Behavioral/Cognitive

Comparison of Object Recognition Behavior in Human and Monkey

Rishi Rajalingham,[1] Kailyn Schmidt,[2] and James J. DiCarlo[1,2]

[1]Department of Brain and Cognitive Sciences and [2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Although the rhesus monkey is used widely as an animal model of human visual processing, it is not known whether invariant visual object recognition behavior is quantitatively comparable across monkeys and humans. To address this question, we systematically compared the core object recognition behavior of two monkeys with that of human subjects. To test true object recognition behavior (rather than image matching), we generated several thousand naturalistic synthetic images of 24 basic-level objects with high variation in viewing parameters and image background. Monkeys were trained to perform binary object recognition tasks on a match-to-sample paradigm. Data from 605 human subjects performing the same tasks on Mechanical Turk were aggregated to characterize "pooled human" object recognition behavior, as well as 33 separate Mechanical Turk subjects to characterize individual human subject behavior. Our results show that monkeys learn each new object in a few days, after which they not only match mean human performance but show a pattern of object confusion that is highly correlated with pooled human confusion patterns and is statistically indistinguishable from individual human subjects. Importantly, this shared human and monkey pattern of 3D object confusion is not shared with low-level visual representations (pixels, V1+; models of the retina and primary visual cortex) but is shared with a state-of-the-art computer vision feature representation. Together, these results are consistent with the hypothesis that rhesus monkeys and humans share a common neural shape representation that directly supports object perception.

Key words: human; monkey; object recognition; vision

Significance Statement

To date, several mammalian species have shown promise as animal models for studying the neural mechanisms underlying high-level visual processing in humans. In light of this diversity, making tight comparisons between nonhuman and human primates is particularly critical in determining the best use of nonhuman primates to further the goal of the field of translating knowledge gained from animal models to humans. To the best of our knowledge, this study is the first systematic attempt at comparing a high-level visual behavior of humans and macaque monkeys.

Introduction

Humans are able to rapidly, accurately and effortlessly perform the computationally difficult visual task of invariant object recognition: the ability to discriminate between different objects in the face of high variation in object viewing parameters and background conditions (DiCarlo et al., 2012). However, it is still unclear how the human brain supports this behavior. To uncover the neuronal mechanisms underlying human visual processing, it has been necessary to study various animal models, including nonhuman primates, felines, and rodents (Hubel and Wiesel, 1962; Van Essen, 1979; Zoccolan et al., 2009). In particular, the rhesus macaque monkey, an Old World primate that diverged from humans ~25 million years ago (Kumar and Hedges, 1998), is one of the most widely used animal models of high-level human visual perception (Mishkin et al., 1983; Tanaka, 1996; Minamimoto et al., 2010; Kravitz et al., 2011; Grill-Spector and Weiner, 2014). There exist strong anatomical and functional correspondences of visual cortical areas between humans and monkeys (Tootell et al., 2003; Orban et al., 2004; Mantini et al., 2012; Miranda-Dominguez et al., 2014). Thus, it has long been assumed that high-level visual behaviors and underlying neural substrates are comparable between monkey and human. However, humans have capacities not found in monkeys, and their

Received Feb. 2, 2015; revised July 21, 2015; accepted July 23, 2015.

Author contributions: R.R. and J.J.D. designed research; R.R. and K.S. performed research; R.R. contributed unpublished reagents/analytic tools; R.R. analyzed data; R.R. and J.J.D. wrote the paper.

This work was supported by the National Institutes of Health and the Natural Science and Engineering Research Council of Canada.

The authors declare no competing financial interests.

Correspondence should be addressed to James J. DiCarlo, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: **dicarlo@mit.edu**.

DOI:10.1523/JNEUROSCI.0573-15.2015

brains differ in important ways (Passingham, 2009). To date, the limits of the similarity in high-level visual behaviors of macaques and humans are unknown because no effort has been made to systematically compare rhesus macaque monkeys with humans in invariant object recognition. In light of recent work showing that rodent models of visual processing display the qualitative ability to perform invariant shape discrimination (Zoccolan et al., 2009), making tight, quantitative comparisons between monkeys and humans is especially critical in determining the best use of nonhuman primates to further the goal of the field of translating knowledge gained from animal models to humans.

To do this, we here systematically compared the behavior of two macaque monkeys with that of normal human subjects on an invariant object recognition paradigm. Our goal was to make direct measurements of object recognition ability, over a very large number of recognition tasks, always under conditions of high object view variation (also known as "invariant" object recognition). We focused on "core invariant object recognition": rapid and reliable recognition during a single, natural viewing fixation (DiCarlo and Cox, 2007; DiCarlo et al., 2012), operationalized as images presented in the central 10° of the visual field for durations under 200 ms. We further restricted our behavioral domain to "basic-level" object categories, as defined previously (Rosch et al., 1976). We do not claim this to be an exhaustive characterization of all possible visual object recognition behaviors but rather a good starting point for that greater goal. Monkeys easily learn such tasks, and, after testing 276 such object recognition tasks, our results show that rhesus monkey and human behavior are essentially indistinguishable and that both species are easily distinguishable from low-level visual representations asked to perform the same 276 tasks. These results show that rhesus monkeys are a very good—and perhaps quantitatively exact—model of human invariant object recognition abilities, and they are consistent with the hypothesis that monkeys and humans share a common neural representation that directly underlies those abilities.

Materials and Methods

Visual images. We examined basic-level object recognition behavior by generating images of a set of 64 objects that we found previously to be highly reliably labeled by independent human subjects, based on the definition proposed previously (Rosch et al., 1976). From this set, three groups of eight objects were sampled for this study; the selection of these 24 objects was random but biased toward groups of objects that exhibited reliable confusion patterns in humans (for a full list of those 24 basic-level objects, see Fig. 1). To enforce true object recognition behavior (rather than image matching), several thousand naturalistic images, each with one foreground object, were generated by rendering a 3D model of each object with randomly chosen viewing parameters (2D position, 3D rotation, and viewing distance) and placing that foreground object view onto a randomly chosen, natural image background. To do this, each object was first assigned a canonical position (center of gaze), scale (~2°), and pose, and then its viewing parameters were randomly sampled uniformly from the following ranges for object translation ([−3°, 3°] in both h and v), rotation ([−180°, 180°] in all three axes), and scale ([0.7×, 1.7×]). Backgrounds images were sampled randomly from 3D high-dynamic range images of indoor and outdoor scenes obtained from Dosch Design (www.doschdesign.com). As a result, these images require any visual recognition system (human, animal, or model) to tackle the "invariance problem," the computational crux of object recognition, because it is highly challenging for low-level visual representations (Ullman and Humphreys, 1996; Pinto et al., 2008). Using this procedure, we generated 2400 "test" images (100 images per object) at 1024 × 1024 pixel resolution with 256-level grayscale and with square apertures for human psychophysics, monkey psychophysics, and model evaluation. A separate set of 4800 "training" images (200 images per object) were generated with the same procedure with circular apertures to initially train the monkeys. Figure 1*A* shows example test images for each of the 24 basic-level objects.

To quantify the overlap between training and test image sets, we computed the pixel Euclidean distance of each test image to the nearest training image of the same object. For this analysis, training and test images were re-rendered on gray backgrounds to measure background-independent distances and resized to 64 × 64 pixel resolution. The resulting distance distribution was compared with that computed from simulated "ideal" generalization conditions. We rendered six different surrogate training image sets, each with identical generative parameters to the background-less training images except for one of the six viewpoint parameters held at its mean value. These sparsely sampled training images simulated six different experimental conditions wherein subjects would not have been exposed to variations in one parameter during the training stage but later tested on full variation images. From these newly generated training images, we computed the corresponding "ideal" pixel Euclidean distances of each test image to the nearest training image of the same object. We found that the distribution of background-independent distances of actual test images from actual training image sets was only slightly less than the corresponding distribution across the simulated ideal generalization conditions (3.4 and 3.7% increase in median and maximum distances, respectively, for the simulated conditions; Fig. 2*D*, bottom). This suggests that our 4800 training images did not sample the image space too "densely" but rather of comparable sparsity as if we had entirely held back particular types of viewpoint variations.

