A rodent model for the study of invariant visual object recognition

Davide Zoccolan[a,b,c,1], Nadja Oertelt[a,1], James J. DiCarlo[b], and David D. Cox[a,2]

[a]The Rowland Institute at Harvard, Harvard University, Cambridge, MA 02142; [b]McGovern Institute for Brain Research and the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]Neurobiology 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.

invariance | vision | rat | behavior

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. 1*A*). 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.

Author contributions: D.Z., N.O., J.J.D., and D.D.C. designed research; D.Z., N.O., and D.D.C. performed research; D.Z., N.O., and D.D.C. analyzed data; and D.Z., N.O., J.J.D., and D.D.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]D.Z. and N.O. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: cox@rowland.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0811583106/DCSupplemental.

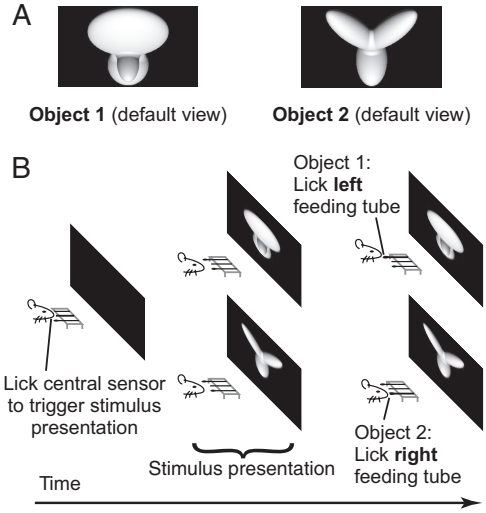

Fig. 1. Visual stimuli and behavioral task. (*A*) Default views (0° in-depth rotation) of the target objects that rats were trained to discriminate during phase I of the study (each object's default size was 40° visual angle). (*B*) Rats were trained in an operant box that was equipped with an LCD monitor, a central touch sensor, and 2 flanking feeding tubes (also functioning as touch sensors). Following initiation of a behavioral trial (triggered by the rat licking the central sensor), 1 of the 2 target objects was presented on the monitor, and the animal had to lick either the left or the right feeding tube (depending on the object identity) to receive reward.

Learning an Invariant Object Recognition Task. Rats were trained to discriminate between 2 three-dimensional synthetic objects (Fig. 1*A*) presented on an LCD monitor. Animals initiated each behavioral trial by inserting their heads through a narrow hole in one of the walls of the training box and touching their tongues to a centrally aligned touch sensor (Fig. 1*B*). This arrangement allowed a reasonably reproducible positioning of the animals' heads during stimulus presentation and good control over the viewing distance.

During phase I of our study, each object was presented in the center of the monitor at a fixed, default size (40° visual angle) and viewpoint (0° in-depth orientation), with a single, consistent light source (in front of the object; see Fig. 1*A*). Phase I typically lasted 25 sessions, during which the animals learned a fixed mapping between object identity (object 1/object 2) and reward port (left/right; see Fig. 1*B*). Phase I was completed when the animals achieved >70% correct object discrimination.

During phase II, rats were required to perform the same object discrimination task, but while the size of the target objects and their azimuth rotation (i.e., in-depth rotation about the objects' vertical axis) were separately varied (i.e., the object sometimes appeared smaller, or rotated, but never appeared smaller and rotated). Variation in object appearance along each dimension was introduced and gradually increased using an adaptive staircase procedure that updated the range of object variation that animals were required to tolerate (for details see *Materials and Methods* and Fig. S1). Consequently, at the end of phase II, subjects were required to discriminate between the objects, in interleaved trials, despite more than an octave of variation in size (i.e., between 40° and 15° visual angle) and ±60° of variation in object azimuth rotation (Fig. S2). Animals rapidly acquired this task, achieving (in 25–30 training sessions) at least 70% correct discrimination performance across a broad set of object appearances (light blue frames in Fig. 2*A*).

Generalization to a Large Set of Novel Object Transformations. Phase II of our study (see Fig. S2) already demonstrates that rats are able to recognize objects despite a large range of image variation, at least for images for which they have received training. However, a still more critical feature of invariant recognition is the ability to generalize to previously unseen transformations of known visual objects. To test rats' generalization abilities, in phase III, we constructed a large set of novel transformations of each target object (Fig. 2*A*; see *Materials and Methods* for details). This image set consisted of novel combinations of objects' size and azimuth rotation (Fig. 2*A*, outside the light blue frames). We then asked whether rats were still able to correctly identify the target objects, despite being presented across this substantially new range of variation in appearance (Fig. S3), in interleaved trials with previously trained appearances.

While generalization performance would ideally be measured in the absence of feedback, in practice, animals will not perform large numbers of unrewarded trials. Thus, to overcome potential training effects over the course of probing generalization performance, we designed phase III with 3 critical features. First, a large number of new transformations were used (80 unique appearances for each target object; Fig. 2*A*) making it unlikely that the rats could memorize the association between each of them and the corresponding response. Second, we withheld feedback (i.e., the animal did not receive any reward or feedback tone) for a fraction of the new transformations. Third, we measured the rats' overall performance on the very first presentation of each of the novel transformation conditions, allowing performance to be assessed before any additional training.

Group mean performances across the full set of object transformations tested during phase III are shown in Fig. 2*B* (*Left*), along with each individual animal's performance across all stimulus conditions (Fig. 2*B*, *Right*). Note that because our experimental rig allowed for the collection of hundreds of trials per day, we were able to collect 70–90 trials per stimulus condition, per animal, allowing us to assess significance without resorting to pooling of animals or stimulus conditions. As Fig. 2*B* shows, both the group mean performance (*Left*; one-tailed t test) and each individual rat's performance (*Right*; 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. 2*B*, *Right*). 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. 3*A*. 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. 3*A*) 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. 3*A*).

