre-trained weights in the first 22
253 layers were kept the same, while the last three layers—fully connected layer, SoftMax layer, and
254 classification layer—were trained to classify each intact image into animate or inanimate
255 categories. The training was conducted on the 500 intact animate and 500 intact inanimate
256 images used in Experiment 1, using the stochastic gradient descent with momentum optimizer,
257 minimum batch size 64, maximum epochs 20, and an initial learning rate of 10^{-4} . After 300

258 iterations, the neural network performance converged on an accuracy of 99.9%. Then the trained
259 neural network was tested to classify the same 1000 synthesized images used in Experiment 2
260 into either the animate or inanimate category.

261 To compute the correlation of the DCNN classification accuracies and monkeys' response
262 accuracies to the synthesized images in Experiment 2, we arranged the responses of the DCNN
263 and each monkey according to the ascending order of curvilinear values of the synthesized
264 images presented in each trial. The ordered responses were then grouped into 40 bins. The
265 monkeys' accuracies used for the correlation analysis were averaged across all three animals.
266 Next, the response accuracy for each bin was calculated for the DCNN and monkeys, resulting in
267 two sets of 40 data points. The significance of the correlation was assessed by a permutation test
268 (10,000 iterations).

269

270

271

272

273

Results

274 Experiment 1: Intact images

275 1) *Overall classification accuracy for individual monkeys*

276 During the testing phase of Experiment 1, in which novel intact images were used for the
277 categorization task, each image was presented only once regardless of the monkeys' responses.
278 This eliminated the option of memorizing test images to perform the task. Across five days of
279 testing, all monkeys performed the task significantly above chance (overall accuracy for M1:
280 80.88%, $p < 0.0001$; M2: 78.38%, $p < 0.0001$; M3: 76.95%, $p < 0.0001$). The statistical
281 significance was determined by the permutation test (see Methods). The overall response rate
282 was 99.64% for M1, 73.43% for M2, and 98.86% for M3.

283 Upon closer inspection of the data we found that M2 memorized all 40 training images to
284 perform the categorization task. Thus, in the first day of testing, M2 was learning the
285 categorization task. After eliminating data from this day, overall performance was 85.64% ($p <$
286 0.001), and overall response rate was 73.3%. Unless stated otherwise, subsequent analyses used
287 M2's testing data from day 2 to day 5 only. Data from all five days of testing are included in
288 Supplementary Figure 1.

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

293

294 2) *Generalization and learning effect for individual monkeys:*

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

304 regression model. We anticipated that, if there were a generalization effect, then the intercept of
305 the logistic regression model would be significantly greater than zero, and if there were a
306 learning effect, then the slope of the regression model would be significantly greater than zero.

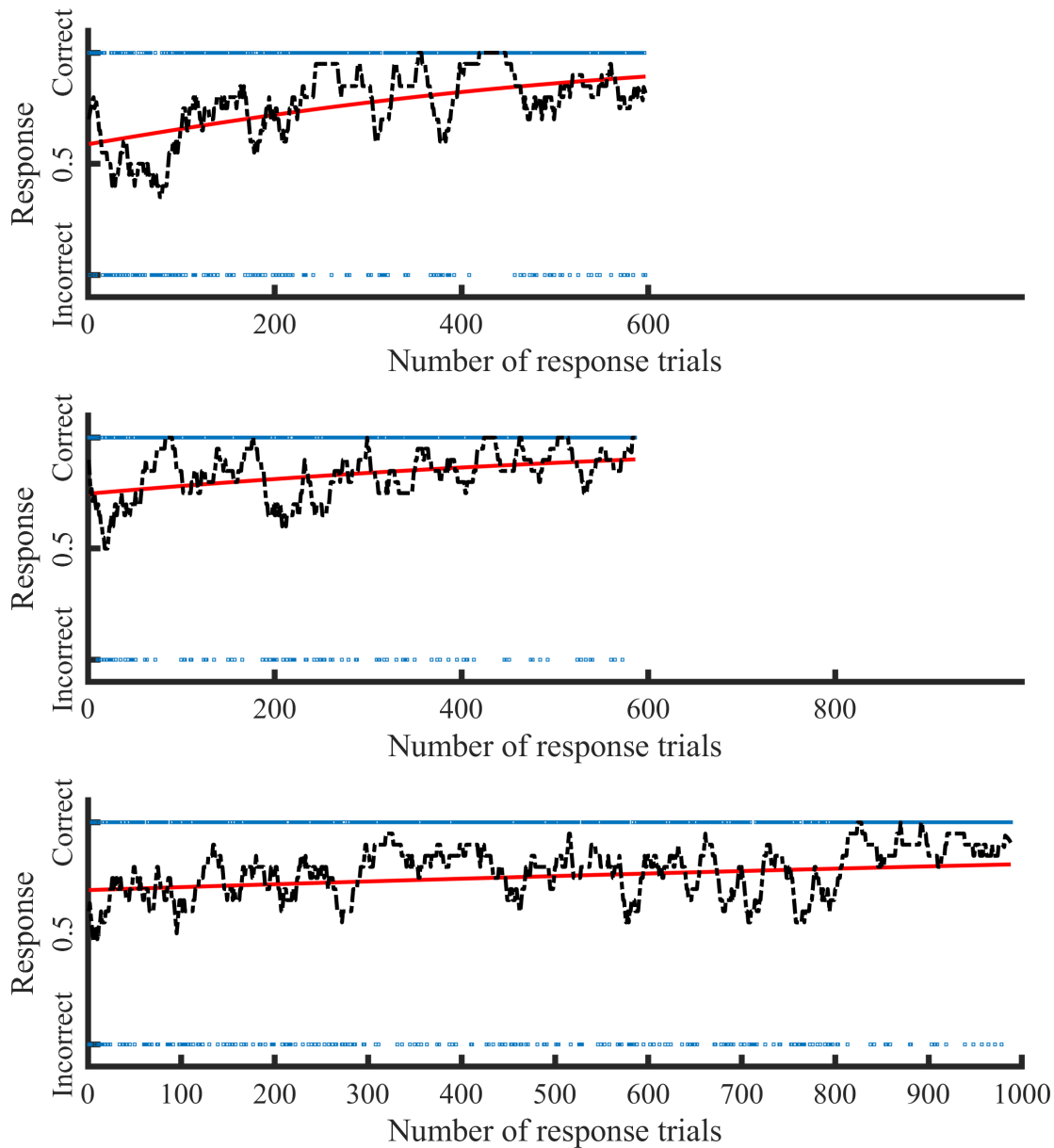
307 Monkeys were able to use the information they learned during training to perform the
308 categorization task on unfamiliar images at the onset of the testing phase, as shown in Table 1,
309 where the intercept of the logistic regression is shown to be significantly above chance for all
310 three monkeys. The slope of the logistic regression was positive and significantly different from
311 zero in all monkeys, indicating that performance improved as testing progressed. All three
312 monkeys' performance was significantly predicated by trial number, as shown in Figure 3 and
313 Table 1 (for M1: $\chi^2(595) = 58.545, p = 1.98 \times 10^{-14}$; M2: $\chi^2(584) = 18.361, p = 1.828 \times 10^{-5}$; M3:
314 $\chi^2(986) = 13.252, p = 2.72 \times 10^{-4}$), further indicating that monkeys continued to learn during the
315 testing phase, improving their performance even though each image was presented only once.