Human behavior. All human behavioral data presented here were collected from 638 human subjects on Amazon Mechanical Turk (MTurk) performing 276 interleaved, basic-level, invariant, core object recognition tasks. Each task consisted of a binary discrimination between pairs of objects from the 24 objects considered. Subjects were instructed to report the identity of the foreground object in each presented image, from two given choices. Because those two choices were provided after the test image and all 276 tasks were interleaved randomly (trial-by-trial), subjects could not deploy feature attentional strategies specific to each task to process the test images. Each trial initiated with a central black point for 500 ms, followed by 100 ms presentation of a test image. The test image contained one foreground object presented under high variation in viewing parameters and overlaid on a random background, as described above (see Visual images). Immediately after extinction of the test image, two choice images, each displaying a single object in a canonical view with no background, were shown to the left and right. One of these two objects was always the same as the object that generated the test image (i.e., the correct choice), and its location (left or right) was chosen randomly on each trial. After mouse clicking on one of the choice images, the subject was given another fixation point before the next stimulus appeared. No feedback was given; subjects were never explicitly trained on the tasks. Under assumptions of typical computer ergonomics, we estimate that images were presented at 6–8° of visual angle in size, and response images were presented at 6–8° of eccentricity.

The online MTurk platform enables efficient collection of reliable, large-scale psychophysical data and has been validated by comparing results obtained from online and in-lab psychophysical experiments (Majaj et al., 2012; Crump et al., 2013). In particular, a previous study from our group directly compared the patterns of behavioral performance on invariant object recognition tasks of in-lab and online subjects. In brief, human subjects were tested in a controlled in-lab setting on eight-alternative forced-choice core object recognition tasks at both basic and subordinate levels. A total of 10, 15, and 22 subjects each performed 600 trials of basic-level object categorization, car identification, and face identification tasks, respectively. Pooling trials from all subjects, the in-lab human behavioral data was highly reliable ($\rho_{\text{in-lab,in-lab}} = 0.95 \pm 0.024$, median ± SE). A separate set of 104 human subjects from MTurk performed trials of the same tasks, resulting in similarly reliable pooled online human data ($\rho_{\text{MTurk,MTurk}} = 0.97 \pm 0.023$, median ± SE). Accounting for noise, the behavioral patterns of in-lab and online human subjects were virtually identical ($\tilde{\rho}_{\text{in-lab,MTurk}} = 0.98$; see below, Analysis; Majaj et al., 2012), supporting the use of MTurk for characterizing hu-

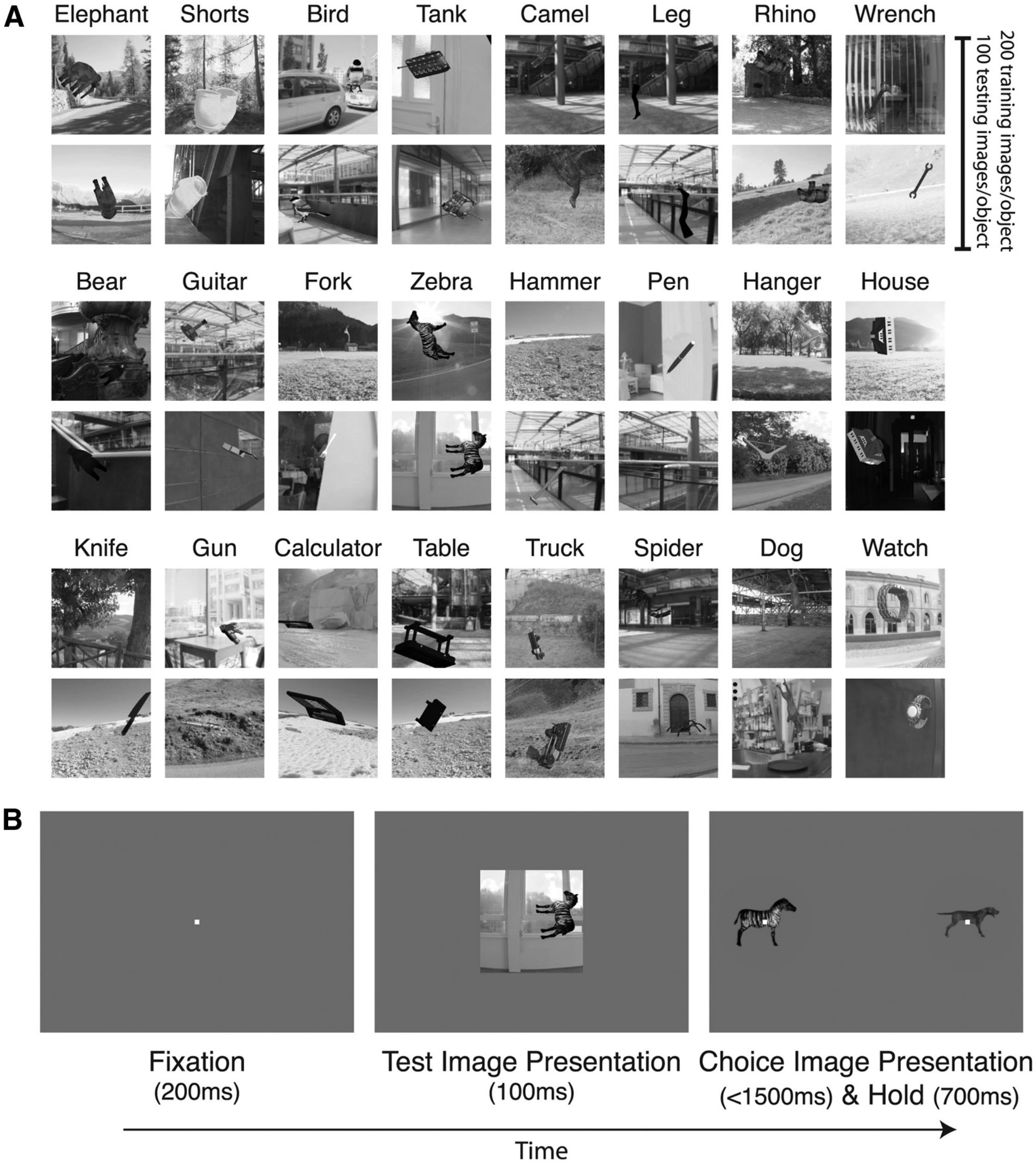

Figure 1. ***A***, Two example images for each of the 24 basic-level objects, sampled from the test set (each row corresponds to a group of 8 objects). To enforce true object recognition behavior (rather than image matching) and tackle the invariance problem, we generated thousands of naturalistic images, each with one foreground object, by rendering a 3D model of each object with randomly chosen viewing parameters (2D position, 3D rotation, and viewing distance) and placing that foreground object view onto a randomly chosen, natural image background. ***B***, Behavioral paradigm (for monkey M). Each trial was initiated when the monkey held gaze fixation on a central fixation point for 200 ms, after which a square test image (spanning 6° of visual angle) appeared at the center of gaze for 100 ms. Immediately after extinction of the test image, two choice images, each displaying the canonical view of a single object with no background, were shown to the left and right (see Materials and Methods). Test and choice images are shown to scale. The monkey was allowed to freely view the response images for up to 1500 ms and responded by holding fixation over the selected image for 700 ms. Monkey Z performed the exact same tasks but used touch to initiate trials and indicate its choice (see Materials and Methods). Successful trials were rewarded with juice, and incorrect choices resulted in timeouts of 1.5–2.5 s.

man core object recognition behaviors. Following the methods of that previous work, we here did not perform eye tracking of online human subjects to measure or control their gaze. Instead, subjects were cued to the location of image presentation with a fixation cue. Subjects detected as obviously guessing were banned from additional experiments, and the corresponding data were excluded from additional analyses (<1% of subjects were excluded). To do this, we quantified this guessing behavior using a choice entropy metric that measured how well the current trial response of a subject was predicted by the previous trial response. For all remaining subjects, we did not observe any significant differences in performance between the first and last halves of behavioral trials ($p = 0.49$, *t* test), suggesting that subjects did not undergo substantial learning. Overall, subjects achieved high performance on all behavioral tasks (88.35 ± 5.6%, mean ± SD; $n = 276$ tasks).

