A rodent model for the study of invariant visual object recognition

Davide Zoccolan*abc,1, Nadja Oertelt*abc,1, James J. DiCarlob, and David D. Cox*2

*The Rowland Institute at Harvard, Harvard University, Cambridge, MA 02142; bMcGovern Institute for Brain Research and the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and cNeurobiology and Cognitive Neuroscience Sectors, International School for Advanced Studies, 34014 Trieste, Italy

Edited by Nancy G. Kanwisher, Massachusetts Institute of Technology, Cambridge, MA, and approved April 2, 2009 (received for review November 17, 2008)

The human visual system is able to recognize objects despite tremendous variation in their appearance on the retina resulting from variation in view, size, lighting, etc. This ability—known as “invariant” object recognition—is central to visual perception, yet its computational underpinnings are poorly understood. Traditionally, nonhuman primates have been the animal model-of-choice for investigating the neuronal substrates of invariant recognition, because their visual systems closely mirror our own. Meanwhile, simpler and more accessible animal models such as rodents have been largely overlooked as possible models of higher-level visual functions, because their brains are often assumed to lack advanced visual processing machinery. As a result, little is known about rodents’ ability to process complex visual stimuli in the face of real-world image variation. In the present work, we show that rats possess more advanced visual abilities than previously appreciated. Specifically, we trained pigmented rats to perform a visual task that required them to recognize objects despite substantial variation in their appearance, due to changes in size, view, and lighting. Critically, rats were able to spontaneously generalize to previously unseen transformations of learned objects. These results provide the first systematic evidence for invariant object recognition in rats and argue for an increased focus on rodents as models for studying high-level visual processing.

We recognize visual objects with such ease, it is natural to overlook what an impressive computational feat this represents. Any given object can cast an infinite number of different images onto the retina, depending on the object’s position relative to the viewer, the configuration of light sources, and the presence of other objects. Despite this tremendous variation, we are able to rapidly recognize thousands of distinct object classes without apparent effort. At present, we know little about how the brain achieves robust, “invariant” object recognition, and reproducing this ability remains a major challenge in the construction of artificial vision systems (1).

Animal models provide a critical tool in the investigation of invariant object recognition by allowing the direct study of the neuronal substrates of invariance. Currently, nonhuman primates are the model-of-choice in the study of the mechanisms underlying object vision, because their visual systems closely mirror our own (2). However, while nonhuman primates have many advantages as a model system, there are many disadvantages as well. Experiments are slow and labor-intensive, typically involving small numbers of subjects, and genetic, molecular, and highly invasive manipulations are often not practical.

In contrast, rodent models have long been valued for their superior experimental accessibility, with a wide range of powerful techniques in widespread use (see Discussion). However, the vision science community has largely overlooked rodents as a model, because their brains are often assumed to lack advanced visual processing machinery. Such assumptions are based, in part, on the observations that rodents have lower visual acuity than primates (e.g., approximately one cycle/degree in pigmented rats) (3–7) and make extensive use of their whiskers (8, 9) and sense of smell (10, 11) when exploring their environment.

At the same time, rodent vision, as a whole, has not been completely ignored; there is substantial literature concerning vision in the pigmented rat, starting in the first half of the last century (7, 12) and extending to more modern investigations of visual development (13–16) and memory using the visual modality (17–27). However, while contributing to our understanding of neuronal plasticity in low-level visual areas and the anatomical substrates of learning and memory, this literature has paid less attention to mid- to high-level processing of visual objects.

The one study that has specifically looked at invariant recognition in rats (24) concluded that rats lack robust, general invariant recognition abilities, reinforcing the idea that rodents are not suitable visual models for complex visual phenomena, except perhaps as a means to some other end (e.g., memory). In the present work, we challenge the notion that rats are incapable of invariant recognition, training rats to recognize visual objects despite substantial variation in their appearance. Critically, we show that rats were able to generalize to novel object appearances, including generalization to types of image changes the animals were never previously exposed to in an experimental context. Our results suggest that the rat visual system contains relatively advanced visual processing machinery that allows invariant representation of visual objects. Given the growing demand for simpler and more experimentally accessible model systems, we argue for an increased focus on the rat and other rodents as models for object recognition research.

