Novel binary categorization task to study high-level visual object classification in macaque monkeys

Han Zhang
(zhangh2022@ion.ac.cn)

Zhihao Zheng

Jiaqi Hu

Qiao Wang

Zixuan Li

Mengya Xu

Gouki Okazawa
(okazawa@ion.ac.cn)

Institute of Neuroscience,
Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology,
Center for Excellence in Brain Science and Intelligence Technology,
Chinese Academy of Sciences, Shanghai, China

Abstract

Humans associate visually perceived objects with abstract concepts, such as animate, mammalian, or artificial. This ability seems to depend on human language, but previous studies have reported that neural representations in higher visual areas of macaque monkeys also reflect some of these concepts. But can macaque monkeys categorize objects at this abstract level? Here, we developed a novel binary categorization task and found that monkeys quickly learned to classify images of natural objects based on abstract concepts, including animate versus inanimate, natural versus artificial, and mammalian versus non-mammalian. They generalized the learned rule to new images and made errors consistent with human classification. Since their choices could be well fit by artificial neural networks, we interpret that they could solve the tasks by extracting higher-order visual features. Our behavioral paradigm is well suited to study the monkey's capacity to visually categorize objects using various rules and stimulus sets.

Keywords: monkey behavior; object recognition; perceptual decision making; abstract concepts

Introduction

Humans classify visual objects in the external world into many basic-level categories such as dogs and tables, but they also recognize abstract concepts associated with them such as animate, artificial (man-made), and mammalian (i.e., superordinate categories). However, neurophysiological investigation of its underlying processes has been limited, probably because these concepts are thought to be tied to human language and thus beyond the scope of pure visual mechanisms (Murphy, 2004). Indeed, many monkey studies have focused on rather specific object identification problems as a major goal of the ventral visual pathway (DiCarlo et al., 2012; Rajalingham et al., 2018). Artificial neural networks (ANNs) have also typically been trained on concrete object categorization; the goal of the ImageNet challenge was to classify images into 1,000 categories.

Are abstract object concepts, then, only available through human language? Separate lines of monkey research suggest otherwise. Representational similarity analysis of neural responses in the monkey IT cortex has revealed a clear segregation of animate and inanimate objects (Kiani et al., 2007; Kriegeskorte et al., 2008), potentially indicating that this is an important conceptual boundary even for the monkey brain. A key question, however, is whether macaque monkeys can actually use such abstract distinctions to guide their behavior. A series of studies by Fabre-Thorpe and others have shown that monkeys can successfully detect the presence of an animal in a rapid sequence of images (Fabre-Thorpe et al., 1998; Fize et al., 2011). Other studies trained monkeys to categorize tree vs. non-tree (Vogels, 1999), cat vs. dog (Setogawa et al., 2021), car vs. truck (Minamimoto et al., 2010), food vs. others (Santos et al., 2001), and human vs. monkey (Roberts & Mazmanian, 1988) among others. However, these studies were often limited to specific categories or used a Go/No-go design to test the detection rather than the classification of objects. Moreover, animal training often requires many weeks, raising the question of whether their behavior is a product of repeated feedback. Here, we developed a novel binary categorization task and found that monkeys learn to classify images according to abstract concepts within a few days.

Results

Monkeys rapidly learned categorization

We developed a version of the binary decision-making task implemented in a touchscreen system installed on the monkey's home cage. In each trial, an object stimulus (natural photograph) and two target boxes were presented, and the monkey had to move the object to one of the two boxes according to a hidden rule learned through juice reward feedback (Fig. 1a). We reasoned that direct contact with object images would encourage monkeys to make associations quickly. The images were grayscale and cropped from the background from natural photographs of various objects, including humans, mammals, reptiles, plants, food, everyday tools, and electronics (Fig. 1b). For each task, we used a fixed stimulus set (~ 80 images) for training and then introduced a new set (~ 80) to test the generalization.

To our surprise, monkeys (n = 3) learned the tasks quickly, reaching approximately 90% correct (chance level: 50%) for the training set in 3-4 days (Fig. 1c; 600-1000 trials per day). They showed similar learning curves in various tasks tested sequentially, including animate vs. inanimate, natural vs. artificial objects, humans vs. monkeys, monkeys vs. other mammals, and mammals vs. non-mammals (Fig. 1d). They also generalized the learned rule to the new set well ($\sim 80\%$ correct).