Most of the human psychophysical subjects were used to characterize "pooled human" behavior. Specifically, data from 605 MTurk subjects each performing a relatively small number of trials (mean of 114 trials/subject) were aggregated to obtain a highly reliable estimate of pooled human object recognition performance on each task. Each subject only performed a subset of the tasks (mean of 67 tasks/subject). All trials of all of these subjects (69,000 trials in total, 250 trials/task) were pooled together to characterize pooled human behavior.

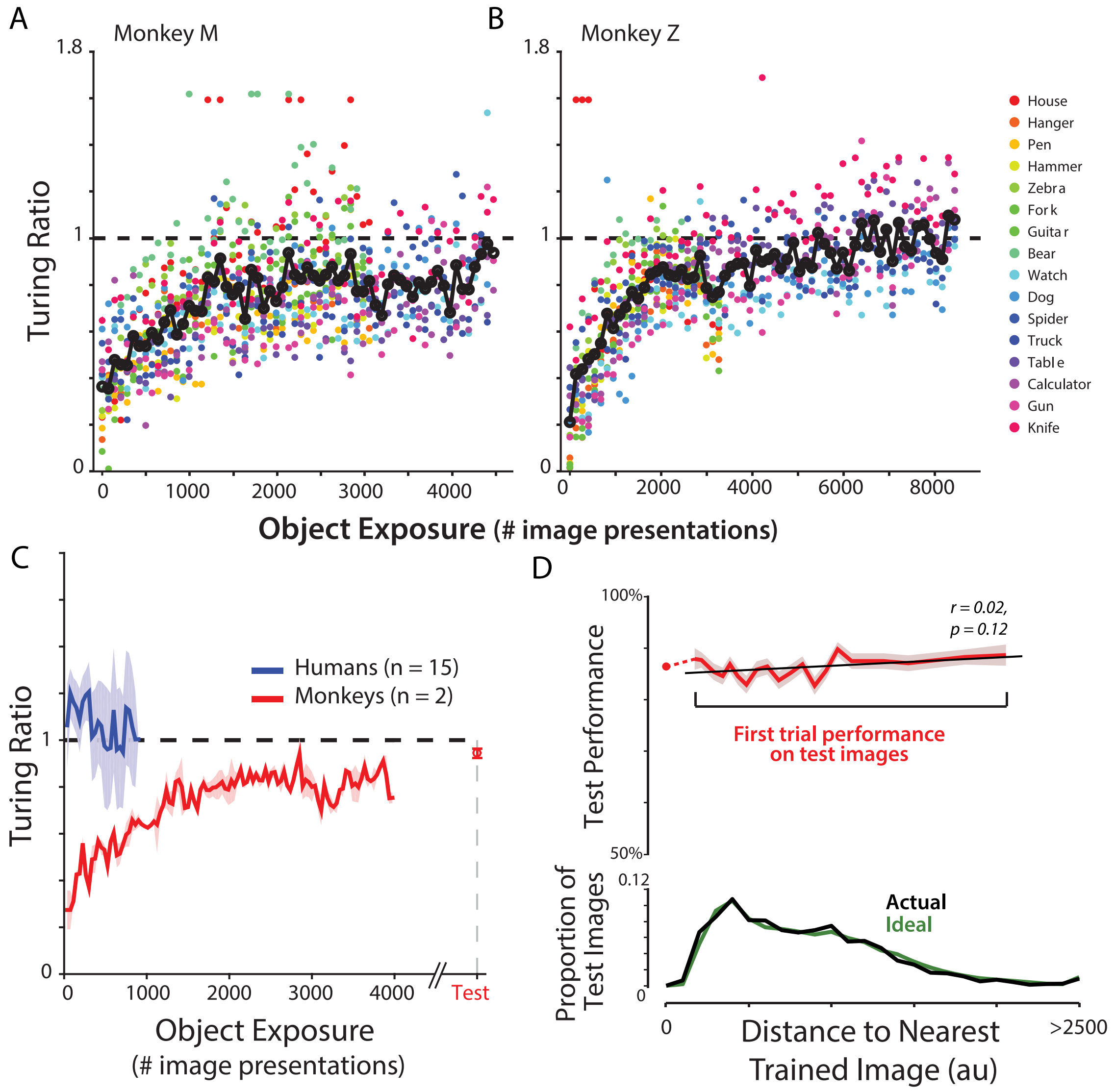

Figure 2. ***A***, ***B***, Learning of novel objects. For each monkey, performance relative to the human pool, quantified as a Turing ratio d'/d' human pool (one-versus-all d'), for each of these 16 objects (colored dots) and the mean over objects (black line) are shown. For these 16 objects, animals were exposed to full-variation images in groups of eight objects and reached high performance in 1000–2000 image presentations per object. ***C***, Performance relative to the human pool, quantified as a Turing ratio, of two monkeys and 15 unique individual humans subjects with sufficient longitudinal data on the same tasks (mean ± SE over subjects). Monkeys learned each new object rapidly, whereas humans performed at a high initial performance and exhibited no change in performance as a function of (unsupervised) experience with the objects. The Test marker indicates monkeys' relative performance on held-out test images, after all behavioral training. ***D***, Generalization to novel images. Top, Pooling data from both monkeys, the first-trial performance of 2400 test images (mean ± SE) is plotted as a function of the corresponding distance to the nearest training image; black line denotes linear regression. Distance is computed on gray background images, as a Euclidean pixel distance of each test image to the nearest training image of the same object (see Materials and Methods, Visual images). The distance values have been discretized into 20 bins of varying lengths with equal number of images. The overall performance, including all subsequent exposures to test images, is shown on the left (at 0 distance). Bottom, Overlap between training and test image sets. The distribution of distances of test images to the nearest trained image is shown relative to actual training images (black line) and to "ideal" generalization surrogate training images (green line). The last histogram bin includes all distances >2500.

A separate set of human subjects was used to characterize the variability in individual human subject behavior. Specifically, these MTurk subjects performed a relatively large number of trials of binary object recognition tasks over groups of only eight of the 24 objects (note that eight objects generate 28 unique binary tasks). Each individual human subject performed trials for all of those 28 tasks. To ensure that sufficiently many trials were collected from each individual subject for reliable measurement of his or her pattern of behavioral performance, we used a predetermined criterion for reliability, defined as split-half internal consistency (see below, Analysis) significantly greater than 0.5 ($p <$ 0.05, one-tailed t test). We then attempted to repeatedly recruit each subject until his or her total pool of data reached this reliability criterion. Of 80 unique subjects that performed this task, 33 were successfully re-recruited a sufficient number of times to reach the reliability criterion on at least one group of objects (five of these subjects performed tasks in two different groups, and three subjects performed in all three groups). In total, we collected data from 16, 16, and 12 subjects for each of three groups of eight objects (mean of 3003 trials/subject within each group).