Results

The goal of our study was to test whether rats are capable of transform-invariant object recognition, i.e., whether they are able to recognize visual objects despite “identity-preserving” changes in their appearance, such as variation in size, orientation, lighting, etc. We tested 6 pigmented rats in an object discrimination task (see Fig. 1) that consisted of 4 phases. During the initial phase, the animals were trained to discriminate between the default (fixed) views of 2 visual objects (shown in Fig. 1A). During the second phase, the rats were trained to perform the task in the face of some amount of variation in each object’s appearance. Finally, during the third and fourth phases, we tested the rats’ capability to spontaneously generalize their recognition to a large collection of novel identity-preserving transformations of the target objects.
broad set of object appearances (light blue frames in Fig. 2) at least 70% correct discrimination performance across a wide range of transformations (i.e., within an octave of variation in size (i.e., between 40° and 15° visual angle) and a 30° difference in azimuth rotation; see diagram in Fig. 3). The animals readily acquired this task, achieving (in 25–30 training sessions) more than an octave of variation in size (i.e., between 40° and 15° visual angle) and a 30° difference in azimuth rotation; see diagram in Fig. 3). Consequently, at the end of phase II, subjects were required to tolerate (for details see Zoccolan et al. PNAS). Animals were tested in phase III. Performance was high and significantly above chance (one-tailed Binomial test) was highly significantly above chance (P < 0.001) for nearly all of the previously unseen transformations (stimulus conditions outside the light blue frames) and, crucially, also for the fraction of transformations for which feedback was withheld (stimulus conditions inside the black frames in Fig. 2B). As expected due to the animal’s relatively low visual acuity, performance was impaired only at small object sizes. A summary group mean performance over the no-feedback conditions is shown in Fig. 3A. Animals’ performance was not significantly different for stimulus conditions where they did not receive feedback (second bar), versus those (size-matched) conditions for which they did (third bar), and, in both cases, was high and not significantly different from the performance over the stimulus conditions that had been trained during phase II (first bar; one-way ANOVA, P = 0.51). This indicates that receiving feedback during training was not critical for achieving high generalization performance. Performance was similarly high for the special case stimulus conditions (white bar in Fig. 3A) that were never fewer than 2 “squares” away from the nearest rewarded condition (i.e., 25% difference in size, 30° difference in azimuth rotation; see diagram in Fig. 3A).

To further explore the degree of automaticity in generalization, the animals’ group mean performance was plotted (Fig. 3B) as a function of presentation number (e.g., first presentation, second, third, etc.) of all of the novel transformations (i.e., outside the trained axes; see diagram in Fig. 3B), with or without feedback, that were tested in phase III. Performance was high and significantly above chance (one-tailed t test, P = 0.002), even for the very first presentation of each novel stimulus, and remained stable over the course of the experiment, with no significant variation as a function of presentation number (one-way ANOVA, P = 0.87). This was true also for the performance of each individual subject (one-tailed Binomial test, P < 0.05). This high initial performance level and stability across time indicates that generalization of rat recognition to previously unseen transformations of known visual objects.
behavior was spontaneous and cannot be explained by learning taking place during phase III. Indeed, it appears that animals immediately achieved their maximal performance and did not significantly improve thereafter.

**Generalization to a Novel Type of Image Variation.** To test rats’ generalization ability further, in phase IVa of our experiment, we created 15 additional stimulus conditions for each target object by varying the position of the virtual light source for 15 arbitrary size-azimuth conjunctions from the previous phase (see Fig. 4A and Fig. S4A). In most cases, this manipulation produced large changes in the pixel-level appearance of the objects, changing the mean luminance level and contrast, and in many cases inverting the relationships of which portions of the objects were light or dark (see examples at the bottom of Fig. 4A). We interleaved these new lighting conditions trials with the full “matrix” of size-azimuth transformations from phase III. Importantly, these novel lighting condition stimuli were never rewarded (nor was any feedback given), and rats received no training with any stimuli under lighting aside from the “default” lighting condition (e.g., as in phases I–III). Performance was high overall (approximately 75%) and significantly above chance for 14 out of 15 of the novel lighting conditions (Fig. 4C; one-tailed t test, see legend for significance levels) and was roughly comparable to performance with the default lighting. These results indicate that the animals were still able to recognize the objects despite the large pixel-wise disruption of images induced by the novel lighting conditions.

While performance was robust overall, we observed some decrement in performance between the matched “default” lighting and novel lighting conditions (Fig. 4C). It is not clear whether this decrement was due to incomplete generalization or simply the fact that the novel lighting condition images were overall substantially darker and lower contrast than their “default” lighting counterparts. Indeed, the one lighting condition for which the rats did not perform above chance (black circle, Fig. 4C) was by far the darkest and lowest-contrast image (Fig. S4A and B). More generally, the decrement between performance in the novel lighting conditions and their “default” lighting counterparts was strongly predicted by the contrast of each novel lighting condition image ($t = 0.88$, $P < 10^{-4}$, two-tailed t test; see Fig. S4B). Given the known properties of the rat’s contrast sensitivity function (3, 4, 6), it is likely that the novel lighting conditions were somewhat more difficult simply because they were somewhat harder to see.

