Neural Substrates of View-Invariant Object Recognition Developed without Experiencing Rotations of the Objects

Jun-ya Okamura,¹ Reona Yamaguchi,¹ Kazunari Honda,¹ Gang Wang,^{1,2} and Keiji Tanaka²

¹Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University, Kagoshima 890-0065, Japan and ²Cognitive Brain Mapping Laboratory, RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan

One fails to recognize an unfamiliar object across changes in viewing angle when it must be discriminated from similar distractor objects. View-invariant recognition gradually develops as the viewer repeatedly sees the objects in rotation. It is assumed that different views of each object are associated with one another while their successive appearance is experienced in rotation. However, natural experience of objects also contains ample opportunities to discriminate among objects at each of the multiple viewing angles. Our previous behavioral experiments showed that after experiencing a new set of object stimuli during a task that required only discrimination at each of four viewing angles at 30° intervals, monkeys could recognize the objects across changes in viewing angle up to 60°. By recording activities of neurons from the inferotemporal cortex after various types of preparatory experience, we here found a possible neural substrate for the monkeys' performance. For object sets that the monkeys had experienced during the task that required only discrimination at each of four viewing angles, many inferotemporal neurons showed object selectivity covering multiple views. The degree of view generalization found for these object sets was similar to that found for stimulus sets with which the monkeys had been trained to conduct view-invariant recognition. These results suggest that the experience of discriminating new objects in each of several viewing angles develops the partially view-generalized object selectivity distributed over many neurons in the inferotemporal cortex, which in turn bases the monkeys' emergent capability to discriminate the objects across changes in viewing angle.

Key words: monkey inferotemporal cortex; object; view invariance

Introduction

An unfamiliar object cannot be distinguished from similar distractors when the viewing angle changes (Rock and DiVita, 1987; Bülthoff and Edelman, 1992; Humphrey and Khan, 1992; Logothetis et al., 1994; Tarr, 1995). The capability to recognize objects across changes in viewing angle gradually develops as the viewer repeatedly sees the object and distractors in rotation. It has been proposed that different views of each object become associated with one another when they are experienced in succession during rotation (Földiák, 1991; Wiskott and Sejnowski, 2002; Wyss et al., 2006; Masquelier and Thorpe, 2007). It has been shown that human subjects, by successively seeing nearby views of different people's faces, formed some degree of false view invariance (Wallis and Bülthoff, 2001; Cox et al., 2005).

However, natural experience of objects also contains ample opportunities to discriminate among objects at each of multiple viewing angles. This component of natural experience has been largely neglected. Single-cell recordings in macaque monkeys have shown that the selectivity of neurons in the inferotemporal cortex, which is a high-level stage in the visual pathway for object vision (Tanaka, 1996), changes as the monkey repeatedly experiences and discriminates among stimulus images (Sakai and Miyashita, 1994; Logothetis et al., 1995; Kobatake et al., 1998; Baker et al., 2002; Sigala and Logothetis, 2002). The experience of discriminating among objects at each of a small number of viewing angles may have impact on the development of view-invariant recognition capability for the objects.

For faces, clusters of cells with consistent selectivity for personal identity over a wide range of viewing angles have been found in the anterior part of the macaque inferotemporal cortex (Freiwald and Tsao, 2010). For objects other than faces, however, the proportion of such cells with nearly perfect view invariance was low (Tanaka et al., 1991; Booth and Rolls, 1998). Instead, many inferotemporal cells showed moderately broad tunings for viewing angle (Logothetis et al., 1995). This partial view invariance distributed over many inferotemporal cells might underlie view-invariant recognition by animals.

By exposing macaque monkeys to a new set of object stimuli in a task that required recognition of the objects for each of four viewing angles at 30° intervals, we had previously found that, immediately after the preparatory experience, the monkeys rec-

Received May 12, 2014; revised Sept. 24, 2014; accepted Sept. 29, 2014.

Author contributions: G.W. and K.T. designed research; J.-y.O., R.Y., K.H., and G.W. performed research; J.-y.O., R.Y., and G.W. analyzed data; G.W. and K.T. wrote the paper.

This work was supported by the Grant-in-Aid for Scientific Research (23500521) from the Japanese Ministry of Education, Sports, Science, and Technology to G.W. and by the Japan Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) to K.T.

The authors declare no competing financial interests.

This article is freely available online through the J. Neurosci. Author Open Choice option.

Correspondence should be addressed to Dr. Gang Wang, Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University, Kagoshima 890-0065, Japan. E-mail: gwang@ibe.kagoshima-u.ac.jp.

DOI:10.1523/JNEUROSCI.1898-14.2014

Copyright © 2014 the authors 0270-6474/14/3415047-13\$15.00/0

ognized the objects across changes in viewing angle of up to 60° without further training (Wang et al., 2005; Yamashita et al., 2010). To determine neural substrates of the monkeys' performance, we here recorded activities of neurons from the inferotemporal cortex after various types of preparatory experience. We found that inferotemporal cells showed responses with object selectivities covering multiple viewing angles for stimulus sets that the monkeys had experienced only during discrimination at each of four viewing angles. These behavioral and cellular results suggest that the experience of discriminating images of new objects in each of several viewing angles develops partially view-generalized object selectivity distributed over neurons in the inferotemporal cortex, which in turn bases the monkeys' emergent capability to discriminate the objects across changes in viewing angle.