Humans, although not explicitly trained on these images, likely get extensive experience with similar objects over the course of their lifetime. To investigate the effect of experimental experience on behavior, we measured the performance of individual human subjects as a function of the number of presentations per object. To allow direct comparison with monkeys, this analysis was constrained to the 16 objects in the second and third groups for which corresponding monkey "training" data were also available. Figure 2C shows the relative performance, quantified as a Turing ratio (d' individual human/d' human pool, one-versus-all d'), of individual humans subjects with sufficient longitudinal data on these tasks (defined as >150 trials/object, 15 unique human subjects). We observe that individual humans perform at a high initial performance and exhibit no change in performance as a function of (unsupervised) experience with the objects, suggesting that humans are already well trained on these tasks.

Monkey training and behavior. Monkey behavioral data on the exact same object recognition tasks were collected from two adult male rhesus macaque monkeys (*Macaca mulatta*) weighing 6 kg (monkey M) and 12 kg (monkey Z). All procedures were performed in compliance with National Institutes of Health guidelines and the standards of the Massachusetts Institute of Technology Committee on Animal Care and the American Physiological Society. To ensure that our behavioral tests were tapping a sensory representation (i.e., did not depend on the reporting effector), we tested one monkey (M) using saccade reports (gaze tracking) and the other monkey (Z) using reaching reports (touch screen).

For monkey M, before behavioral training, a surgery using sterile technique was performed under general anesthesia to implant a titanium head post to the skull. After head-post implant surgery, monkey M was trained on a match-to-sample paradigm under head fixation and using gaze as the reporting effector. Eye position was monitored by tracking the position of the pupil using a camera-based system (SR Research Eyelink II). Images were presented on a 24-inch LCD monitor (1920 × 1080 at 60 Hz; GD235HZ; Acer) positioned 42.5 cm in front of the animal. At the start of each training session, the subject performed an eye-tracking calibration task by saccading to a range of spatial targets and maintaining fixation for 800 ms. Calibration was repeated if drift was noticed over the course of the session. Figure 1*B* illustrates the behavioral paradigm. Each trial was initiated when the monkey acquired and held gaze fixation on a central fixation point for 200 ms, after which a test image appeared at the center of gaze for 100 ms. Trials were aborted if gaze was not held within ±2°. After extinction of the test image, two choice images, each displaying a single object in a canonical view with no background, were shown immediately to the left and right (each centered at 6° of eccentricity along the horizontal meridian; Fig. 1*B*). One of these two objects was always the same as the object that generated the test image (i.e., the correct choice), and its location (left or right) was chosen randomly on each trial. The monkey was allowed to view freely the choice images for up to 1500 ms and indicated its final choice by holding fixation over the selected image for 700 ms. During initial training, the monkey typically visually explored both objects before making a selection but quickly transitioned to selecting its choice covertly in that it often directed its first saccade in the direction of the final choice.

Monkey Z performed the same task using a touch screen. Other than the differences noted below, the task was identical to monkey M. Monkey Z underwent no surgical procedures and was instead seated head free in front of a 15-inch LCD touch screen (1024 × 768 at 60 Hz; ELO Touch 1537L) at a distance of 34.2 cm. The subject interacted with the task by touching the screen with its left hand through an opening in the primate chair. Monkey Z initiated each trial by touching a fixation point 4° below the center of the screen for 250 ms, which triggered the presentation of the test image at the center of the screen (i.e., this ensured that the hand and finger rising from below did not occlude any of the test image). After the appearance of the choice images, the monkey indicated its choice by touching the selected image. Gaze was not controlled or measured in monkey Z, but we instead assumed that touch-point acquisition would correspond to gaze being directed at the screen. Because the test image screen location was fixed over trials and the test image content was required for successful reward, we assumed that the animal's gaze would be directed reliably at the center of each test image. This assumption is supported by the finding that monkey Z showed a very similar pattern of performance as monkey M (Fig. 3*D*).

The images were sized so that they subtended 6 × 6° for each monkey. Real-time experiments for all monkey psychophysics were controlled by open-source software (MWorks Project http://mworks-project.org/). Animals were rewarded with small juice rewards for successfully completing each trial and received timeouts of 1.5–2.5 s for incorrect choices. Animals were trained to work with each group of eight objects to criterion, defined as a stable pattern of behavioral performance over at least four behavioral sessions, before moving on to the next group. For the first group, both monkeys were exposed to images with gradually increasing variation in viewing parameters (pose, position, and viewing distance) over several behavioral sessions. For each subsequent group of eight objects, animals were exposed immediately to full variation images and reached high performance in 1000–2000 image presentations per object (~10–15 behavioral sessions). Figure 2 shows the monkeys' performance relative to the human pool, quantified as a Turing ratio (d' monkey/d' human pool, one-versus-all d'), for each of these 16 objects. When presented with these novel objects, monkeys were able to reach high-level performance relatively quickly (Fig. 2*A*,*B*). After training of all binary tasks in each group of eight objects, animals were trained to criterion on the remaining pairs of objects. Once animals reached criterion on all 276 possible binary object recognition tasks, complete behavioral data were collected in a fully interleaved manner, first using training images and subsequently switching to held-out test images. Importantly, monkeys immediately generalized to new images of previously learned objects, with comparable high performance on the very first trial of a new image for both monkeys (monkey M, 88%; monkey Z, 85%). Furthermore, the monkeys' performance on the first trial of novel test images was not dependent on the similarity of the test image to previously seen training images (see above, Visual images). We observed no significant negative correlation between first-trial performance of test images and their background-independent distance to the nearest training images ($r = 0.036$, $p = 0.07$ and $r = 0.010$, $p = 0.63$ for monkeys M and Z, respectively), as shown in the top of Figure 2*D* (mean ± SE, pooling both monkeys). Subsequent exposures to these test images did not further increase behavioral performance (Fig. 2*D*, zero distance marker). Together, this suggests that monkeys did not rely simply on the similarity to previously seen images. Furthermore, the object confusion patterns were found to be primarily independent of the image set; the consistency (computed as a noise-adjusted correlation; see Analysis) between confusion patterns of the training and test image sets was 0.9566 ± 0.0253 and 0.9489 ± 0.0157 (mean ± SD, for monkeys M and Z, respectively). Thus, complete behavioral data collected in a fully interleaved manner from both images sets were pooled. A total of 106,844 trials were collected from both monkeys (51,096 from monkey M and 55,748 from monkey Z) and used for the analyses below.

Machine behavior. We tested different machine systems on our 276 tasks by computing the feature population output of each machine for each of our images and using trained classifiers to make behavioral choices based on the test images.

Low-level visual representations of pixel and V1+ (Pinto et al., 2008) features were used as control models. These features approximate the representations of the retina and primary visual cortex, respectively.

High-performing feature representations from state-of-the-art computer vision models were also tested for comparison. HMAX (Riesenhuber and Poggio, 1999; Serre et al., 2007) is a model inspired by the tolerance and selectivity properties of the ventral visual stream. We used the publicly available FHLib implementation (Mutch and Lowe, 2008). We trained the model on 5760 synthetic images of 3D objects drawn from eight natural categories (animals, boats, cars, chairs, faces, fruits, planes, and tables; Yamins et al., 2014) that did not overlap with the 24 objects used in this study. CNN2013 refers to a model based on a state-of-the-art deep convolutional neural network model (Zeiler and Fergus, 2014). Using an architecture and learning parameters based on a previous study (Zeiler and Fergus, 2014), we trained a deep convolutional neural network for 15 epochs on images drawn from the ImageNet 2013 challenge set, adjusting the learning rate in accordance with the heuristics described in this publication. We used a publicly available implementation (Wan et al., 2013), itself based on CudaConvnet (Krizhevsky et al., 2012), that allowed for dropout regularization. Training the model took 2 weeks on a Tesla K40 GPU, provided by NVIDIA.