As a final test of rats’ generalization abilities, in phase IVb of our experiment, we created 15 novel stimulus conditions by varying the elevation rotation of each target object (i.e., in-depth rotation about the horizontal axis) by $\pm 10^\circ$ or $\pm 20^\circ$ and then combining these new transformations with 15 arbitrary size-azimuth conjunctions from phase III (see Fig. 4B and Fig. S5). Despite the fact that these transformations resulted in substantial changes of the objects’ bounding contour and were novel to the animals, performance was significantly above chance for all new elevation conditions and was close to their performance with the default ($0^\circ$) elevation conditions (Fig. 4D). Taken together, these results indicate that rats were able to generalize to novel object transformations.
Discussion

Our study provides systematic evidence that rats are capable of invariant visual object recognition, an advanced visual ability that has only been ascribed to a select few species. While the “front end” of rat vision is clearly of lower acuity than primates, rats nonetheless possess at least some sophisticated “back end” processing that enables the recognition of complex form in the face of real-world image variation. This finding opens up largely unexplored experimental avenues for probing the mechanisms of invariant visual object recognition.

Fig. 3. Generalization of recognition performance. (A) Mean performances obtained by pooling across different subsets of object conditions tested during phase III (error bars indicate SEM). The first grouping of bars (in gray) shows performance with previously trained object transformations (first bar), with the set of novel transformations for which feedback was withheld (second bar), and with a size-matched (i.e., acuity-matched) subset of novel transformations for which feedback was provided (third bar). Diagrams below each bar show which conditions were included in each subset according to the convention set forth in Fig. 2. Performances over these 3 groups of conditions were all significantly higher than chance (one-tailed t test; *** $P < 0.001$) but not significantly different from each other. The white (fourth) bar shows the performance over the special case “no-feedback” condition that was always separated from the nearest “feedback” condition by at least 10° in size and 30° in azimuth. Such a condition existed only within the top-left and the top-right no-feedback quadrants (see diagram) and was tested for rats R2, R5, R3, and R6 (see Fig. 2 B, Right). (B) Group mean performance ($n = 6$; black line) over the full set of novel object transformations tested during phase III, computed for the first, second, third, etc., presentation of each object in the set (shaded area shows the SEM). All performances along the curve were significantly above chance (one-tailed t test, $P < 0.005$) and were not significantly different from each other.

Limitations and Implications of Our Findings and Comparison with Previous Studies.

Our findings contradict the one report that previously looked at invariant object recognition in rats (24), which concluded that rats rely on low-level image cues (e.g., luminance in the lower half of the stimulus display) rather than on more advanced, invariant shape processing. In addition to numerous methodological differences, a crucial difference between our study and that of Minini and Jeffrey (24) is the fact that our animals were extensively trained in a task that required them to discriminate the target objects despite variation in their size and viewpoint (phase II). Although our subjects were ultimately tested with novel object appearances (phase III and IV), an extended period of familiarization with a transformation-invariant task was likely a key factor in inducing them to adopt an object-invariant or, at least, feature-invariant strategy. On the contrary, in the task used by Minini and Jeffrey (24), computation of lower hemifield luminance was a perfectly “valid” solution to the task at hand (i.e., it was one of many possible strategies for maximizing reward within the context of the experiment). Thus, within a highly limited experimental context, it
may be difficult to distinguish between an inability to perform invariant recognition, and an effective but less general choice of strategy. Interestingly, even monkeys (which are widely assumed to possess advanced visual abilities) sometimes default to a simpler, "lower-level" strategy if it is effective for solving a given task. For instance, in the work of Nielsen and Logothetis (28), monkeys chose to rely on small, diagnostic patches of a phase-scrambled noise background, rather than focus on a foreground object, as humans did when performing the same task. It has also been repeatedly demonstrated that monkeys do not spontaneously generalize to significant rotations in-plane or in-depth without exposure to multiple views along the axis of rotation (29–31). Adoption of a simpler strategy in an experimental task does not imply that an animal is incapable of adopting more complex strategies as needed.