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

321

322 Table 1. Logistic regression results from Experiment 1.

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



323

324 Figure 3. The logistic regression results of Experiment 1 for M1 (top), M2 (middle), and M3
325 (bottom). The x-axis represents the number of response trials (trials without responses were
326 removed), and the y-axis represents the monkey's response. The monkeys' responses for each
327 trial are shown as blue dots, which appears as a blue line because of the large number of trials.
328 The red line represents the predicted response probability produced from the logistic regression
329 analysis. The black dotted line represents the response accuracy of a moving average of 20 trials,
330 which is for illustration purposes only and not used for calculating logistic regression. The
331 intercepts of the regression lines for all three monkeys are larger than 0.5, indicating that all three
332 monkeys were able to generalize from the training set to the testing set. The regression line
333 increased along with the trial numbers, suggesting that monkeys continued to learn during the
334 testing phase to improve their performance. M1 was tested only for three days; therefore, it has
335 only 600 trials. M2 was tested for five days, but data from the first day were removed from the
336 logistic regression.

337

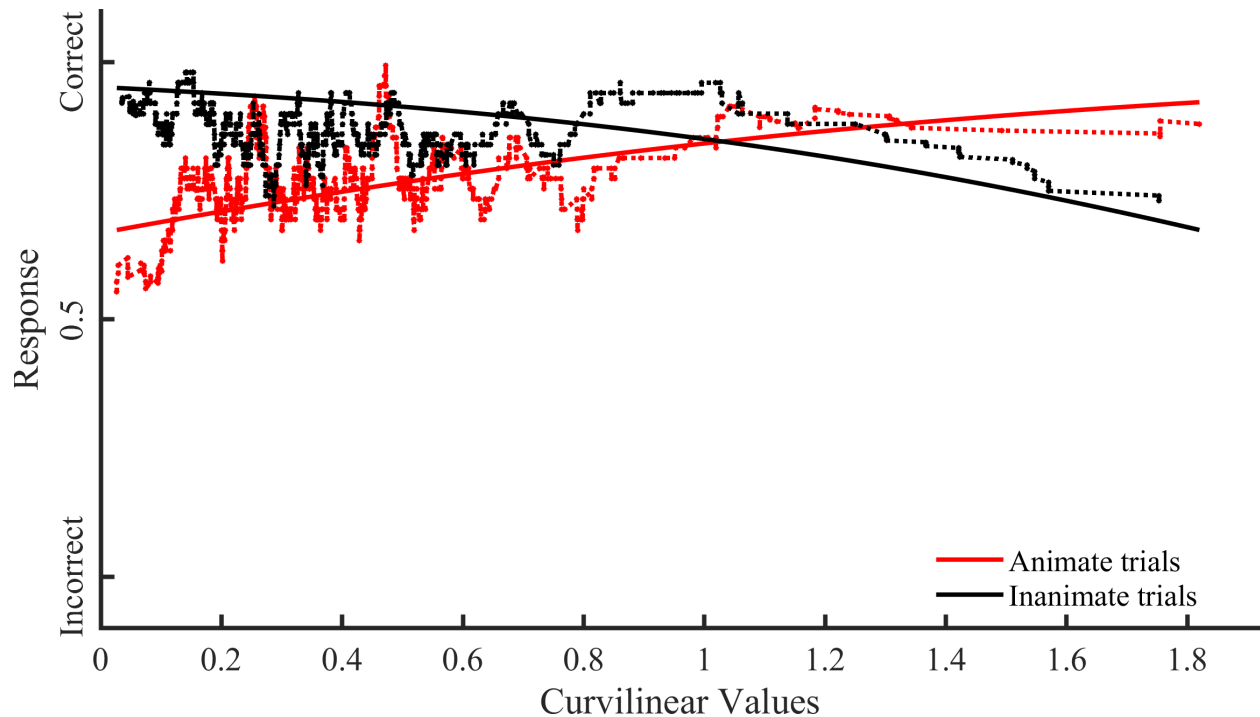
338 3) Contribution of curvilinear and rectilinear features to monkeys' performance at the group
339 level.

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

346 We found that the amount of intermediate image features in the intact images
347 significantly predicted monkeys' performance (main effect: $\chi^2(2768) = 107.4, p = 1.450 \times 10^{-21}$),
348 suggesting that the amount of intermediate image features might assist them in categorizing
349 intact images into animate and inanimate groups. Furthermore, we found that curvilinear values
350 of intact images significantly predicted monkeys' performance (beta = 0.974, $p = 0.031$), but
351 rectilinear values did not (beta = -0.4817, $p = 0.272$). There was a significant interaction
352 between the curvilinear values and the stimulus category (beta = -2.21, $p = 1.118 \times 10^{-4}$),
353 indicating that curvilinear values predicted monkeys' performance in animate trials differently
354 than on inanimate trials. Figure 4 shows the functional relationship between curvilinear values
355 and monkeys' performance across animate and inanimate trials, which was produced from the
356 logistic regression model. As the amount of curvilinear information in an image increased,
357 monkeys' performance increased for animate images and decreased for inanimate images.