To further explore the degree of automaticity in generalization, the animals' group mean performance was plotted (Fig. 3*B*) 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. 3*B*), 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

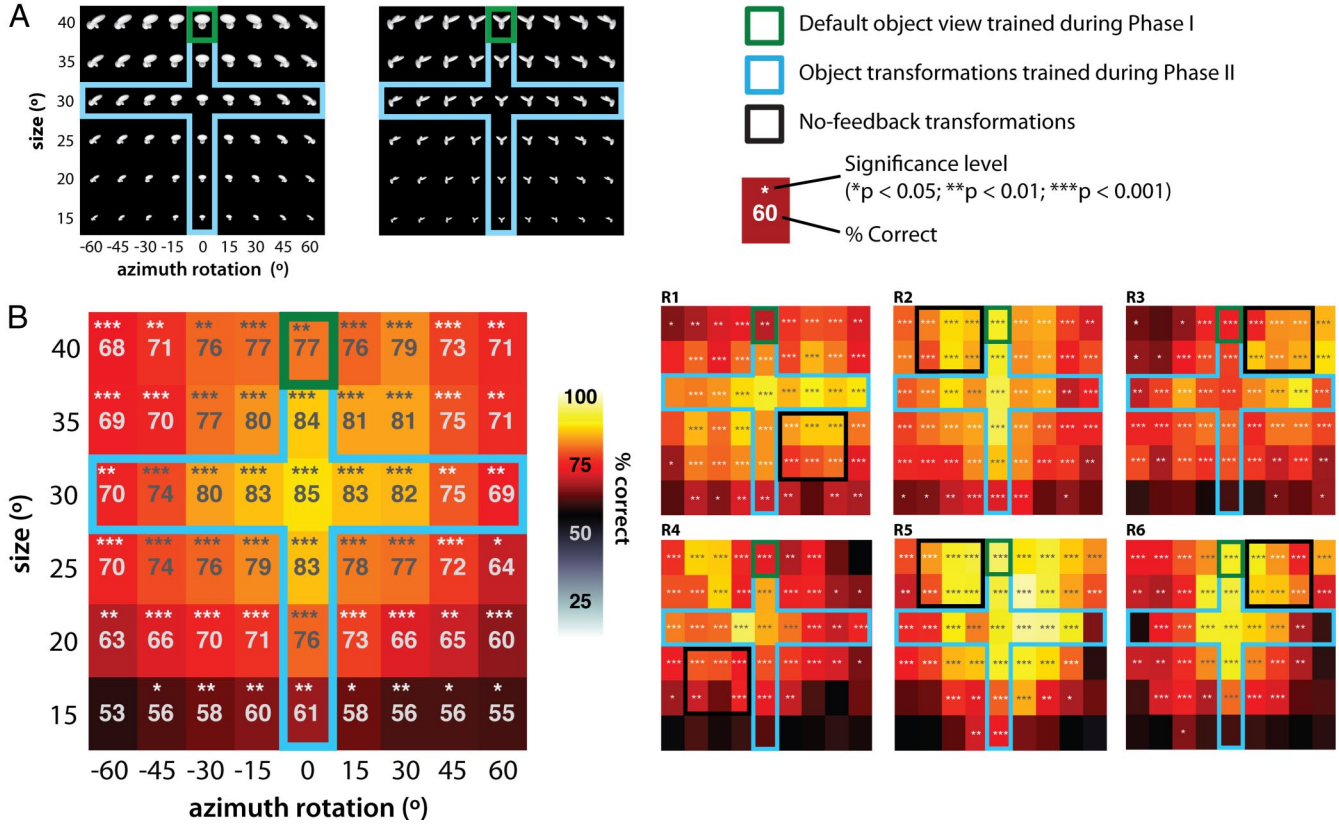

Fig. 2. Rats' group mean and individual performances across the range of object transformations tested during phase III of the study. (*A*) The set of object transformations used in phase III, consisting of all possible combinations of 6 sizes and 9 in-depth azimuth rotations of both target objects. Green frames show the default object views trained during phase I. Light blue frames show the subset of transformations (14) for each object during phases II. All of the remaining transformations (40 for each object) were novel to the animals. (*B*) Left plot shows the animals' group mean performance (n = 6) for each of the tested object transformations depicted in *A*—the percentage of correct trials is both color-coded and reported as a numeric value, together with its significance according to a 1-tailed *t* test (see key for significance levels). The right plots show the performance of each individual subject for each object condition, and its significance according to a one-tailed Binomial test (see key for significance levels). Black frames show the quadrants of adjacent transformations for which feedback was not provided to the subjects, to better assess generalization (counterbalanced across animals).

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. 4*A* and Fig. S4*A*). 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. 4*A*). 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. 4*C*; 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. 4*C*). 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. 4*C*) was by far the darkest and lowest-contrast image (Fig. S4*A* 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 ($r = 0.88$, $P < 10^{-4}$, two-tailed *t* test; see Fig. S4*B*). 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 ±10° or ±20° and then combining these new transformations with 15 arbitrary size-azimuth conjunctions from phase III (see Fig. 4*B* 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°) elevation conditions (Fig. 4*D*). Taken together, these results indicate that rats were able

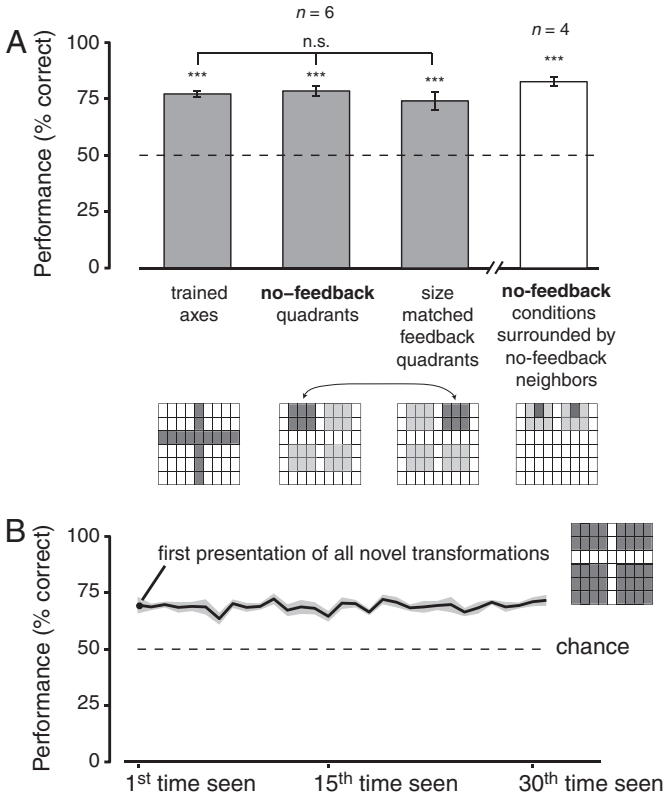

Fig. 3. Generalization of recognition performance. (*A*) Mean performances obtained by pooling across different subsets of object conditions tested during phase III (error bars indicate SEM). The first grouping of bars (in gray) shows performance with previously trained object transformations (first 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.