More generally, our results extend the existing literature concerning rats' visual abilities (3, 4, 6, 22, 24) and memory processes (17–27, 32). While these studies have not looked at invariant recognition specifically, they nonetheless show that rats are adept at using vision to solve a variety of tasks involving complex shapes. However, because these previous studies primarily used two-dimensional shape stimuli, without any variation in individual object appearance, it is difficult to rule out "low-level" accounts of the animals' performance, such as those offered by Minini and Jeffrey (e.g., luminance confounds). With only a small number of target stimuli, it is impossible to simultaneously control for all such low-level confounds, since equalizing one low-level property (e.g., lower-field luminance) invariably produces a difference in another (e.g., upper-field luminance). In our study, we required rats to recognize objects in the face of substantial variation in object view (both azimuth and elevation), size, and lighting. Variation along these axes produced large and complex changes in the pixel-level appearance of the objects, disrupting low-level confounds, and resulting in a situation where the pixel-wise image differences within the different appearances (sizes, views, and lightings) of the same object were much greater than the image differences between the 2 objects when matched for size, view, and lighting (Fig. S3; notably, this was true also when the within- and the between-object image differences were computed over the responses of a population of simulated V1-like simple cells, see Fig. S6).

More generally, while our results cannot speak to whether rats employ similar neuronal hardware as humans or monkeys do, their abilities nonetheless meet a reasonable operational definition of "invariant object recognition." It is important to note that this need not imply that the rat visual system builds complex three-dimensional geometric representations of whole objects. Indeed, it is possible, and even likely, that rats achieve invariant object recognition through the use of the invariant representation of subfeatures that are smaller or less complex than the entire object. That is, rats could use feature detectors that respond to only a subportion of the object, or are sensitive to key two-dimensional, rather than three-dimensional, features. Critically, however, whatever feature is used, it must be tolerant to the wide range of image-level variation in position, scale, orientation, lighting, etc., found in our stimulus set. The potential use of such invariant subfeatures in no way diminishes the sophistication of the rat visual system—such invariant subfeature detectors are not found in the early visual system, and many leading computational models of how invariant recognition is achieved in primates explicitly rely on such subfeatures (33–36).

Given the above results, the rat seems an attractive model system for the study of object recognition, complementing work done in primate species. In particular, rat studies allow the application of a powerful array of techniques that are currently very difficult or impossible to apply in nonhuman primates, including molecular (37, 38) and histological (27) approaches, two-photon imaging (39), large-scale recordings from multiple brain areas (40), and in vivo patch clamp in awake animals (41, 42). In addition, our findings raise the possibility that other rodent species, with higher accessib-

bility to genetic approaches (mice) or more developed visual systems (e.g., squirrels) (43), might also be valuable models for higher-level vision.

The brains of rats are clearly less advanced as compared to nonhuman primates; however, this is potentially an asset. Indeed, when attempting to understand a particular phenomenon, it is often wise to seek out the simplest system that demonstrates the properties of interest. While we would not suggest that the rat should or could replace "higher" organisms such as monkeys in the study of object recognition, the availability of a simpler system that possesses many of the same fundamental features provides an important additional window onto the computational problem of object recognition.

Materials and Methods

Subjects. Six adult male Long-Evans rats (Charles River Laboratories) were used for behavioral testing. Animals weighed approximately 250 g at the onset of training and grew to over 500 g. Rats were water-restricted throughout the training period (see the schematic in Fig. 1B). Water was provided in a bottle kept at approximately 5 cm from the bottom of the cage and was available at all times. Rats were water deprived to 70% of their free feeding weight. Rats were then returned to free feeding weight 1–2 d before being placed in their home cages. Animals weighed approximately 250 g at the onset of training and grew to over 500 g. Rats were water-restricted throughout the training period (see the schematic in Fig. 1B). Water was provided in a bottle kept at approximately 5 cm from the bottom of the cage and was available at all times. Rats were water deprived to 70% of their free feeding weight. Rats were then returned to free feeding weight 1–2 d before being placed in their home cages.

Behavioral Rig and Task. The training/testing apparatus consisted of an operant box that was equipped with a 19" LCD monitor for presentation of visual stimuli, as well as automated fluid dispensers for liquid reward delivery ("reward ports"), and capacitive contact sensors for initiation of behavioral trials and collection of responses (see the schematic in Fig. 1B). All stimuli were presented on a Samsung SyncMaster 940BX LCD monitor for presentation of visual stimuli, and capacitive contact sensors for initiation of behavioral trials and collection of responses (see the schematic in Fig. 1B). All stimuli were presented on a Samsung SyncMaster 940BX LCD monitor for presentation of visual stimuli, and capacitive contact sensors for initiation of behavioral trials and collection of responses (see the schematic in Fig. 1B).