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

362



363
364 Figure 4. Functional relationship between the amount of curvilinear information present in visual
365 stimuli and monkeys' performance across stimulus category in Experiment 1. The x-axis
366 represents the curvilinear values of the stimuli. The y-axis represents the response probability of
367 the monkeys' performance. The solid lines represent the response probability to visual stimuli
368 calculated with the logistic regression model that was created using the monkeys' group raw
369 response. The dotted lines represent a moving average of 60 trials, which is for illustration
370 purposes only and was not used for fitting the logistic regression model. The red line represents
371 the response probability resulting from the logistics regression fitting for the animate trials. The
372 black line represents the response probability resulting from the logistics regression fitting for
373 the inanimate trials

374

375 Experiment 2: Synthesized images

376 1) Overall classification accuracy for individual monkeys:

377 The monkeys were never trained to categorize the synthesized images presented in
378 Experiment 2. Furthermore, the synthesized images were each shown only once, regardless of
379 the monkeys' responses. As shown in Figure 3B, all three monkeys performed the categorization
380 task significantly above chance (overall accuracy for M1, 64.48%, $p < 0.0001$; M2, 59.10%, $p <$
381 0.0001; M3, 60.27%, $p < 0.0001$). The overall response rate was 99.6% for M1, 92.7% for M2,

382 and 85.1% for M3. Although the overall classification accuracies were lower than those for the
383 intact images in Experiment 1, the significant above-chance performances suggest that the image
384 features distinguishing the two groups of synthesized images provided sufficient information for
385 monkeys to classify the images into the two categories.

386

387 2) *Generalization and learning effect for individual monkeys:*

388 To provide a parallel analysis to the one performed in Experiment 1, we ran a logistic
389 regression to evaluate if the monkeys' overall accuracies for categorizing the synthesized images
390 resulting from generalizing visual features learned from the intact images to the synthesized
391 images and/or continuous learning. We found that the intercept, but not the slope, of the logistic
392 regression model was significant for all three monkeys, as shown in Table 2. Performance was
393 not significantly determined by test trial number for any monkeys (for M1: $\chi^2(994) = 0.365$, $p =$
394 0.546 ; M2: $\chi^2(925) = 0.340$, $p = 0.560$; M3: $\chi^2(849) = 0.032$, $p = 0.859$), indicating that
395 monkeys' performance did not improve as testing progressed. These results reveal that, at the
396 onset of Experiment 2, all three monkeys used information they learned on the categorization
397 task in Experiment 1 to classify the synthesized images as animate and inanimate objects.

398

399

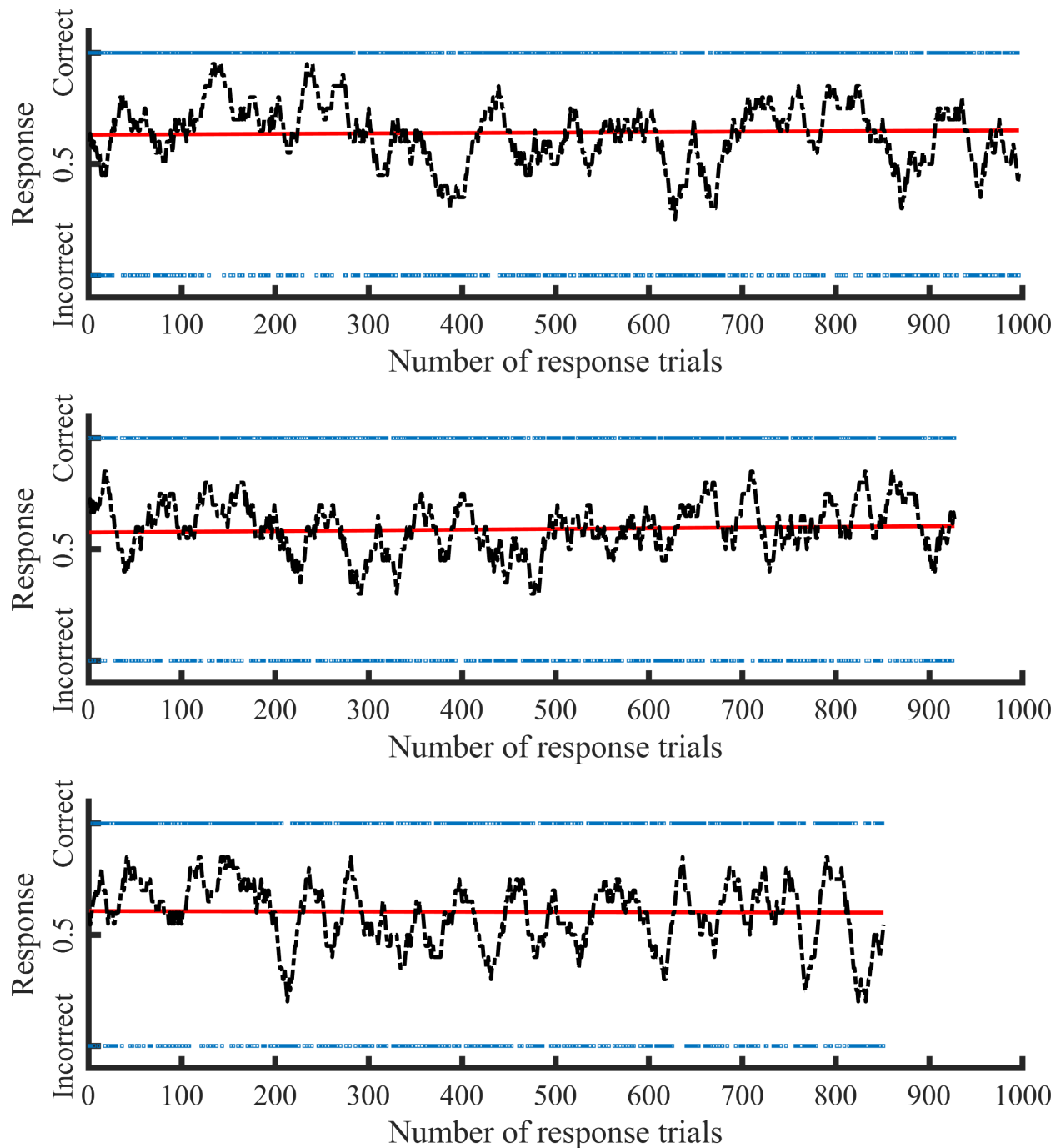
Table 2. Logistic regression result of Experiment 2.