For each machine representation, features were extracted from the same images that were presented to humans and monkeys. As with humans and monkeys, each machine representation was tested on the same 276 interleaved binary object recognition tasks. For each machine feature representation, performance on each binary task was measured using

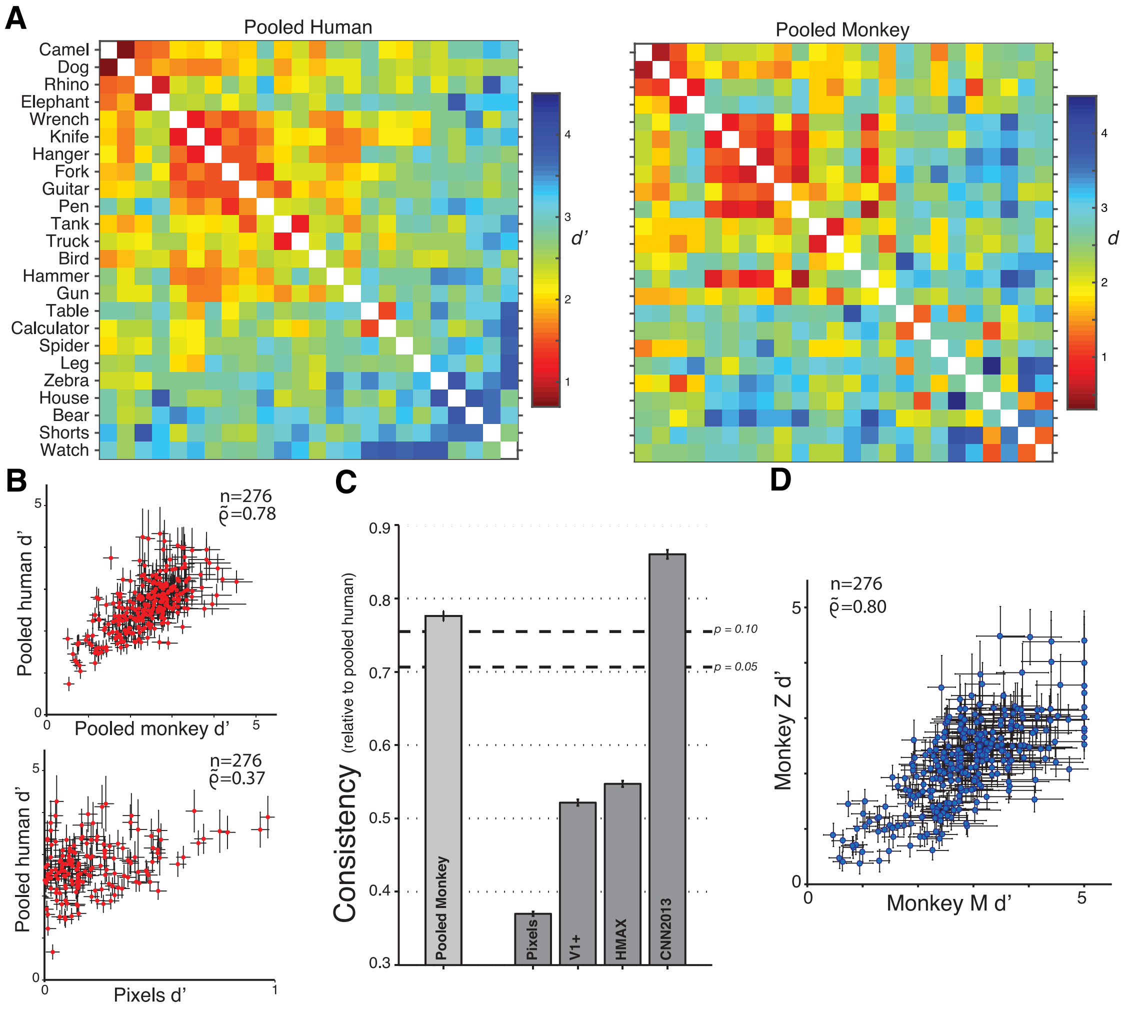

Figure 3. ***A***, Pattern of behavioral performances for the pooled human and pooled monkey. Each 24 × 24 matrix summarizes confusions of all two-way tasks: the color of bin (*i*,*j*) indicates the unbiased performance (d') of the binary recognition task with objects *i* and *j*. Objects have been reordered based on a hierarchical clustering of human confusion patterns to highlight structure in the matrix. We observe qualitative similarities in the confusion patterns. For example, (camel, dog) and (tank, truck) are two often confused object pairs in both monkeys and humans. ***B***, Comparison of d' estimates of all 276 tasks (mean ± SE as estimated by bootstrap, 100 resamples) of the pooled human with that of the pooled monkey (top) and a low-level pixel representation (bottom). ***C***, Quantification of consistency as noise-adjusted correlation of d' vectors. The pooled monkey shows patterns of confusions that are highly correlated with pooled human subject confusion patterns (consistency of pooled monkey, 0.78). Importantly, low-level visual representations do not share these confusion patterns (pixels, 0.37; V1+, 0.52). Furthermore, a state-of-the-art deep convolutional neural network representation was highly predictive of human confusion patterns (CNN2013, 0.86), in contrast to an alternative model of the ventral stream (HMAX, 0.55). The dashed lines indicate thresholds at $p = 0.1, 0.05$ confidence for consistency to the gold-standard pooled human, estimated from pairs of individual human subjects. ***D***, Comparison of d' estimates of all 276 tasks (mean ± SE as estimated by bootstrap, 100 resamples) between the two monkeys.

twofold cross-validation using a maximum correlation classifier, repeated 50 times over permutations of classifier training and testing data partitions.

Analysis. For each of the 276 binary object recognition tasks, an unbiased measure of performance was estimated using a sensitivity index d' (Macmillan, 1993): $d' = Z(\text{hit rate}) - Z(\text{false-alarm rate})$, where $Z(\ldots)$ is the inverse of the cumulative Gaussian distribution. All d' estimates were constrained to a range of [0, 5]. Bias was estimated using a criterion index c (Macmillan, 1993): $c = 0.5 \times (Z(\text{hit rate}) + Z(\text{false-alarm rate}))$. We refer to the 276-dimensional vector of d' values over all binary object recognition tasks as the "pattern of behavioral performance" (**b**).

The reliability (also known as internal consistency) of behavioral data was computed as the Spearman's correlation between patterns of behavioral performance patterns computed on separate halves of the data (random split-halves of trials); this process was repeated across 100 draws. Because this estimate is measured using only half of the data, the Spearman–Brown prediction formula (Brown, 1910; Spearman, 1910) was applied to allow for comparisons with correlations measured using all trials.

Consistency between different behavioral patterns $\mathbf{b_1}$, $\mathbf{b_2}$ was then computed as a noise-adjusted rank correlation between patterns of behavioral performances (d' or c vectors):

$$\tilde{\rho}_{b_1,b_2} = \frac{\rho_{b_1,b_2}}{\sqrt{\rho_{b_1,b_1} \times \rho_{b_2,b_2}}},$$

where ρ_{b_1,b_2} is the raw Spearman's rank correlation, and ρ_{b_1,b_1}, ρ_{b_2,b_2} are the Spearman–Brown corrected internal consistency estimates of each behavioral pattern. Our rationale for using a noise-adjusted correlation measure for consistency was to account for variance in the behavioral patterns that arises from "noise," i.e., variability that is not replicable by stimulus object identity. We obtained a distribution of consistency values using the 100 resampled estimates of internal consistencies of each behavioral pattern (i.e., from the 100 random draws of split-halves of trials of $\mathbf{b_1}$, $\mathbf{b_2}$).