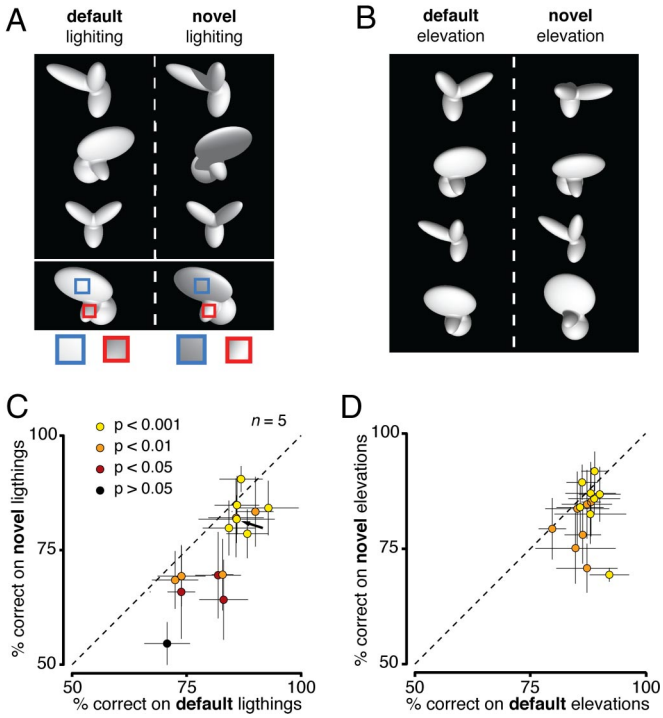

Fig. 4. Generalization to novel lighting and elevation conditions. (*A*) Examples of lighting conditions tested during phase IVa of our study, with the first column showing default lighting conditions (i.e., same as during phase I–III) and the second column showing novel lighting conditions. The bottom examples show how manipulating lighting often produced a reversal in the relative luminance of different areas over the surface of an object. Under default lighting, the blue-framed image region was brighter than the red-framed region (top), but this relationship was reversed under the novel lighting condition (bottom). (*B*) Examples of elevation conditions tested during phase IVb of our study, with the first column showing default (0°) elevation conditions (i.e., same as during phase I–IVa) and the second column showing novel (±10° and ±20°) elevation conditions. Note the variation in the objects' silhouette produced by changing the elevation. (*C*) Rats' mean performance with the novel lighting conditions (ordinate) is plotted against performance with the matching ''default'' lighting conditions (abscissa). Performance on the novel lighting conditions was high overall, and in all but one condition was significantly above chance (one-tailed *t* test; see key for significance levels). The black arrow indicates the performance over the bottom example conditions shown in *C*. (*D*) Rats' mean performance with the novel elevation conditions (ordinate) is plotted against performance with the matching ''default'' elevation conditions (abscissa). Color convention as in *C*. In both *C* and *D*, error bars indicate standard errors of the means.

to generalize across a wide range of previously unseen object transformations, including novel combinations of size and in-depth rotations of the learned objects (phase III), new lighting/shading patterns over the objects' surface (phase VIa), and also substantial variation in object silhouette (phase IVb).

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.

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 accessibility 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 experiments, with each animal typically receiving several milliliters of water (4–10 mL depending on the weight) as a reward during each training/testing session (ad libitum water was additionally available for 1 hour after each session). All animal procedures were performed in accord with the National Institute of Health guidelines, the Massachusetts Institute of Technology Committee on Animal Care and the Harvard Institutional Animal Care and Use Committee.

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. 1*B*). One wall of the box contained a 3-cm diameter hole (27) that allowed the animal to extend its head outside the box and frontally face the monitor, approximately 250 mm in front of the rat's eyes. An array of 3 contact sensors was also placed at 3 cm from the opening (the sensors were approximately 10 mm apart from each other). The terminal part of each sensor was made out of hypodermic tubing and could deliver a flavored milk reward via computer-controlled syringe pumps (New Era Pump Systems).

Rats were trained in a visual object recognition task that required them to discriminate between 2 target objects (Fig. 1*A*). Animals were trained to trigger the onset of a behavioral trial by licking the central contact sensor. This prompted 1 of the 2 target objects to be displayed on the monitor (see Fig. 1*B*), and the rat was required to correctly identify which object was presented (i.e., object 1 or object 2; Fig. 1*B*), by licking either the left or the right contact sensor (depending on object identity). For each animal, the mapping from object identity to contact sensor was kept fixed across experimental sessions. Reward was delivered from the corresponding reward port in case of correct identification, and a positive reinforcement sound was played. An incorrect choice yielded no reward and a 1–3 s time out (during which a failure tone sounded and the monitor flickered from black to middle gray at a rate of 5 Hz). Further details are provided in the *SI*.

Visual Stimuli. Each subject was required to discriminate between a pair of 3-lobed visual objects. These objects were renderings of three-dimensional models that were built using the ray tracer POV-Ray (http://www.povray.org/). Fig. 1*A* shows the default (''frontal'') object views used during phase I. Both objects were illuminated from the same light source location and, when rendered at the same in-depth rotation, their views were approximately equal in height, width and area (Fig. 2*A*). Objects were rendered against a black background. Two rats were trained with vertically flipped versions of these same stimuli.

Each object's default (initial) size was 40° of visual angle, and their default position was the center of the monitor (horizontally aligned with the position the rats' head). Stimuli were presented on a Samsung SyncMaster 940BX LCD monitor (1280 × 1024 pixels resolution; 60 Hz refresh rate; 5 ms response time; 300 cd/m^2 maximal brightness; 1000:1 contrast ratio).

Experimental Design. Phase I. Initially, rats were trained to discriminate between the default views of the 2 target objects shown in Fig. 1*B*. 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. S2*A*), 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 30° visual angle (Fig. S2*B*). 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. 2*A*) 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. 2*A*), 40 of which were novel (objects outside the light blue frames in Fig. 2*A*). The average pixel-level difference between the appearances of the same 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. 2*B*, *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 not-rewarded trials, as evidenced by the high recognition performance achieved for these no-feedback transformations (see black frames in Fig. 2*B*, *Right* and Fig. 3*A*).

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. 4*A* and Fig. S4*A*). 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. 2*A*, 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 $\pm 10°$ or $\pm 20°$, and then combining this new in-depth rotation with 15 arbitrary size-azimuth conjunctions from phase III (see Fig. 4*B* 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 Accademia Nazionale dei Lincei—Compagnia di San Paolo Grant. 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. Pritchett, P. Reinagel, J. Ritt, and G. Westmeyer for helpful discussions.