Monkeys	Logistic regression	
	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} ($p = 0.919$)

400

401

402



403

404 Figure 5: The logistic regression results of Experiment 2 for M1 (top), M2 (middle), and M3
405 (bottom). Axes are the same as those used in Figure 3. As shown in Table 2, all three monkeys
406 showed significant generalization but no learning effects. These results suggest that the monkeys
407 employed some image features distinguishing intact animate images from intact inanimate
408 images to categorize the synthesized images as animate or inanimate.

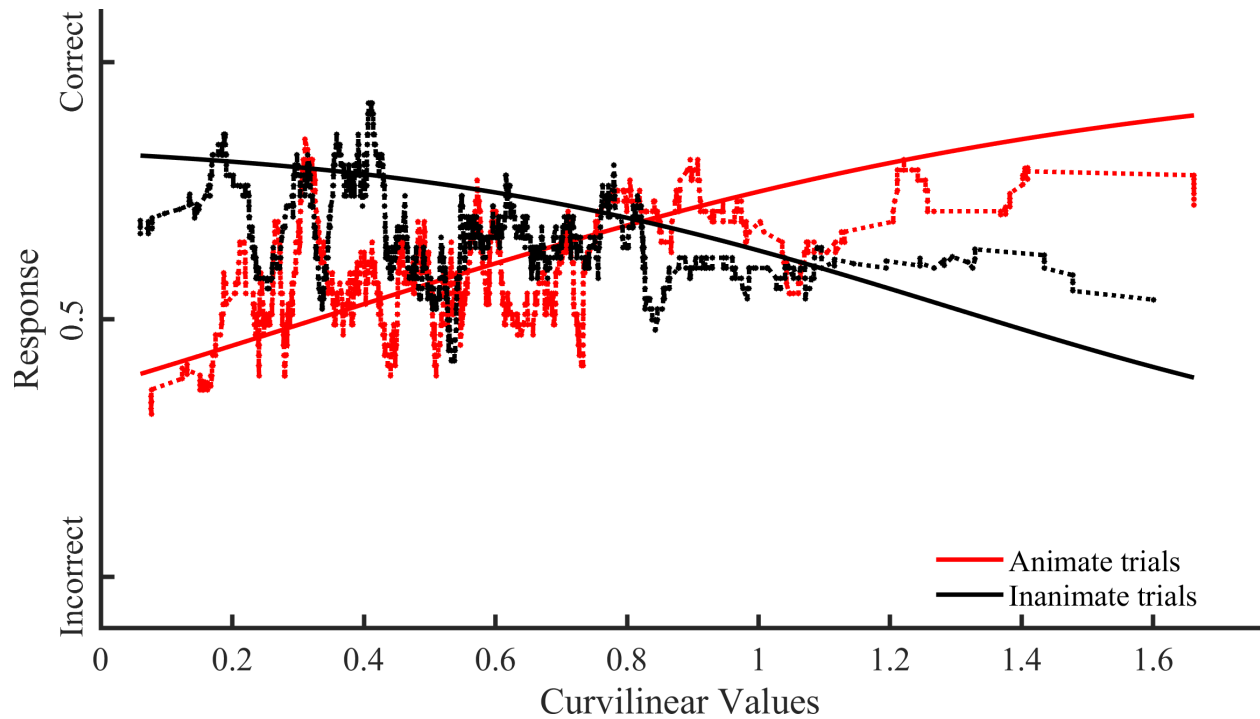
409

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

412 To examine the extent to which the amount of intermediate visual features contributed to
413 monkeys' performance in Experiment 2, we used the same testing procedure as Experiment 1 but
414 with synthesized images.

415 We found a significant main effect of the amount of curvilinear and rectilinear image
416 features on monkeys' performance ($\chi^2(2768) = 177.160, p = 2.160 \times 10^{-36}$). Furthermore, both
417 curvilinear and rectilinear values of synthesized images significantly predicted monkeys'
418 performance (curvilinear: $\beta = 1.617, p = 2.615 \times 10^{-7}$; rectilinear: $\beta = -1.257, p = 5.865 \times$
419 10^{-4}). However, the data suggested that the amount of curvilinear image features present in the
420 synthesized images played a more dominant role than the amount of rectilinear image features.
421 To test this hypothesis, we performed a regression Wald test to examine whether the curvilinear
422 coefficient was significantly different from the rectilinear coefficient. The curvilinear coefficient
423 was significantly larger than the rectilinear coefficient (Wald test: $\chi^2(1) = 19.938, p = 7.994 \times 10^{-$
424 6), indicating that the amount of curvilinear image features present in the synthesized images was
425 more informative for the categorization task than the amount of rectilinear image features. As
426 such, the following analysis of interaction between the amount of intermediate image features
427 and stimulus category was focused on the contribution of the amount of curvilinear image
428 features on monkeys' performance across stimulus categories. Results of the analysis of the
429 interaction effect between the amount of rectilinear image features with stimulus category are
430 shown in Supplementary Figure 2.

431 We observed a significant interaction between the curvilinear values of stimuli and
432 stimulus category ($\beta = -4.040, p = 1.672 \times 10^{-20}$). Monkeys' performance on synthesized
433 images increased when curvilinear values increased in the animate trials but decreased in the
434 inanimate trials (Figure 6); similar to what we observed in Experiment 1 (Figure 4). These data
435 indicate that the more curvilinear information present in an animate image, the more likely it was
436 to be categorized correctly, whereas the opposite is true for inanimate images.