Results

As stated in Introduction, our primary goal was to measure the difficulty of hundreds of basic-level invariant object discrimina-

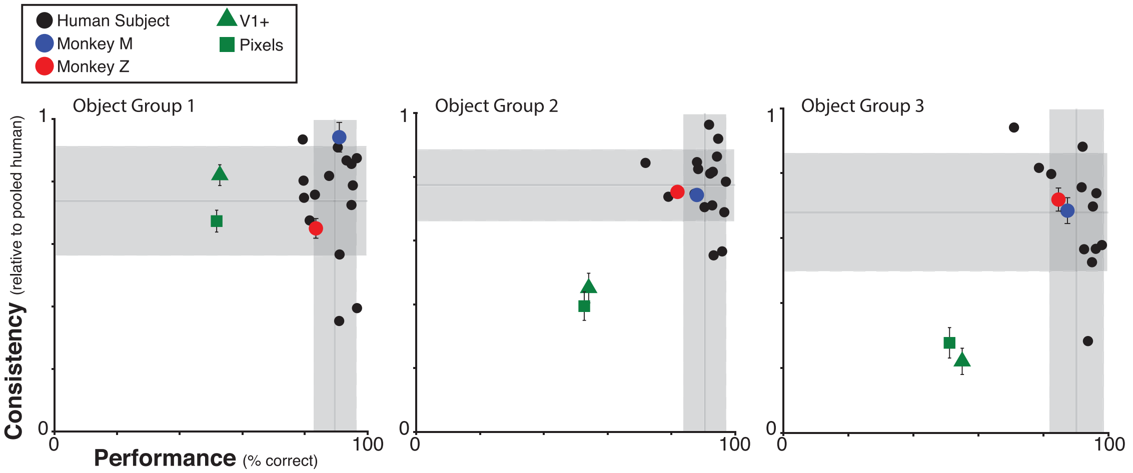

Figure 4. Accounting for intersubject variability. For each of three groups of eight objects, the absolute performance and consistency of individual human subjects, individual monkeys, and machine features are shown. Error bars for consistency relative to pooled human (mean ± SD) are shown for individual monkeys and machine features for each group (error bars for monkeys are not visible in object group 2 because of small variability). The shaded gray areas indicate the distribution of performance/consistency over individual human subjects (mean ± SD). There is significant intersubject variability: individual human subjects are on average not perfectly correlated with the pooled human (average consistency, 0.74, 0.77, and 0.68 for the 3 groups). As a result, monkeys are statistically indistinguishable from individual human subjects in their correlation to the human pool (monkey M, $p = 0.4$, 0.21, and 0.22; monkey Z, $p = 0.16$, 0.22, and 0.27). In contrast, low-level visual representations were falsified on both performance and consistency grounds for two of three groups of objects (V1, $p = 0.30$, 0.03, and 0.03; pixels, $p = 0.17$, 0.02, and 0.03).

tion tasks in human subjects and to compare that pattern of behavioral performance with that measured in monkeys performing the exact same tasks. In summary, the comparisons presented here (Figs. 3, 4) are based on data obtained from a pool of 605 human subjects (69,000 total trials) and two monkeys (106,844 total trials; see Materials and Methods). The monkey data pool only includes behavioral trials collected after the monkey learned all 24 objects. This corresponds to a total of 250 human behavioral trials for each of the 276 tasks and a total of 362–417 monkey trials for each task.

As described in Materials and Methods, monkeys rapidly learned each new object. Although we do not know the humans' or monkeys' previous lifetime experiences with these objects or related image statistics, we reasoned that, after monkey training, both species might be in a comparable lifetime training regimen. To assess this, we examined the effects of experience in both humans and monkeys by tracking the performance of individual human and monkey subjects as a function of the number of exposures to each object. Figure 2, *A* and *B*, shows each monkey's performance, relative to the human pool, when presented with 16 novel objects; both monkeys were able to reach high performance relatively quickly (~1000–2000 image presentations). Figure 2*C* directly compares both monkeys with individual humans subjects on the exact same tasks ($n = 15$ human subjects with sufficient longitudinal data). Unlike monkeys, individual human subjects initially behaved at a high level of performance and exhibited no increase in performance as a function of exposures to objects, suggesting that humans have previous experience with similar objects.

After training, monkeys maintained high performance when tested on novel images of these previously learned objects (Fig. 2*C*, Test marker). Importantly, this generalization was immediate, with comparable high performance on the very first trial of a new image for both monkeys. Furthermore, the generalization performance did not depend on the similarity of test images to previously seen training images (see Materials and Methods; Fig. 2*D*, top). Indeed, monkeys maintained high behavioral performance even for the subset of test images that were nearly as far from the training images as they would have been if we had completely restricted training with each single axis of variation (see Materials and Methods). Finally, subsequent exposures to these test images did not further increase behavioral performance (Fig. 2*D*, zero distance marker). Together, these results suggest that monkeys were not simply memorizing previously learned images and that they could generalize to significantly different images of the same object, even when the training images only sparsely sample the view space of the object. Importantly, although it is impossible to guarantee or expect that humans and monkeys have identical lifetime experience, we find that, once monkeys were trained, additional experience had no observable effect on the behavioral performance of either species.

The difficulty of each of the 276 invariant object discrimination tasks was determined by measuring the unbiased performance (d') of monkeys/humans. That performance metric is shown for all 276 tasks in Figure 3*A*. We refer to these data as the pattern of behavioral performance for pooled human (605 subjects, 69,000 trials) and pooled monkey (two subjects, 106,844 trials). Note that these matrices are not standard confusion matrices, but they are closely related in that they express the pairwise difficulty (confusion) of each pair of objects. Objects in Figure 3*A* have been reordered based on a hierarchical clustering of human error patterns to highlight structure in the matrix. Because difficulty is measured via unbiased performance in discriminating pairs of objects, matrices are symmetric by construction. We noted high qualitative similarities in these patterns of performance (Fig. 3*A*, compare pooled human and pooled monkey). For example, (camel, dog) and (tank, truck) are two examples of object pairs that are often confused by humans and monkeys alike.

To quantify the similarity of these patterns of behavioral performance, we took the pooled human pattern as the gold standard. We first compared the pooled monkey pattern by computing the noise-adjusted correlation of the 276 d' values (termed consistency relative to the pooled human; see Materials and Methods). We found this number to be 0.78 ± 0.007 (Fig. 3*B*, top). Using the same human gold standard and the same methods, we also computed the consistency of each monkey, a low-level pixel representation (Fig. 3*B*, bottom), and computer vision representations (Fig. 3*C*). Both monkeys show patterns of confusions that are highly correlated with pooled human confusion patterns (monkey M, 0.80 ± 0.009; monkey Z, 0.66 ± 0.005). Importantly, low-level visual representations do not share these patterns of behavioral performance (pixels, 0.37 ± 0.003; V1+, 0.52 ± 0.004). Of the two high-performing computer vision image representations, we found that that from the top layer of a state-of-the-art deep convolutional neural network model optimized for invariant object recognition performance was highly predictive of human confusion patterns (CNN2013, 0.86 ± 0.006), in contrast to a previous, alternative model of the ventral stream (HMAX, 0.55 ± 0.004).