Encouraged by these findings, we next tested their behavior using a large scale natural image set with color and background. We selected images from the THINGS database (Hebart et al., 2019; Stoinski et al., 2023) and created three tasks: animate vs. inanimate, natural vs. artificial objects, and mammalian vs. non-mammalian animals. For each task, we first trained the three monkeys with 100 training images and then tested their performance on new images without repetition (Fig. 1e). The new images were either from the same object categories as the training set defined in the THINGS database ("same category"; e.g., both images were dogs) or images from categories that were not presented during the training ("new category"; $>$ 1,000 images each). For both types of novel images, the monkeys successfully generalized their learned rule, achieving 80%$-$90% correct (Fig. 1f).

Comparison with human behavior

We observed that the monkeys tended to respond incorrectly to images that also looked atypical to humans (e.g., a snake in a coil; Fig. 1g). To confirm the similarity between monkey and human judgment, we asked human participants to perform the same tasks with cropped image sets ($\sim 9,500$ trials

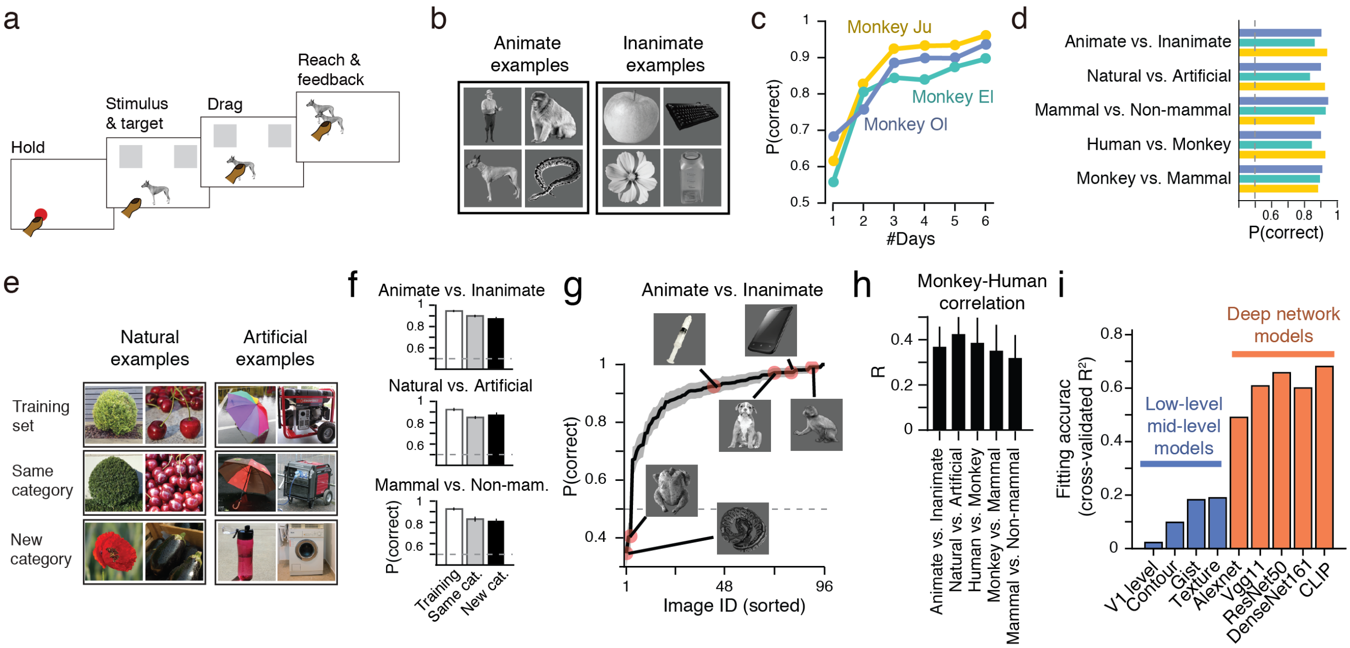

Figure 1: (a) Macaque monkeys moved an object image to one of two target boxes on a touchscreen according to a hidden rule learned through juice reward feedback. (b) For example, two boxes were associated with abstract concepts such as animate vs. inanimate. We first used grayscale images with no background to train and test monkey's performance. (c) By day 3-4, monkeys' performances reached $\sim$ 90% correct. The chance level was 50%. (d) They showed high performance for various concept classifications tested one after another. Bar colors indicate different monkeys. (e) We then tested monkey performance with a large-scale image set with color and background (THINGS; Hebart et al. (2019)). Test images were shown only once and contained object categories either presented ("same category") or not presented ("new category") during training. (f) The monkeys showed high performance in various tasks. (g) Monkeys made frequent errors on certain images. (h) This behavior was correlated with human performance. (i) Monkey behavior was better fit by deep learning models than by low- or mid-level image models. Error bars indicate S.E.M. Images were either obtained from Kiani et al. (2007) or used under a CC0 license.