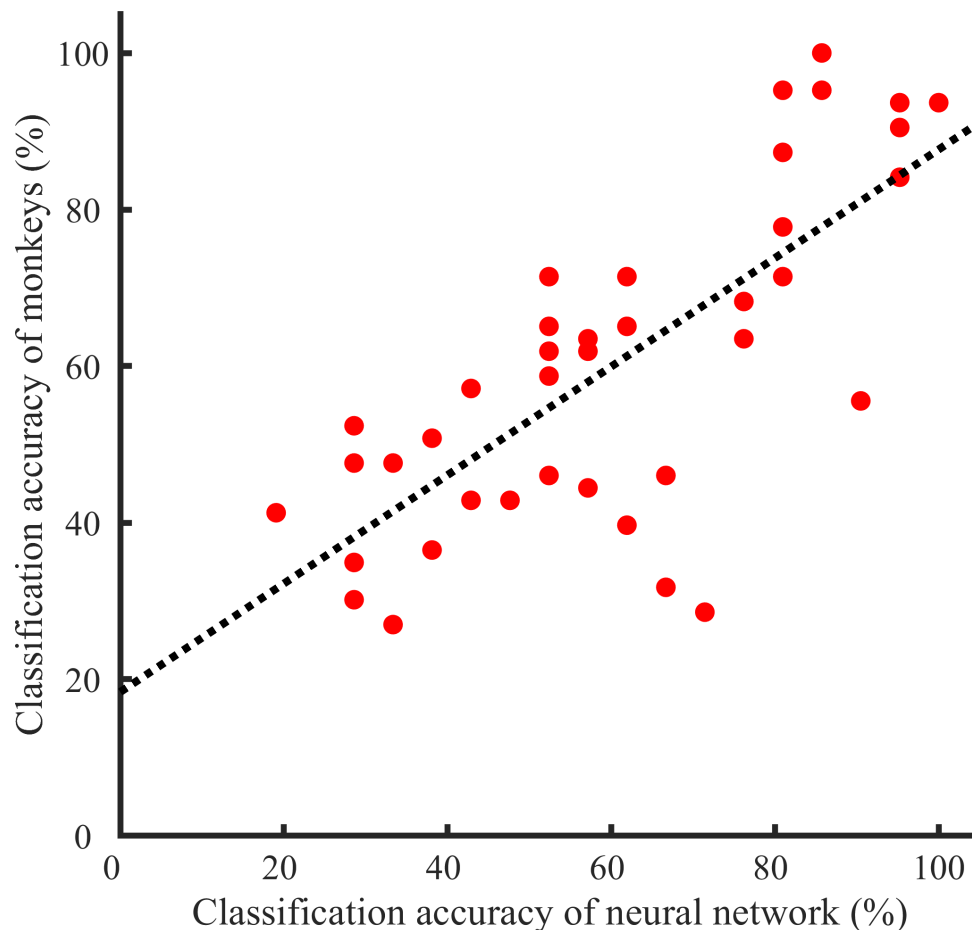


437
438 Figure 6. Functional relationship between amount of curvilinear information present in the visual
439 stimuli and monkey's group performance across stimulus category in Experiment 2. The x-axis
440 represents the curvilinear values of visual stimuli. The y-axis represents the response probability
441 of the monkeys' performance. The solid lines represent the response probability to visual stimuli
442 calculated with the logistic regression model that was created using the monkeys' group raw
443 response. The dotted lines represent a moving average of 60 trials, which is for illustration
444 purposes only. The red line represents the response probability resulting from the logistics
445 regression fitting for the animate trials. The black line represents the response probability
446 resulting from the logistics regression fitting for the inanimate trials.

447
448 4) *Correlation of monkeys' performance with DCNN performance at the group level*

449 Because monkeys were never trained to classify synthesized images into animate and
450 inanimate categories, the possibility remained that monkeys categorized the images into two
451 groups using differences between synthesized images that were entirely unrelated to the animate
452 and inanimate category but happened to coincide with the two categories in the set of testing
453 images used. As such, we used the DCNN to address this concern (see Methods). The network
454 was trained to classify the 1000 intact images used in Experiment 1 into animate and inanimate
455 categories and then tested on the categorization task with the 1000 synthesized images used in

456 Experiment 2 (see Methods). We found a significant positive correlation of the DCNN's
457 categorization performance with the monkeys' group performance ($r = 0.739$, $p = 5.0502 \times 10^{-8}$)
458 (Figure 7), suggesting that the monkeys performed the animate vs. inanimate categorization in
459 Experiment 2, when the global form in the images was distorted beyond recognition. These data
460 provided further evidence that the monkeys used image features distinguishing intact animate
461 and inanimate images to categorize the synthesized images.
462



463
464 Figure 7: Correlation of monkeys' response accuracies with DCNN classification accuracies.
465 To compute the correlation of the DCNN classification accuracies and monkeys' response
466 accuracies to the synthesized images, we arranged the responses of the DCNN and each monkey
467 according to the ascending order of curvilinear values of the synthesized images. The monkeys'
468 accuracies used for the correlation analysis were averaged across all three animals. The ordered
469 responses were then grouped into 40 bins. Next, the response accuracy for each bin was
470 calculated for the DCNN and monkeys separately, resulting in two sets of 40 data points. Each
471 red dot represents the classification accuracy for each bin. We observed a significant correlation
472 between monkeys' response accuracies and DCNN classification accuracies ($r = 0.739$, $p =$
473 5.0502×10^{-8}), indicating that monkeys performed the animate vs. inanimate categorization.