Although both monkeys' patterns of behavioral performance were highly consistent with the pooled human pattern, they were not perfectly correlated (i.e., the consistency value is not 1.0). We asked whether this reflected a true species difference between human and monkey behavior or whether it might be explained by within-species subject variability. To do so, 16, 16, and 12 separate MTurk subjects were recruited to characterize individual human subject behavior on three groups of eight objects. First, we observed that there is significant intersubject variability in the tested human population; the median consistency of behavioral patterns between pairs of individual humans subjects is only 0.76. Consequently, individual human subjects are on average not perfectly correlated with the pooled human. Figure 4 shows both absolute performance (percentage correct) and consistency (relative to the pooled human) of individual human subjects, individual monkeys, and low-level machine features on the three tested groups of eight objects. Note that these data account for only 30% ($3 \times 28/276$) of all tasks presented in Figure 3. The solid and dashed lines indicate the mean ± SD performance/consistency of individual human subject population; for the three groups of eight objects, the mean ± SD consistency of individual human subjects were 0.74 ± 0.18, 0.77 ± 0.11, and 0.68 ± 0.18. Importantly, this variability is sufficiently small to reject some representations. Indeed, low-level visual representations that model the retina and primary visual cortex fall outside the distribution of consistency over human subjects for two of three groups of objects (Fig. 4; V1+, $p = 0.30$, 0.03, and 0.03; pixels, $p = 0.17$, 0.02, and 0.03; Fisher's exact test for the three groups of objects, respectively). However, both monkeys remain statistically indistinguishable from individual human subjects in their consistency to the pooled human (monkey M, $p = 0.4$, 0.21, and 0.22; monkey Z, $p = 0.16$, 0.22, and 0.27 for the three groups; Fisher's exact test). Additionally, the CNN2013 model could not be falsified in any of the three groups of objects ($p = 0.38$, 0.35, and 0.36 for the three groups of objects, respectively), whereas HMAX was rejected in one of the three groups ($p = 0.26$, 0.07, and <0.01, respectively).

We next asked whether intersubject variability might also explain the imperfect consistency of the pooled monkey relative to the pooled human (Fig. 3*C*). To account for the small sample size of monkeys ($n = 2$), we randomly sampled pairs of individual human subjects and measured the consistency relative to the pooled human of their pooled behavioral data. This process was repeated 50 times for each of the three groups of eight objects to obtain a distribution of consistency of $n = 2$ pooled human subjects. Figure 3*C* shows the $p = 0.1$ and $p = 0.05$ confidence thresholds of this distribution (dashed lines). The pooled monkey cannot be rejected relative to this distribution, i.e., the pooled monkey's patterns of performance are statistically indistinguishable from patterns of similarly sampled pools of human subjects.

We also estimated biases in object confusion patterns using the criterion index *c* (see Materials and Methods). We found that this measure was significantly less replicable across subjects: the median consistency of bias (*c*) between pairs of individual humans subjects was 0.40 compared with 0.76 for consistency of unbiased performance (d'), suggesting that biases are significantly less meaningful behavioral signatures on which to compare humans and monkeys.

Discussion

Previous work has revealed quantitative similarity of human and macaque monkey behavior in low-level visual behaviors (De Valois et al., 1974a,b; Vogels and Orban, 1990; Vázquez et al., 2000; Kiorpes et al., 2008; Gagin et al., 2014), suggesting that an understanding of the neural mechanisms underlying those tasks in macaques will directly translate to humans. However, such correspondences for high-level visual behaviors, such as view-invariant object recognition, have not been demonstrated. Although many studies have shown that monkeys can learn to perform tasks based on object shape (Mishkin et al., 1983; Minamimoto et al., 2010; Okamura et al., 2014), this is taken by some as evidence of the powerful learning abilities of monkeys in experimenter-chosen tasks rather than a tight correspondence with humans in their behavioral patterns in object discrimination abilities. In this study, we systematically compared the basic-level core object recognition behavior of two rhesus macaque monkeys with that of human subjects. Our results show that monkeys not only match human performance but show a pattern of object confusion that is highly correlated with pooled human confusion patterns and that individual monkey subjects are statistically indistinguishable from the population of individual human subjects. Importantly, these shared patterns of basic-level object confusions are not shared with low-level visual representations (pixels, V1+).

We characterized average human population and individual human subject behavior using high-throughput online psychophysics on MTurk. This method allowed us to efficiently gather datasets of otherwise unattainable sizes and has been validated previously by comparing results obtained from online and in-lab psychophysical experiments (Majaj et al., 2012; Crump et al., 2013). In particular, patterns of behavioral performance on object recognition tasks from in-lab and online subjects were equally reliable and virtually identical (Majaj et al., 2012). Although we did not impose experimental constraints on subjects' acuity and we can only infer likely gaze position, the observed high performance and highly reliable confusion patterns suggest that this sacrifice in exact experimental control is a good tradeoff for the very large number of tasks (276) that could be tested.

We characterized monkey object recognition behavior from two subjects using two different effector systems. This modest sample size is typical for systems neuroscience experiments, given the cost and difficulty of monkey psychophysics. As a result, it is unclear whether the differences observed between monkeys (consistency between monkeys, 0.80) reflect true intersubject variability or are attributable to differences in the effector system. Monkey Z's overall performance (83.4%) was lower than monkey M's (89.2%), and, for an equal number of confusions, confusion patterns from monkey Z were significantly less reliable than those from monkey M ($p \ll 0.001$, two-sample *t* test). These differences suggest additional variance ("noise") in monkey Z's behavioral data, potentially attributable to less gaze control than monkey M, that may partly account for the differences in behavioral patterns between monkeys. However, this additional behavioral variance did not significantly influence the result; each monkey subject was highly correlated with the human pool and statistically indistinguishable from individual humans.

Additionally, it is possible that we failed to observe a significant difference between monkeys and humans because of a lack of statistical power from a sample of just two monkeys. In principle, one cannot prove that there is absolutely no difference between monkeys and humans, because ever-increasing power would be required for the testing of an ever-smaller proposed difference. Here, we showed that our behavioral tests do have sufficient power to falsify other models (pixels, V1+) as matching human core object recognition behavior but failed to falsify monkeys as a model of that domain of human behavior. Testing additional monkeys on this behavioral domain or additional behavioral tests beyond this domain may, in principle, reveal differences between monkey and human object recognition behavior.

We argue that the observed results are not attributable to overtraining of animals. Monkeys were trained using a standard operant conditioning paradigm to report object identity in visual images. Objects were novel to monkeys before behavioral training. When presented with these novel objects, monkeys were able to reach high-level performance relatively quickly (Fig. 2*A*–*C*). Furthermore, by sampling from a large pool of images, we were able to ensure that monkeys were exposed to any given image at most once per behavioral session on average. Finally, we collected behavioral data on a set of held-out images (test set, 100 images/object) after monkeys were trained fully to criterion on all tasks. Importantly, both monkeys successfully generalized to these new images of previously learned objects; performance on the very first trial of a new image was high for both monkeys (monkey M, 88%; monkey Z, 85%), and the first-trial performance was not predicted by the similarity of test images to previously seen training images (Fig. 2*D*). As a result, the measured patterns of behavioral performance reflect the monkeys' ability to discriminate between pairs of basic-level objects rather than memorized or overtrained behavior. Importantly, humans, although not explicitly trained on these images, likely get significant previous experience with similar objects over the course of their lifetimes. We observed that individual humans perform at a high initial performance and exhibit no change in performance as a function of (unsupervised) exposure to objects (Fig. 2*C*), suggesting that humans are already well "trained" on these tasks. In summary, although it is impossible to guarantee or expect that humans and monkeys have identical lifetime experience, we find that, once monkeys were trained, additional experience has little to no effect on the patterns of behavioral performance of either species. We note that this does not imply that monkeys and humans learn new objects at a similar rate, only that their steady-state patterns of behavior are highly comparable.

Object categories consisted of basic-level objects with a single object instance (a single 3D model) per category. Consequently, our results do not speak to the monkeys' ability to generalize across multiple object instances within semantic categories but are instead constrained to judgments of visual similarity of 3D objects. Species differences at the category level are possible. Similarly, past studies have debated about differences in the specificity of the "face-inversion effect" between macaque monkeys and chimpanzees/humans (Bruce, 1982; Vermeire and Hamilton, 1998; Parr, 2011). Our results do not rule out the possibility of such species differences for subordinate-level object recognition behaviors. Future work with semantic object categories or subordinate-level object recognition tasks would thus be important for discerning the limits of the species comparison over all of object recognition behavior.