per task with six participants). Humans immediately understood the rules for each task and performed almost perfectly, but their reaction times were systematically longer for some images. We therefore developed a metric to estimate the difficulty of each stimulus from choices and reaction times using the drift-diffusion model (Gold & Shadlen, 2007). This metric had moderate correlations between humans and monkeys across the images (Fig. 1h; $R = 0.3 - 0.43$; $P < 0.006$), indicating that images difficult for humans were also difficult for monkeys. Given the variability in behavioral responses, the noise ceiling of this correlation is estimated to be $0.5 - 0.6$.

How could monkeys solve these tasks?

The monkeys engaged in our experiments probably never saw many objects in the image sets, but then how could they categorize these images at this abstract level? We believe that high-level visual processing alone is largely sufficient for solving these classification problems (c.f., Long et al. (2018)). To demonstrate this, we compared monkey behavior with that of pre-trained ANNs. Monkey behavior was best fit by deep layers of various ANNs, whereas low- or mid-level image features consistently showed poorer fits (Fig. 1i). Indeed, linear classifiers trained on ANN outputs could solve our tasks with high accuracy ($> 90\%$). Many of these ANNs were only trained to categorize the ImageNet 1,000 categories, thus direct training with abstract concepts seems unnecessary to form representations that reflect these concepts (c.f., Jozwik et al. (2017)).

Discussion

The observed speed of learning may be in part owing to our task design, which allowed the monkeys to touch an image to report their decision. The task could be leveraged to study monkey behavior using various categorization rules and stimulus sets. Successful learning of our tasks does not mean that monkeys internally possess the kinds of abstract concepts we operationally defined, such as natural or mammalian. Their behavior could be well fit by ANNs and is thus explainable as a categorization of high-level image features. Nevertheless, the surprisingly fast learning speed suggests that the acquisition of abstract concepts is subserved by the capacity of the primate brain to extract high-level image features and quickly associate them with behavioral responses without the aid of language.

Acknowledgments

This work was supported by the National Science and Technology Innovation 2030 Major Program (grant 2021ZD0203703) and the National Natural Science Fund for Excellent Young Scientists Fund Program (overseas).

References

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434.

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, *9*(2), 303–308.

Fize, D., Cauchoix, M., & Fabre-Thorpe, M. (2011). Humans and monkeys share visual representations. *Proceedings of the National Academy of Sciences*, *108*(18), 7635–7640.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535–574.

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS one*, *14*(10), e0223792.

Jozwik, K. M., Kriegeskorte, N., Storrs, K. R., & Mur, M. (2017). Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Frontiers in psychology*, *8*, 285087.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of neurophysiology*, *97*(6), 4296–4309.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., . . . Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141.

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*, *115*(38), E9015–E9024.

Minamimoto, T., Saunders, R. C., & Richmond, B. J. (2010). Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron*, *66*(4), 501–507.

Murphy, G. (2004). *The big book of concepts*. MIT press.

Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience*, *38*(33), 7255–7269.

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

Santos, L. R., Hauser, M. D., & Spelke, E. S. (2001). Recognition and categorization of biologically significant objects by rhesus monkeys (macaca mulatta): the domain of food. *Cognition*, *82*(2), 127–155.

Setogawa, T., Eldridge, M. A., Fomani, G. P., Saunders, R. C., & Richmond, B. J. (2021). Contributions of the monkey inferior temporal areas te and teo to visual categorization. *Cerebral Cortex*, *31*(11), 4891–4900.

Stoinski, L. M., Perkuhn, J., & Hebart, M. N. (2023). Thingsplus: New norms and metadata for the things database of 1854 object concepts and 26,107 natural object images. *Behavior Research Methods*, 1–21.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. part 1: behavioural study. *European Journal of Neuroscience*, *11*(4), 1223–1238.