474

Discussion

475 This study investigated the contributions of both training and image-based features to the
476 perceptual categorization of animacy. In Experiment 1, we found that naïve monkeys trained to
477 categorize a small set of animate and inanimate images classified a large set of unfamiliar images
478 into animate and inanimate categories with high accuracy. In Experiment 2, we tested whether
479 image-based features that differ between the two object categories in the statistics of natural
480 environments, i.e. curvilinear and rectilinear information (Kurbat, 1997; Levin et al. 2001;
481 Perrinet and Bednar, 2015; Long et al. 2017; Zachariou et al., 2018), determined the monkeys'
482 classification accuracy. We created sets of synthetic animate and inanimate images using an
483 algorithm that significantly distorted the global shape of the original images while maintaining
484 the original images' intermediate features (Portilla and Simoncelli, 2000). The monkeys'
485 classification accuracy on these synthesized images was still significantly above chance and
486 correlated with the amount of curvilinear information present in the stimuli. These data indicate
487 that image-based features, in this case curvilinearity, can be used to distinguish animate from
488 inanimate objects in the absence of global shape information without prior training.

489 As monkeys raised in the laboratory have limited experiences with objects that humans
490 are otherwise familiar with, they are ideal candidates to study the contribution of experiences and
491 image-based features to the emergence of perceptual categorization (e.g. Arcaro & Livingstone,
492 2017). Our results show that monkeys performed an animacy categorization task with intact
493 images significantly above chance at the very beginning of the test phase of Experiment 1,
494 suggesting that monkeys used what they had learned during training to classify novel images of
495 objects, with which they had no previous experience, into animate and inanimate categories.
496 Further, the curvilinear values of intact images had a significant interaction with stimulus
497 category, and significantly predicted the monkeys' performance. These findings indicate that
498 image-based features that are predictive of each category provide substantial information that
499 monkeys can use to distinguish the two categories with little training. In other words, experience
500 interacting with objects may not be the only origin of behavioral categorization of animacy in
501 monkeys.

502 To confirm this, using the synthesized images in Experiment 2, we eliminated local
503 features (faces, ears, etc.) that monkeys might have been familiar with and could have used to
504 classify the images into animate and inanimate categories. We found that the monkeys were able

505 to perform the categorization of the synthesized images significantly above chance, which
506 indicates that the image-based features were sufficient for the emergence of perceptual
507 categorization. It is worth noting that human participants also classified synthesized images
508 similar to those used in this experiment into animate and inanimate categories with significant
509 above-chance accuracy (Zachariou, et al., 2018; Long et al., 2017). Although humans and
510 monkeys do not share the collective experience of what and how objects are encountered in daily
511 life, they perform similarly when classifying synthesized images into animate and inanimate
512 categories (Figure 6, Figure 3 in Zachariou, et al., 2018), which suggests that image-based
513 feature differences could play a critical role in the emergence of perceptual categorization
514 abilities across species. Together, our findings provide strong evidence in support of the
515 hypothesis that perceptual categorization can emerge from image-based features that are
516 predictive of each category in the natural statistics of the visual environment.

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

523 Overall, monkeys categorized the intact object images with significantly greater accuracy
524 than the synthesized images. However, for synthesized images with high curvilinear values (in
525 the range of 1.4 – 1.6), monkeys' classification accuracy for the animate category could reach
526 above 80% which is comparable to the classification accuracy for intact images (Figure 6). This
527 illustrates that monkeys could achieve high accuracy when synthesized images with extreme
528 curvilinear values were used as stimuli. Thus, the overall difference in classification accuracy
529 between the intact and synthesized images does not argue against the idea that image-based
530 features play a significant role in determining perceptual categorization.

531 The primate visual system takes significant time to fully mature postnatally (Gilmore et al.,
532 2018; Ellemberg et al., 1999; Kovacs et al., 1999). During development, young infants view the
533 world as consisting not of coherent objects but instead visual pieces that move in unpredicted
534 ways (Hyvärinen, et al., 2014). In such a fragmented visual world, differentiating animate from
535 inanimate objects would be challenging. Infants who can differentiate animate from inanimate

536 objects would have a better chance to avoid being harmed by animals to survive than those who
537 cannot. Through natural selection, our brains may have evolved the capacity to differentiate
538 animate and inanimate objects quite quickly, first based on sensory information that represents
539 visual statistics of the natural environment. Experience with objects would play a significant
540 role in later life to further differentiate categories. Our data provide evidence to support this
541 hypothesis by showing that monkeys (as well as humans (Zachariou, et al., 2018)) are able to
542 classify synthesized images that: 1) neither species has experience with; and 2) have similar
543 statistics as the natural original images, into animate and inanimate categories significantly
544 above chance by using the degree of curvilinearity in the images. This hypothesis raises many
545 interesting questions. For which object categories and with which image features is the primate
546 brain biased to use image-based differences for perceptual categorization, and under what
547 conditions? The answers to such questions are critical to understand the functional and
548 anatomical organization of the primate visual system.