1. Pinto N, Cox DD, DiCarlo JJ (2008) Why is real-world visual object recognition hard? *PLoS Comput Biol* 4:e27.
2. Logothetis NK, Sheinberg DL (1996) Visual object recognition. *Ann Rev Neurosci* 19:577–621.
3. Prusky GT, Harker KT, Douglas RM, Whishaw IQ (2002) Variation in visual acuity within pigmented, and between pigmented and albino rat strains. *Behav Brain Res* 136:339–348.
4. Prusky GT, West PW, Douglas RM (2000) Behavioral assessment of visual acuity in mice and rats. *Vision Res* 40:2201–2209.
5. Birch D, Jacobs GH (1979) Spatial contrast sensitivity in albino and pigmented rats. *Vision Res* 19:933–937.
6. Keller J, Strasburger H, Cerutti DT, Sabel BA (2000) Assessing spatial vision—automated measurement of the contrast-sensitivity function in the hooded rat. *J Neurosci Methods* 97:103–110.
7. Lashley KS (1930) The mechanisms of vision: III. The comparative visual acuity of pigmented and albino rats. *J Genet Psychol* 37:481–484.
8. Diamond ME, von Heimendahl M, Knutsen PM, Kleinfeld D, Ahissar E (2008) 'Where' and 'what' in the whisker sensorimotor system. *Nat Rev Neurosci* 9:601–612.
9. von Heimendahl M, Itskov PM, Arabzadeh E, Diamond ME (2007) Neuronal activity in rat barrel cortex underlying texture discrimination. *PLoS Biol* 5:e305.
10. Uchida N, Mainen ZF (2003) Speed and accuracy of olfactory discrimination in the rat. *Nat Neurosci* 6:1224–1229.
11. Rubin BD, Katz LC (2001) Spatial coding of enantiomers in the rat olfactory bulb. *Nat Neurosci* 4:355–356.
12. Lashley KS (1938) The mechanisms of vision: XV. Preliminary studies of the rat's capacity for detail vision. *J Gen Psychol* 18:123–193.
13. Prusky GT, West PW, Douglas RM (2000) Experience-dependent plasticity of visual acuity in rats. *Eur J Neurosci* 12:3781–3786.
14. Pizzorusso T, et al. (2006) Structural and functional recovery from early monocular deprivation in adult rats. *Proc Natl Acad Sci USA* 103:8517–8522.
15. Sale A, et al. (2007) Environmental enrichment in adulthood promotes amblyopia recovery through a reduction of intracortical inhibition. *Nat Neurosci* 10:679–681.
16. Lu W, Constantine-Paton M (2004) Eye opening rapidly induces synaptic potentiation and refinement. *Neuron* 43:237–249.
17. Davies M, Machin PE, Sanderson DJ, Pearce JM, Aggleton JP (2007) Neurotoxic lesions of the rat perirhinal and postrhinal cortices and their impact on biconditional visual discrimination tasks. *Behav Brain Res* 176:274–283.
18. Eacott MJ, Machin PE, Gaffan EA (2001) Elemental and configural visual discrimination learning following lesions to perirhinal cortex in the rat. *Behav Brain Res* 124:55–70.
19. Gaffan EA, Eacott MJ, Simpson EL (2000) Perirhinal cortex ablation in rats selectively impairs object identification in a simultaneous visual comparison task. *Behav Neurosci* 114:18–31.
20. Bussey TJ, et al. (2008) The touchscreen cognitive testing method for rodents: How to get the best out of your rat. *Learn Mem* 15:516–523.
21. Prusky GT, Douglas RM, Nelson L, Shabanpoor A, Sutherland RJ (2004) Visual memory task for rats reveals an essential role for hippocampus and perirhinal cortex. *Proc Natl Acad Sci USA* 101:5064–5068.
22. Simpson EL, Gaffan EA (1999) Scene and object vision in rats. *Q J Exp Psychol B* 52:1–29.
23. Forwood SE, Bartko SJ, Saksida LM, Bussey TJ (2007) Rats spontaneously discriminate purely visual, two-dimensional stimuli in tests of recognition memory and perceptual oddity. *Behav Neurosci* 121:1032–1042.
24. Minini L, Jeffery KJ (2006) Do rats use shape to solve ''shape discriminations''? *Learn Mem* 13:287–297.
25. Gaffan EA, Bannerman DM, Warburton EC, Aggleton JP (2001) Rats' processing of visual scenes: Effects of lesions to fornix, anterior thalamus, mamillary nuclei or the retrohippocampal region. *Behav Brain Res* 121:103–117.
26. Gaffan EA, Healey AN, Eacott MJ (2004) Objects and positions in visual scenes: Effects of perirhinal and postrhinal cortex lesions in the rat. *Behav Neurosci* 118:992–1010.
27. Wan H, Aggleton JP, Brown MW (1999) Different contributions of the hippocampus and perirhinal cortex to recognition memory. *J Neurosci* 19:1142–1148.
28. Nielsen KJ, Logothetis NK, Rainer G (2006) Discrimination strategies of humans and rhesus monkeys for complex visual displays. *Curr Biol* 16:814–820.
29. Nielsen KJ, Logothetis NK, Rainer G (2008) Object features used by humans and monkeys to identify rotated shapes. *J Vis* 8:9.1–9.15.
30. Wang G, Obama S, Yamashita W, Sugihara T, Tanaka K (2005) Prior experience of rotation is not required for recognizing objects seen from different angles. *Nat Neurosci* 8:1768–1775.
31. Logothetis NK, Pauls J, Bulthoff HH, Poggio T (1994) View-dependent object recognition by monkeys. *Curr Biol* 4:401–414.
32. Zhu XO, Brown MW, Aggleton JP (1995) Neuronal signalling of information important to visual recognition memory in rat rhinal and neighbouring cortices. *Eur J Neurosci* 7:753–765.
33. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.
34. Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci USA* 104:6424–6429.
35. Ullman S, Vidal-Naquet M, Sali E (2002) Visual features of intermediate complexity and their use in classification. *Nat Neurosci* 5:682–687.
36. Biederman I (1987) Recognition-by-components: A theory of human image understanding. *Psychol Rev* 94:115–147.
37. Tan EM, et al. (2006) Selective and quickly reversible inactivation of mammalian neurons in vivo using the *Drosophila* allatostatin receptor. *Neuron* 51:157–170.
38. Griffiths S, et al. (2008) Expression of long-term depression underlies visual recognition memory. *Neuron* 58:186–194.
39. Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC (2005) Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433:597–603.
40. Ji D, Wilson MA (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat Neurosci* 10:100–107.
41. Lee AK, Manns ID, Sakmann B, Brecht M (2006) Whole-cell recordings in freely moving rats. *Neuron* 51:399–407.
42. Margrie TW, Brecht M, Sakmann B (2002) In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflugers Arch* 444:491–498.
43. Van Hooser SD, Nelson SB (2006) The squirrel as a rodent model of the human visual system. *Vis Neurosci* 23:765–778.

Supporting Information

Zoccolan et al. 10.1073/pnas.0811583106

SI Methods

Timing of Stimulus Presentation. Rats initiated behavioral trials and reported the stimulus (object) identity as described in *Materials and Methods* (see Behavioral Rig and Task). Once presentation of a visual stimulus was prompted by the animal licking the central touch sensor, its duration depended on the animal's response. The default presentation time (in the event that the animal made no response after initiating a trial) was 3 s. However, if the animal responded correctly before these 3 s expired, the stimulus remained on the monitor for an additional 4 s from the time of the response (e.g., if the animal responded correctly after 500 ms from the stimulus onset, the stimulus was displayed on the monitor for a total of 4.5 s). In the event of an incorrect response, the stimulus was removed immediately and the time-out sequence started. If the animal did not make any response during the default presentation time of 3 s, it still had 1 s, after the offset of the stimulus presentation and before the end of the trial, to make a response.