The observed similarities in monkey and human object recognition behavior are consistent with comparative functional imaging of macaque and human brains that reveal strong species homologies of visual cortical areas (Orban et al., 2004), particularly object-selective regions, based on activity correlations (Mantini et al., 2012) and connectivity (Miranda-Dominguez et al., 2014). Although strict anatomical homologies may be imperfect because of evolution-driven reorganization, functional measures reveal a near-perfect conservation of the ventral visual stream, a hierarchy of visual cortical areas thought to directly underlie object recognition behavior, between macaque monkey and human (Mantini et al., 2012). In particular, the neuronal representations of object categories in the end stage of the ventral stream are matched between monkey and human (Kriegeskorte et al., 2008). Together, the anatomical, physiological, and behavioral similarities between monkeys and humans are consistent with the hypothesis that monkeys and humans share similar neural representations underlying the visual perception of basic-level objects.

Recent advances in machine learning have uncovered high-performing representations for object recognition using deep convolutional neural network models (Krizhevsky et al., 2012; Zeiler and Fergus, 2014). Interestingly, these computational models rival the primate brain for core object recognition behavior (Cadieu et al., 2014) and accurately predict neural responses of high-level visual cortical representations of monkey and humans (Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014). Here, we report that monkey and human behavioral patterns were well predicted by a state-of-the-art deep convolutional neural network model (CNN2013), in contrast to alternative models of the ventral stream (HMAX) and low-level control models (pixels, V1+). Together, these results suggest that current high-performing deep convolutional neural network models may accurately capture the shared representation that directly underlies core basic-level object recognition in both humans and monkeys.

To conclude, systematic comparisons of animal model and human behavior are, to date, mostly lacking in the domain of invariant visual object recognition. Here, we investigated whether this behavior is quantitatively comparable across rhesus monkeys and humans. Our results show that monkeys and humans are statistically indistinguishable on a large battery of basic-level visual object recognition tasks, suggesting that rhesus monkeys and humans may share a neural "shape" representation that directly underlies object perception and supporting the use of the monkey as a closely matched model system for studying ventral stream visual processing and object representation in humans.

References

Brown W (1910) Some experimental results in the correlation of mental abilities1. Br J Psychol 3:296–322.

Bruce C (1982) Face recognition by monkeys: absence of an inversion effect. Neuropsychologia 20:515–521. CrossRef

Cadieu CF, Hong H, Yamins DL, Pinto N, Ardila D, Solomon EA, Majaj NJ, DiCarlo JJ (2014) Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS Comput Biol 10:e1003963. CrossRef Medline

Crump MJ, McDonnell JV, Gureckis TM (2013) Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 8:e57410. CrossRef Medline

De Valois RL, Morgan H, Snodderly DM (1974a) Psychophysical studies of monkey vision-III. Spatial luminance contrast sensitivity tests of macaque and human observers. Vision Res 14:75–81. CrossRef Medline

De Valois RL, Morgan HC, Polson MC, Mead WR, Hull EM (1974b) Psychophysical studies of monkey vision—I. Macaque luminosity and color vision tests. Vision Res 14:53–67. CrossRef Medline

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11:333–341. CrossRef Medline

DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434. CrossRef Medline

Gagin G, Bohon KS, Butensky A, Gates MA, Hu JY, Lafer-Sousa R, Pulumo RL, Qu J, Stoughton CM, Swanbeck SN, Conway BR (2014) Color-detection thresholds in rhesus macaque monkeys and humans. J Vis 14:12. CrossRef Medline

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci 15:536–548. CrossRef Medline

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154. CrossRef Medline

Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput Biol 10:e1003915. CrossRef Medline

Kiorpes L, Li D, Hagan M (2008) Crowding in primates: a comparison of humans and macaque monkeys. Perception 37 ECVP Abstract Supplement, p. 37.

Kravitz DJ, Saleem KS, Baker CI, Mishkin M (2011) A new neural framework for visuospatial processing. Nat Rev Neurosci 12:217–230. CrossRef Medline

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141. CrossRef Medline

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inform Process Syst 25:1097–1105.

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920. CrossRef Medline

Macmillan NA (1993) Signal detection theory as data analysis method and psychological decision model. In: A handbook for data analysis in the behavioral sciences: methodological issues (Keren G, Lewis C, eds.), pp. 21–57. Hillsdale, NJ: Lawrence Erlbaum Associates.

Majaj NJ, Hong H, Solomon EA, DiCarlo JJ (2012) A unified neuronal population code fully explains human object recognition. Presented at the Ninth Annual Computational and Systems Neuroscience (COSYNE) Meeting, Salt Lake City, UT, February 23–26.

Mantini D, Hasson U, Betti V, Perrucci MG, Romani GL, Corbetta M, Orban GA, Vanduffel W (2012) Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. Nat Methods 9:277–282. CrossRef Medline

Minamimoto T, Saunders RC, Richmond BJ (2010) Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. Neuron 66:501–507. CrossRef Medline

Miranda-Dominguez O, Mills BD, Grayson D, Woodall A, Grant KA, Kroenke CD, Fair DA (2014) Bridging the gap between the human and macaque connectome: a quantitative comparison of global interspecies structure-function relationships and network topology. J Neurosci 34:5552–5563. CrossRef Medline

Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. Trends Neurosci 6:414–417. CrossRef

Mutch J, Lowe DG (2008) Object class recognition and localization using sparse features with limited receptive fields. Int J Comp Vis 80:45–57. CrossRef

Okamura JY, Yamaguchi R, Honda K, Wang G, Tanaka K (2014) Neural substrates of view-invariant object recognition developed without experiencing rotations of the objects. J Neurosci 34:15047–15059. CrossRef Medline

Orban GA, Van Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys and humans. Trends Cogn Sci 8:315–324. CrossRef Medline

Parr LA (2011) The inversion effect reveals species differences in face processing. Acta Psychologica 138:204–210. CrossRef Medline

Passingham R (2009) How good is the macaque monkey model of the human brain? Curr Opin Neurobiol 19:6–11. CrossRef Medline

Pinto N, Cox DD, DiCarlo JJ (2008) Why is real-world visual object recognition hard? PLoS Comput Biol 4:e27. CrossRef Medline

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025. CrossRef Medline

Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P (1976) Basic objects in natural categories. Cognit Psychol 8:382–439. CrossRef

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T (2007) Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 29:411–426. CrossRef Medline

Spearman C (1910) Correlation calculated from faulty data. Br J Psychol 3:271–295.

Tanaka K (1996) Inferotemporal cortex and object vision. Ann Rev Neurosci 19:109–139. CrossRef Medline

Tootell RB, Tsao D, Vanduffel W (2003) Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. J Neurosci 23:3981–3989. Medline

Ullman S, Humphreys GW (1996) High-level vision: object recognition and visual cognition. Cambridge, MA: Massachusetts Institute of Technology.

Van Essen DC (1979) Visual areas of the mammalian cerebral cortex. Annu Rev Neurosci 2:227–263. CrossRef Medline

Vázquez P, Cano M, Acuña C (2000) Discrimination of line orientation in humans and monkeys. J Neurophysiol 83:2639–2648. Medline

Vermeire BA, Hamilton CR (1998) Inversion effect for faces in split-brain monkeys. Neuropsychologia 36:1003–1014. CrossRef Medline

Vogels R, Orban GA (1990) How well do response changes of striate neurons signal differences in orientation: a study in the discriminating monkey. J Neurosci 10:3543–3558. Medline

Wan L, Zeiler M, Zhang S, Cun YL, Fergus R (2013) Regularization of neural networks using DropConnect. In: Proceedings of the 30th International Conference on Machine Learning, Atlanta, June 16–21.

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci U S A 111:8619–8624. CrossRef Medline

Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. arXiv:1311.2901 [cs.CV].

Zoccolan D, Oertelt N, DiCarlo JJ, Cox DD (2009) A rodent model for the study of invariant visual object recognition. Proc Natl Acad Sci U S A 106:8748–8753. CrossRef Medline