549

550

551

552 **Acknowledgments**

553 This work was supported by the NIMH Intramural Research Program.

554

555 **Conflict of Interest**

556 The authors declare no competing financial interests.

557

558

559

References

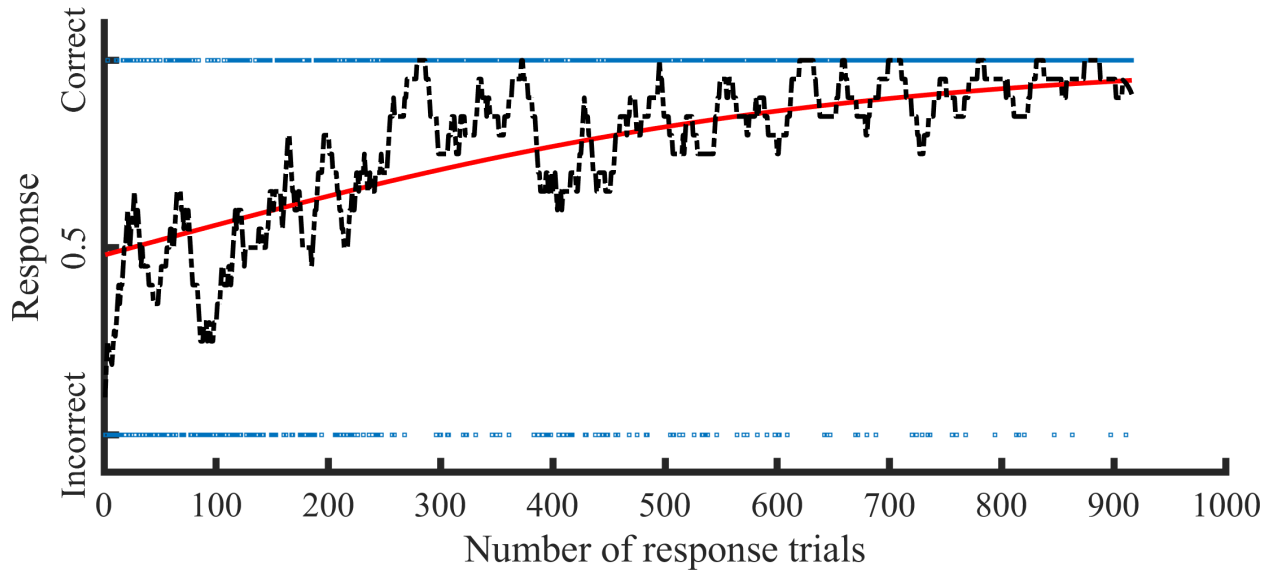
- 560 Arcaro M. J., Livingstone M. S. (2017) A hierarchical, retinotopic proto-organization of the
561 primate visual system at birth. *eLife* 6:e26196.
- 562 Blake, C. E., Bisogni, C. A., Sobal, J., Devine, C. M., & Jastran, M. (2007). Classifying foods in
563 contexts: how adults categorize foods for different eating settings. *Appetite*, 49(2): 500-
564 510.
- 565 Bovet, D., & Vauclair, J. (1998). Functional categorization of objects and of their pictures in
566 baboons (*Papio anubis*). *Learn Motiv*, 29(3): 309-322.
- 567 Calvillo, D. P., & Hawkins, W. C. (2016). Animate objects are detected more frequently than
568 inanimate objects in inattentive blindness tasks independently of threat. *J Gen Psychol*,
569 143(2): 101-115.
- 570 Chiara, T., Bulf H., & Simion, F. (2008). Newborns' face recognition over changes in viewpoint.
571 *Cognition*, 106: 1300-1321.
- 572 Deng, J., Dong, W., Socher, R., Li, L-J., Li, K., & Fei-Fei, L.(2009). Imagenet: A large-scale
573 hierarchical image database. In 2009 IEEE *CVMP* (pp. 248-255).
- 574 Elleberg D, Lewis TL, Liu CH, Maurer D. (1999) Development of spatial and temporal vision
575 during childhood. *Vision Res.* 39(14):2325–2333.
- 576 Gilmore JH, Knickmeyer RC, Gao W. (2018) Imaging structural and functional brain
577 development in early childhood. *Nat Rev Neurosci.* 19(3):123–137.
- 578 Hays, A. V., B. J. Richmond and L. M. Optican (1982). "A UNIX-Based Multiple-Process
579 System for Real-Time Data Acquisition and Control." *WESCON Conference*
580 *Proceedings*(2): 1-10
- 581 Heron-Delaney, M., Wirth, S., & Pascalis, O. (2011). Infants' knowledge of their own species.
582 *Philos T R Soc B*, 366(1571): 1753-1763.
- 583 Hyvärinen, L., Walther, R., Jacob, N., Chaplin, K.N., Leonhardt, M (2014). Current
584 understanding of what infants see. *Curr Ophthalmol Rep.* 2(4): 142-149.
- 585 Kalénine, S., Bonthoux, F., & Borghi, A. M. (2009). How action and context priming influence
586 categorization: a developmental study. *Brit J Dev Psychol*, 27(3): 717-730.
- 587 Kalénine, S., Shapiro, A. D., Flumini, A., Borghi, A. M., & Buxbaum, L. J. (2014). Visual
588 context modulates potentiation of grasp types during semantic object categorization.
589 *Psychon B Revi*, 21(3): 645-651.

- 590 Kovacs I, Kozma P, Fehér A, Benedek G. Late maturation of visual spatial integration in
591 humans. *Proc Natl Acad Sci U S A*. 1999;96(21):12204–12209
- 592 Krizhevsky, A., Sutskever, I., Hinton. GE. (2012) Imagenet classification with deep
593 convolutional neural networks. *Adv Neur In*, 1097-1105.
- 594 Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category:
595 Specifying the features that mark the difference between artifacts and animals in
596 preattentive vision. *Atten Percept Psycho*, 63(4): 676-697.
- 597 Long, B., Störmer, V. S., & Alvarez, G. A. (2017) Mid-level perceptual features contain early
598 cues to animacy. *J. Vis.*, 17:20-20.
- 599 Long. B, Yu, C. P., Konkle, T. (2018) Mid-level visual features underlie the high-level
600 categorical organization of the ventral stream. *Proc Natl Acad Sci USA* 115:E9015-9024.
- 601 LoBue, V., & DeLoache, J. S. (2011). What’s so special about slithering serpents? Children and
602 adults rapidly detect snakes based on their simple features. *Vis Cogn*, 19: 129–143.
- 603 Long, B., Moher, M., Carey, S. E., & Konkle, T. (2019). Animacy and object size are reflected in
604 perceptual similarity computations by the preschool years. *Vis Cogn*, 27(5-8): 435-451.
- 605 Lipp, O. V. (2006). Of snakes and flowers: Does preferential detection of pictures of fear-
606 relevant animals in visual search reflect on fear-relevance? *Emotion*, 6: 296–308.
- 607 Livingstone, M. S., Vincent, J. L., Arcaro, M. J., Srihasam, K., Schade, P. F., Savage, T. (2017).
608 Development of the macaque face-patch system, *Nat Commun*, 8:14897.
- 609 Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychol Rev*, 99(4), 587
- 610 Meyerhoff, H. S., Schwan, S., & Huff, M. (2014). Perceptual animacy: Visual search for chasing
611 objects among distractors. *J Exp Psychol Human*, 40(2): 702-717.
- 612 Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic
613 memory is tuned to animacy. *Curr Dir Psychol Sci*, 26(1): 22-27.
- 614 Opfer, J. E., & Gelman, S. A. (2011). Development of the animate-inanimate distinction. *The*
615 *Wiley-Blackwell handbook of childhood cognitive development*, 2: 213-238.
- 616 Perrinet, L. U., & Bednar, J. A. (2015). Edge co-occurrences can account for rapid categorization
617 of natural versus animal images. *Sci Rep*, 5: 11400.
- 618 Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of
619 complex wavelet coefficients. *Int J Comput Vision*, 40(1): 49-70.