To prevent rats from making very quick (presumably random) responses, a trial was aborted if the animal's reaction time was lower than 350 ms. In such a case, the animal's response was not evaluated (neither reward or time-out was administered), the stimulus was immediately turned off, and a brief tone was played.

Pseudorandom Stimulus Presentation. In each trial, each of the 2 target objects (shown in Fig. 1*A*) had a 50% chance to be randomly selected for presentation, but the same object (e.g., object 1) was allowed to be presented in a sequence of no more than n consecutive trials (n was set equal to 3 or 4, depending on the training session and the animal). Therefore, every time such a sequence occurred by chance, the other object (e.g., object 2) was forced to be presented in the next (the $n + 1$) trial. This pseudorandom presentation strategy was adopted to prevent the rats from developing a bias for a particular reward port based on the occurrence of a long sequence of consecutive trials with objects having the same identity. All results presented in this study were appreciably the same even if these "predictable" trials were removed (see below *Data Analysis*).

Data Analysis. Our behavioral rig allowed the collection of hundreds of behavioral trials per session (between 200 and 600). As a result, over the course of 5–20 sessions, we were able to collect the response to 50–90 presentations of each object appearance that was tested during phases III and IV of our study. This allowed assessing the significance of the recognition performance of each individual subject for any tested object condition, without the need to pool across animals and/or conditions (1-tailed Binomial test, under the null hypothesis that each animal response is a Bernoulli trial with 0.5 probability of being correct; see Fig. 2*B*, *Right*). To provide a more compact description of the data, we also assessed the significance of the animals' group mean performance over individual (Figs. 2*B*, *left*, and 4*C* and *D*) or pooled (Fig. 3) object conditions (1-tailed t test).

As explained above, target objects were presented in a pseudorandom order. As a consequence, in a fraction of trials, the identity of the presented objects could theoretically be predicted from the number of previous occurrences of consecutive trials with the same object identity. Although this fraction of predictable trials was small (12%), and although it seems very unlikely that rats could exploit them to boost their performance (they would need to constantly count the number of consecutive trials with the same object that happen during a session), we verified that the performances obtained by taking into account only the predictable trials were not significantly different from the performances obtained after removal of the predictable trials (combined χ^2 test; $P > 0.05$). Therefore, the rats' performances over the tested object transformation could be safely computed by taking into account all of the collected trials (as was done in Figs. 2–4).

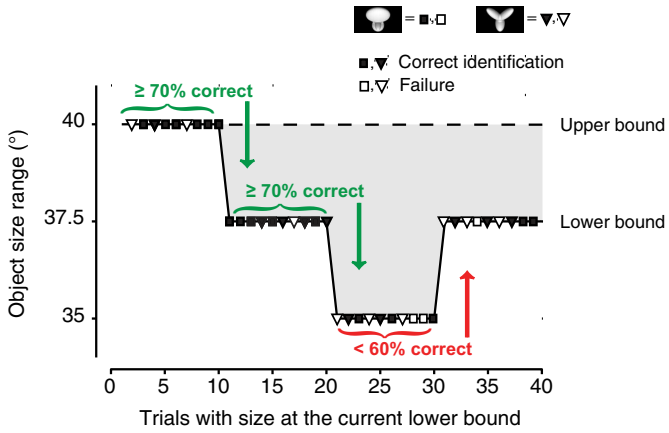

Fig. S1. Illustration of the staircase procedure used to update the range of object sizes that were presented to a subject during a training session in phase II of our study. At any given time during the session, sizes were sampled from a range (gray area) defined by a fixed upper bound of 40° (dashed line; this is the default object size used during phase I of the study) and a lower bound (solid line and symbols) that was determined by the staircase according to the animal's performance. Symbols (squares and triangles) show the identity of the object presented in a given trial (see key in the top of the figure) and the animal's response (filled symbols mean correct identification, while empty symbols mean failure). Note that, for clarity, only trials in which the target objects were presented at the current lower bound of the size range are shown in the figure (e.g., when the sizes' lower bound was 35°, objects were presented with size of 40°, 37.5°, and 35°, but only trials in which the object size was 35° are shown here). The arrows show how the sizes' lower bound was updated (in steps of 2.5°) according to the animal performance over the last 10 presentations of objects with size at the lower bound. The lower bound was decreased if the animal performance was equal or higher than 70% correct (green arrows), while was increased if the performance was equal or lower than 50% correct (red arrow). A similar staircase procedures was used to train the rats to tolerate variation along the other dimension tested in this study (i.e., in-depth azimuth rotation).