Experimental Design. Phase I. Initially, rats were trained to discriminate between the default views of the 2 target objects shown in Fig. 1B. During this stage of training, the target objects were not subject to any image variation, i.e., their size, in-depth rotation, and lighting were kept constant.

Phase II. Once a rat achieved >70% correct discrimination in phase I, we gradually introduced variations in the appearance of the target objects and required the animal to make the discrimination despite such variations. Each animal was first trained to tolerate an increasing amount of size variation using an adaptive staircase procedure (Fig. S1) that, based on the animal performance, updated the
lower bound of the range from which the object size was sampled (the upper bound was 40° visual angle). Object size was randomly sampled within the range defined by the lower and upper bound, so as to force the animals to learn a truly size-invariant recognition task (see Fig. S1). While an animal was trained with the size staircase, the objects' in-depth rotation was held fixed at its default 0° value. Once the sizes' lower bound reached a stable (asymptotic) value across consecutive training sessions (typically 10–15 sessions; Fig. S2A), a similar staircase procedure was used to train the rats to tolerate azimuth rotations of the target objects (i.e., in-depth rotations about objects' vertical axis), while object size was fixed at the lower bound (i.e., S52B). After completion of the staircase training sessions, each animal's overall performance across the trained axes of variations (i.e., size and azimuth) was assessed by uniformly sampling object conditions along these axes (light blue frames, Fig. 2A) over the course of 2–5 additional sessions.

Phase III. In this phase, we tested the rats' ability to recognize the target objects across a large range of novel object appearances. This was possible because, during phase II, while a rat was trained to tolerate variation along one transformation axis (e.g., size), the other object properties (e.g., in-depth rotation and lighting) were kept fixed. Therefore, it was possible to build many combinations of the object transformations that had not been previously seen. We combined 6 sizes and 9 azimuths to yield a total of 54 unique transformations of each target object (shown in Fig. 2A), 40 of which were novel (objects outside the light blue frames in Fig. 2A). The average pixel-level difference between the appearances of the target object (i.e., within object identity) was almost 2 times larger than the average image difference between the 2 objects, when matched for size and azimuth (i.e., between object identities; see Fig. S3). Importantly, for a fraction of transformations (approximately 11%), covering a contiguous quadrant of the size-view matrix, see Fig. 2B, Right), the animal was not rewarded, and no reinforcement sounds were played (after the animal made any response, the stimulus was removed, the trial ended, and the animal was allowed to immediately trigger the next trial). Interestingly, rats were not disturbed by this fraction of non-rewarded trials, as evidenced by the high recognition performance achieved for these no-feedback transformations (see black frames in Fig. 2B, Right and Fig. 3A).

Phase IVa. We further tested rats' generalization abilities by creating 15 new appearances of each target object by varying the position of the virtual light source for 15 arbitrary size-azimuth conjunctions from the previous phase (see Fig. 4A and Fig. S4A). New lighting conditions were divided in 3 subsets of 5, and each subset was presented, interleaved with 45 default lighting conditions from the previous phase (i.e., the full matrix of size-azimuth conjunctions shown in Fig. 2A, with the exception of the 15° size), for 5–10 sessions. For all of the new lighting conditions, no feedback on the correctness of response was provided to the animals.

Phase IVb. Similarly to Phase IVa, we built 15 novel appearances of each target object by varying its elevation (i.e., the in-depth rotation about its horizontal axis) by either ±10° or ±20°, and then combining this new in-depth rotation with 15 arbitrary size-azimuth conjunctions from phase III (see Fig. 4B and Fig. S5). The presentation protocol was identical to that used in phase IVa. Five rats participated to phases IVa and IVb.

For further information, see SI Text.

ACKNOWLEDGMENTS. This work was supported by the McGovern Institute for Brain Research (MIT), The McKnight Foundation, and the Rowland Institute at Harvard. D.Z. was supported by a Charles A. King Trust Postdoctoral Fellowship and by an Academia Nazionale dei Lincei—Compagnia di San Paolo Grant. We thank Mark Everitt, Paul Ungerleider, Deepak Srivastava, Siva Musial, Rushil Panacek, John Pierce, and the Zoccolan lab members for helpful discussions. We thank Pauline Gassman, Crystal Paul-Laughinghouse, Basma Radwan, and Sabrina Tsang for help in training animals. We also thank E. Flister, P. Meier, D. Pichette, P. Reinagel, J. Ritt, and G. Westermeier for helpful discussions.