- 620 Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard?
621 *PLoS Comput Biol*, 4(1): e27.
- 622 Rakison, D. H. (2003). Parts, motion, and the development of the animate–inanimate distinction
623 in infancy. *Early category and concept development*, 159-192.
- 624 Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the
625 newborn baby. *Proc Natl Acad Sci USA*, 105(2): 809-813.
- 626 Srihasam, K., Vincent, J.L., Livingstone, M.S. (2014) Novel domain formation reveals proto-
627 architecture in inferotemporal cortex. *Nat Neurosci*, 17:1776-1783.
- 628 Sugita Y. (2008) Face perception in monkeys reared with no exposure to faces. *Proc Natl Acad*
629 *Sci USA*, 105(1): 394-398.
- 630 Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization.
631 *Cognition*, 105(2): 362-379.
- 632 Wang, P. and D. Nikolic (2011). "An LCD Monitor with Sufficiently Precise Timing for
633 Research in Vision." *Front Hum Neurosci* 5: 85.
- 634 Yue, X., Pourladian, I.S., Tootell, R.B.H., Ungerleider, L.G. (2014) Curvature-processing
635 network in macaque visual cortex. *Proc Natl Acad Sci USA* 111: E3467-E3475
- 636 Yue, X., Robert, S., Ungerleider, L.G. (2020). Curvature processing in human visual cortex.
637 *NeuroImage*, submitted.
- 638 Zachariou, V., Del Giacco, A.C., Ungerleider, L.G., Yue, X, (2018) Bottom-up processing of
639 curvilinear visual features is sufficient for animate/inanimate object categorization. *J Vis*,
640 18:388-398.
- 641
- 642
- 643
- 644
- 645

646
647

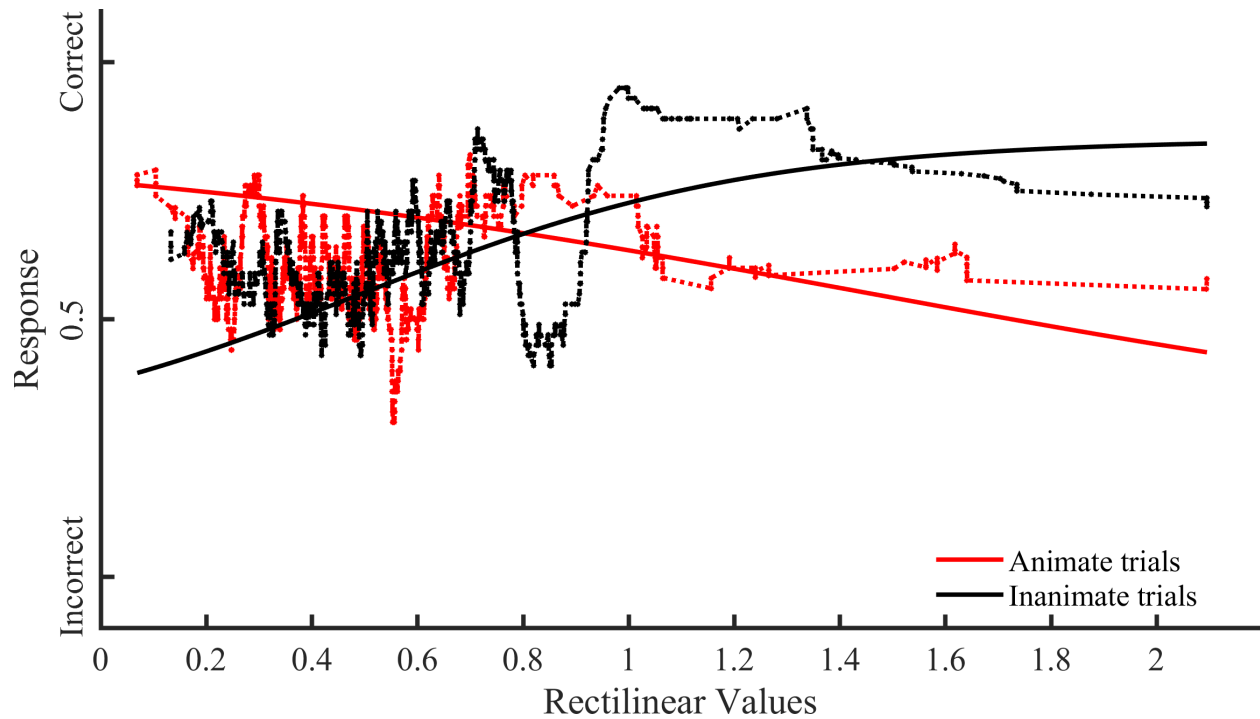
Supplementary Materials



648

649 Figure 1. The logistic regression results of the experiment 1 for M2. The x-axis represents the
650 number of response trials for all five days, and the y-axis represents the monkey's response. The
651 monkey responses for each trial are shown as blue dots, which appears as a blue line because of
652 the large number of trials. The red line represents the predicted response probability produced
653 from the logistic regression analysis. The black dot line represents the response accuracy of a
654 moving average of 20 trials, which is for illustration purposes only and not used for calculating
655 logistic regression.

656
657
658



659

660 Figure 2. Functional relationship between amount of rectilinear information present in the visual
661 stimuli and monkey's group performance across stimulus category in Experiment 2. The x-axis
662 represents the rectilinear values of visual stimuli. The y-axis represents the response probability
663 of the monkeys' performance. The solid lines represent the response probability to visual stimuli
664 calculated with the logistic regression model that was created using the monkeys' group raw
665 response. The dotted lines represent a moving average of 60 trials, which is for illustration
666 purposes only.

667

668

669