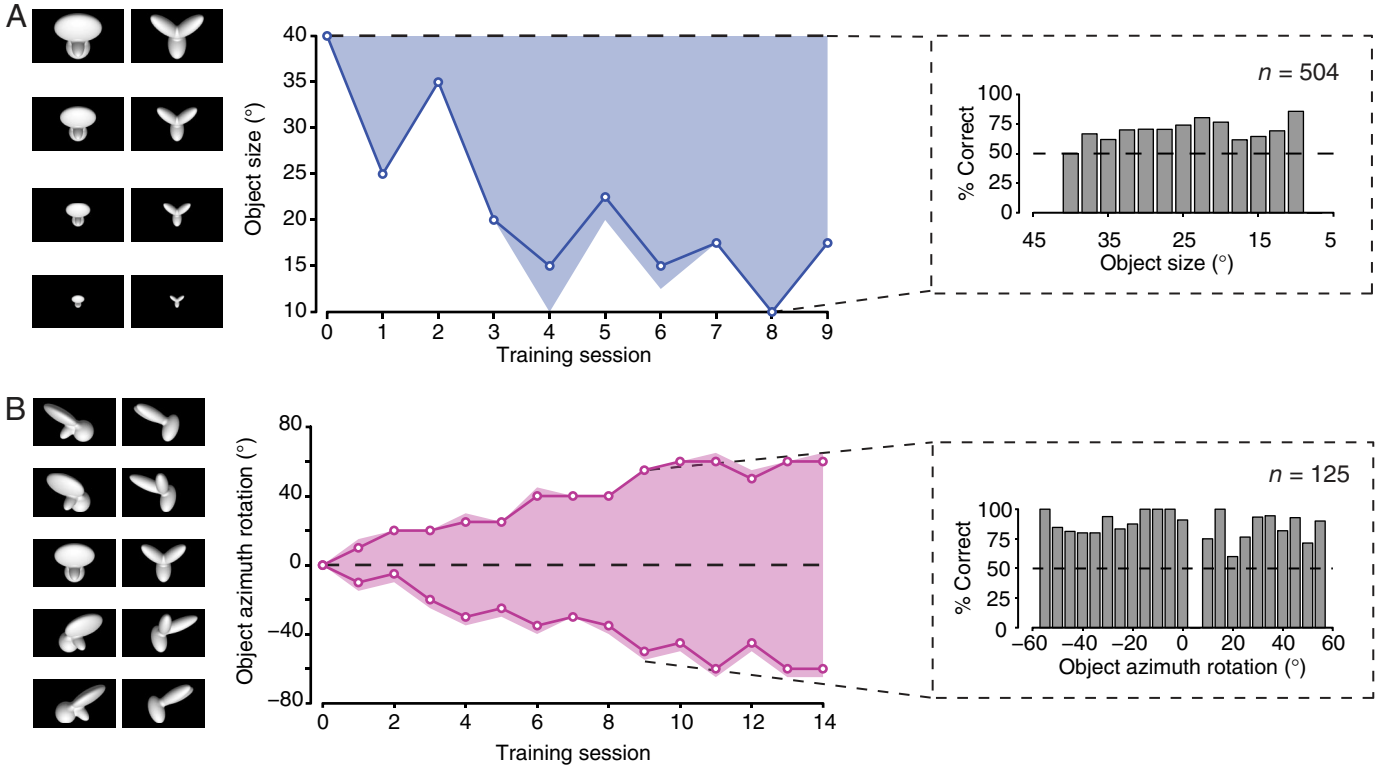

Fig. S2. Increasingly larger variation in objects' size and azimuth rotation that one of our subjects learned to tolerate during phase II of our study. Each plot on the left shows the range of image variation (colored areas) produced by one of the object transformations to which the animal was exposed across consecutive, staircase-controlled (see Fig. S1 and *Materials and Methods*) training sessions of the recognition task. The solid lines (and circles) show the minimal/maximal amount of image variation of the target objects over which the rat was able to maintain a performance equal or higher than 70% correct for at least 10 object presentations. That is, a solid line represents the minimal (maximal) lower (upper) bound reached by a staircase over the course (typically at the end) of a training session. The figure shows how, over the course of 10–15 days of training, the rat learned to recognize the target objects across sizes ranging from 40° to 10° (*A*) and azimuth rotations spanning ±60° (*B*). The dashed lines show the default values of the object properties (i.e., size and azimuth rotation) used during phase I. *Right Insets* show the rat's performance over the range of transformations that were tested during one of the training sessions (*n* is the total number of trials presented during the session).

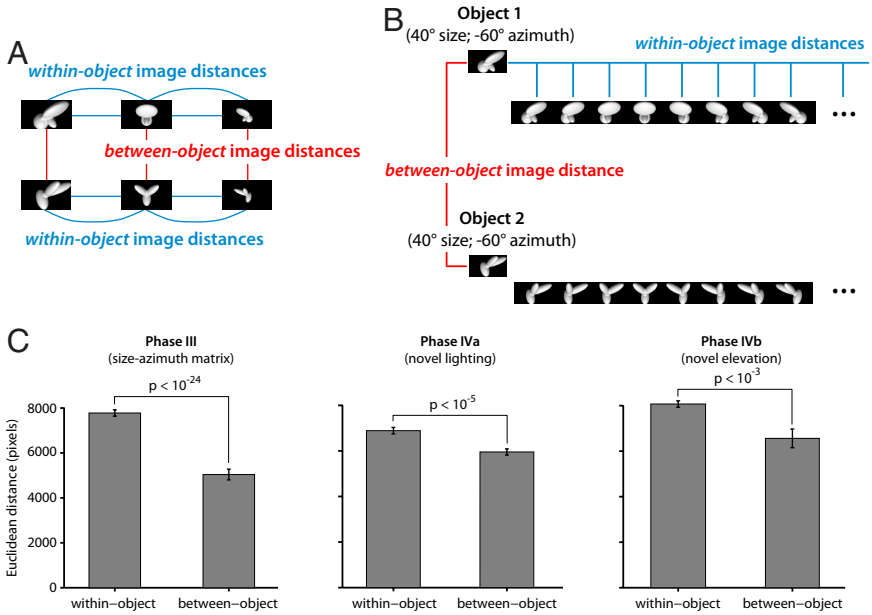

Fig. S3. Comparison of the within-object with the between-object image differences, for the object conditions used in phases III, IVa, and IVb of our study. (*A*) This conceptual diagram helps understanding the rationale of our analysis. Our goal was to measure how much image variation was produced by either changing the appearance (e.g., size and azimuth rotation) of a given object (blue lines), or, instead, the identity of the object, while maintaining size and azimuth fixed (red lines). (*B*) This operational diagram shows how the within-object and the between-object image differences were computed for the object conditions used in phase III. Given an object (object 1, in the example) in a particular appearance (i.e., a size-azimuth conjunction; 40° size and -60° azimuth, in the example), we computed the following metrics: (*i*) the within-object image distance, i.e., the average of the pixel-wise Euclidean distances between this object appearance (image) and all other appearances of the same object that were presented to the subjects during phase III (blue lines); and (*ii*) the between-object image distance, i.e., the pixel-wise Euclidean distance between this object appearance and the appearance of the other object (object 2, in the example), when presented at the same size and azimuth (red line). Both metrics were computed for every object appearance used in phase III, so to obtain 2 sets of values that could be compared pair-wise. A similar procedure was used to compute the within-object and the between-object image differences for the novel lighting conditions used in phase IVa and the novel elevation conditions used in phase IVb. (*C*) The histograms show that, for the sets of object conditions used in phases III (*left*), IVa (*center*), and IVb (*right*), the average within-object distance is larger than the between-object image distance. For each set, this difference was highly significant according to a 2-tailed, paired *t* test.

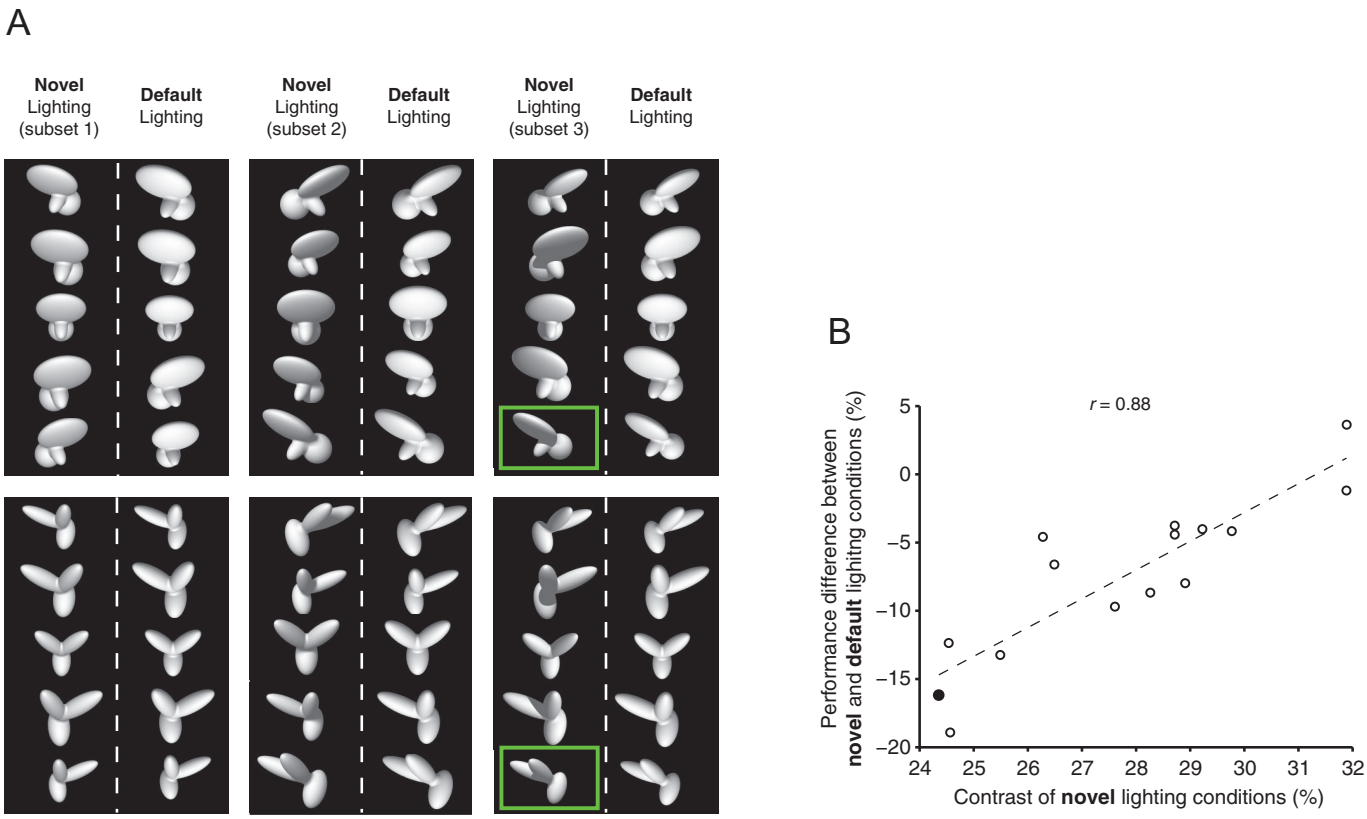

Fig. S4. The novel lighting conditions used during phase IVa of our study. (*A*) The full set of 15 arbitrary size-azimuth conjunctions of the target objects that, during phase IVa of our study, were presented to the rats both under novel lighting conditions and under default lighting conditions (i.e., the same used during phases I–III). As explained in *Materials and Methods*, these 15 novel lighting conditions were divided in 3 subsets of 5 (as shown in the figure), and each subset was presented, interleaved with the default lighting conditions from the previous phase, for 5–10 sessions. Note the large pixel-level image variation produced by the lighting manipulation and how the novel lighting condition images were overall substantially darker and lower contrast than their default lighting counterparts (the lowest-contrast condition for the 2 objects is indicated by the green frames). (*B*) The difference between performance over the novel lighting conditions and performance over their default counterparts (ordinate) is plotted against the contrast of the novel lighting conditions (abscissa). The contrast of an image was quantified by the ratio between the standard deviation of its pixel intensity values and the maximum of the pixel intensity scale (this ratio was multiplied by 100 to obtain a percentage value). For each novel lighting condition, the contrast of the resulting images of objects 1 and 2 was computed and then averaged (this is the value reported on the abscissa of the scatter plot). The performance difference and the contrast of the novel lighting conditions were strongly correlated ($r = 0.88$, $P < 10^{-4}$, 2-tailed t test). The black circle refers to the lowest-contrast novel lighting condition (see green frames in *A*), that is the only novel lighting condition for which the animals' performance was not significantly above change (corresponding to the black circle in Fig. 4*C*).

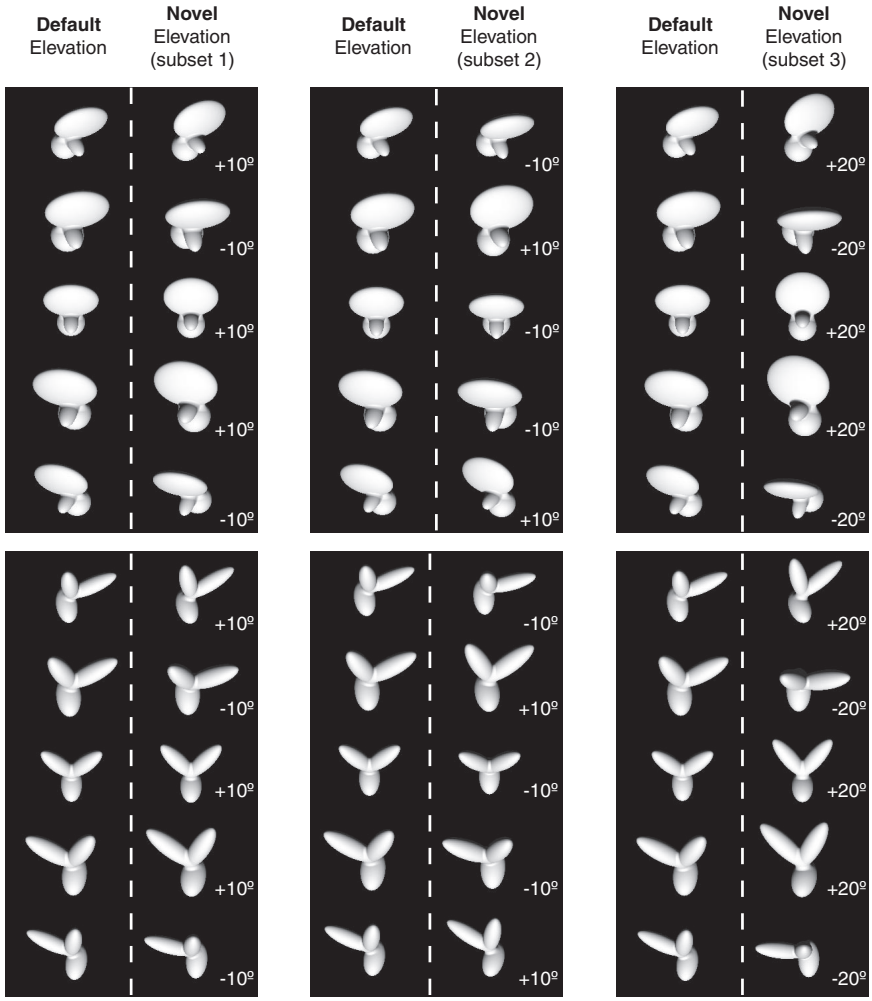

Fig. S5. The novel elevation conditions used during phase IVb of our study. As explained in *Materials and Methods*, 15 novel appearances of each target object were built by varying its elevation (i.e., the in-depth rotation about its horizontal axis) by either ±10° or ±20° and then combining this new in-depth rotation with 15 arbitrary size-azimuth conjunctions from phase III. All these new elevation conditions for both objects are shown in the figure, together with their default elevation counterparts. The 15 novel elevation conditions were divided in 3 subsets of 5 (as shown in the figure), and each subset was presented, interleaved with 45 default elevation conditions from phase III (i.e., the full matrix of size-azimuth conjunctions shown in Fig. 2*A*, with the exception of the 15° size), for 5–10 sessions. Note the substantial variation in the objects' silhouette produced by manipulating the objects' elevation.

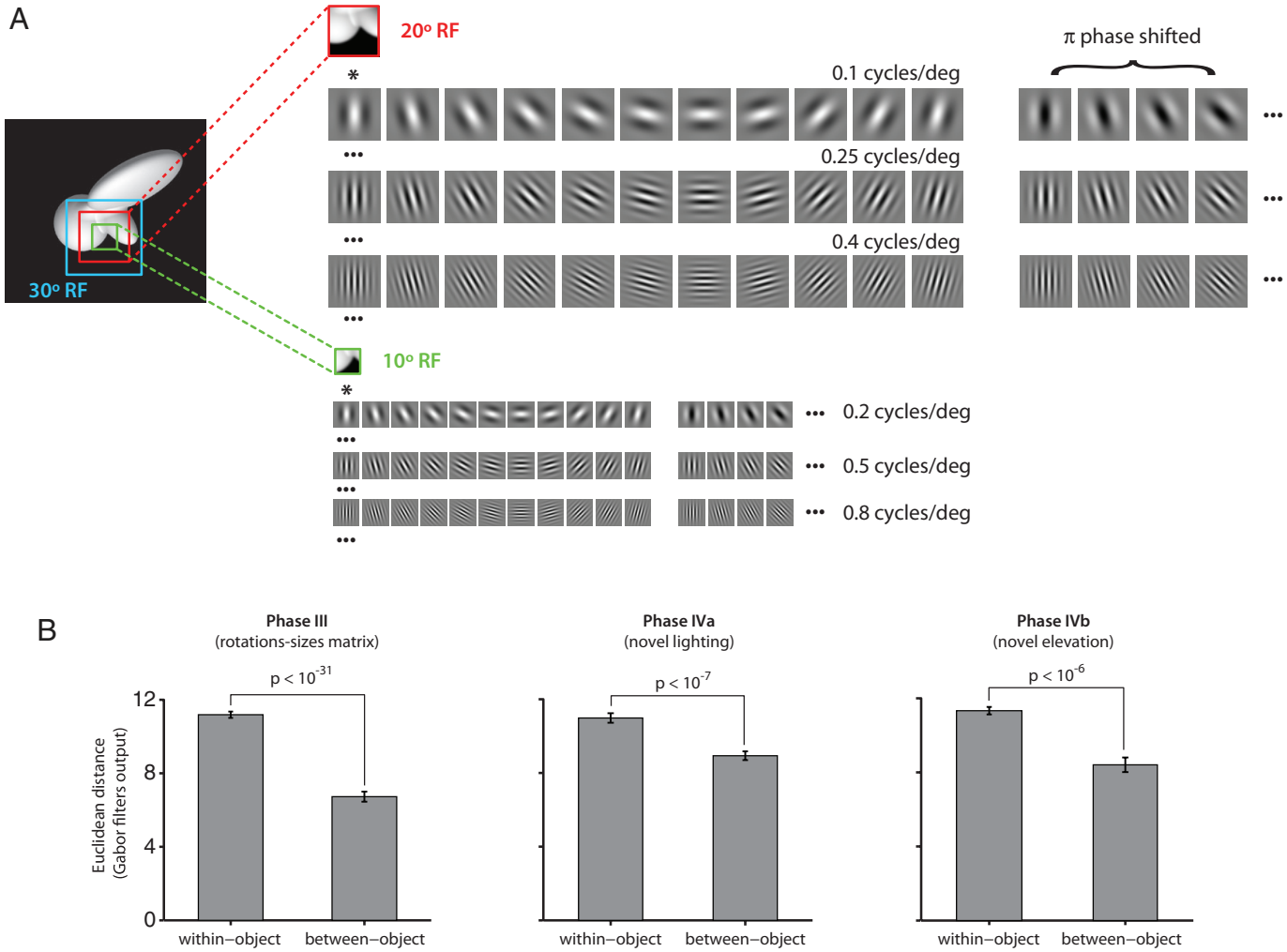

Fig. S6. Within- and between-object image differences computed over the responses of a population of simulated V1-like simple cells to our image set. (*A*) This diagram shows how the simulated V1 population was constructed. The V1 simple cells were simulated using a bank of Gabor filters with orientations, spatial frequencies, and receptive field (RF) sizes matching those reported for rat primary visual cortex [Girman et al. (1999) *J Neurophysiol* 82:301–311] and RF centers tiling the visual field. More precisely, we built an array of Gabor filters resulting from all possible combinations of: 3 RF sizes (10°, 20°, and 30°; shown, respectively, in green, red, and cyan in *Left*); 11 orientations (evenly spaced around the clock; see *Right*); 2 phases (0 and π; shown, respectively, as the left and right sets of filters in *Right*); 10 spatial frequencies, ranging from 1 to 10 cycles per RF size and resulting in a 0.03–1 cycles per degree range (see examples in *Right*). This array of Gabor filters was replicated every 5° in both the vertical and horizontal direction over the 60° × 40° span of visual field occupied by our image stimuli (for sake of simplicity, only filters at one particular visual filed location are shown in the figure). The response of a Gabor filter to a given image was computed as the dot product of the filter and the image patch with the same visual field location and size. To simulate the nonlinear response properties of V1 simple cells (i.e., saturation, luminance and contrast normalization, and non-negative firing rates), both the filter and the image patch were normalized to 1 before computing their dot product and negative responses were clipped to 0. (*B*) For each image in our stimulus set, we computed its representation in the space of the simulated V1 population, and we obtained the within-object and the between-object image differences in this space, using the same rationale described in Fig. S3. As shown by the histograms, for the sets of object conditions used in phases III (*Left*), IVa (*Center*), and IVb (*Right*), the average within-object distance was larger than the between-object image distance, and this difference was highly significant according to a 2-tailed, paired *t* test.