Materials and Methods

We used three male macaque monkeys (*Macaca fuscata*). One of them participated in both the purely behavioral and cell-recording experiments (Monkey 1), a second participated only in the purely behavioral experiments (Monkey 2), and the third participated only in the cell-recording experiments (Monkey 3). All procedures were performed in accordance with the guidelines of the Japan Neuroscience Society and were approved by the Animal Experiment Committee of Kagoshima University. In a preparatory surgery, which was performed under aseptic conditions with sodium pentobarbital anesthesia (35 mg/kg, i.p.; supplemented by 10 mg/kg if necessary), a titanium head holder was implanted on the skull using titanium screws. To perform tasks, the monkeys sat in a chair, with their head position fixed by the titanium holder, facing a CRT display at a distance of 50 cm. A lever was placed in front of the monkey's body. Eye position was measured by an infrared system (<http://staff.aist.go.jp/k.matsuda/eye/>).

Object images used as stimuli. Stimulus objects were created using 3D graphic software (Shade 9; e-frontier). The details have been described previously (Wang et al., 2005). Briefly, we created four artificial objects in a set by deforming a prototype in four different directions in a 3D feature space. Six or seven parameters of the object shape were combined into three parameters that spanned the feature space. To create the four different views, each object was rotated with a 30° interval around an axis perpendicular to the visual axis that connected the viewer's eyes and the object. The 16 images (4 views $\times$ 4 objects) formed a stimulus set. Starting from different prototypes, we created different stimulus sets. Figure 1*a* shows the six prototypes from which the six sets used in the cell-recording experiments of the present study were created, and Figure 1*b* shows the whole set of stimulus images in one set. The differences in the object shapes between the different sets were much larger than the differences within each set. The sizes of object images along the longest axis were 6.5° on average (range was from 4.9 to 8.0°). They were presented to the fovea: their centers were located at the position of the gaze fixation spot.

We used human psychophysics to adjust both the direction of the deformations from the prototype and the relative amount of deformation in the four directions so that the discrimination between the different pairs of daughter objects was equally difficult within a set. We also used human psychophysics to make the difficulty of discrimination within each set comparable among different stimulus sets, with the percentage of correct responses $\sim$ 80% (Fig. 1*c*). Based on previous experiments that tested monkeys with stimulus sets prepared using human psychophysics, we adjusted the presentation condition in the human experiments so that the percentage of correct responses approximately corresponded to those of the monkeys during one of the preparatory tasks (Within-set Image task, see below). Therefore, in the Within-set Image task, we expected that the monkeys would be able to discriminate among the daughter objects at $\sim$ 80% correct.

During the object creation procedure, we calculated the dissimilarity of images in the domain of primitive features and modified the object shapes so that different views of different objects were not more different from the different views of the same object within each set. We calculated

the dissimilarity between two images using the Euclidean distance between the coefficients of their 2D wavelet transformations. We used a discrete type of wavelet transformation (MATLAB Daubechies Daub4 wavelet functions) to analyze the frequency components of the images. With the coefficients given by the Daub4 wavelet functions, images (the signals) can be reconstructed with relatively small errors from the transformed frequency components (Daubechies, 1990). In this calculation, one of the two images was translated and rotated relative to the other and the minimum value of dissimilarity was taken as the final value of dissimilarity. When the dissimilarity between within-objects pairs and across-objects pairs was compared for the same viewing angle difference, there was a good overlap between the two distributions for viewing angle differences of 30, 60, and 90° (Fig. 1*d*). There were no significant differences for these dissimilarity distances ($p > 0.05$, by Mann–Whitney U test). The same results were obtained when the similarity between the two images was calculated as the Euclidean distance between the luminosity values in individual pixels after adjustment of position and orientation.

Object task. The monkeys' ability to recognize objects across viewing angle changes was measured using a task in which they had to detect a change in the identity of an object while neglecting changes in the viewing angle without object change (Object task). The monkey initiated a trial by pressing a lever, which made a fixation spot appear at the center of the screen (Fig. 2*a*). The monkey had to fixate its eyes on the fixation point within 1.5 s. After continuously pressing the lever and fixating for 500 ms, the first stimulus appeared. Two to five stimuli were presented on each trial. Each stimulus was presented for 500 ms with 500 ms interstimulus intervals. With an exception of the last stimulus, all were different views of the first object (Fig. 2*b*, top). The monkey had to keep the lever pressed during the presentation of the different views of the first object and then release the lever within 1 s when a view of a second object appeared. Correct responses were rewarded with a drop of water. The second object was always selected from the same set to which the first object belonged. This task examined the monkeys' ability to recognize objects across viewing angle changes, as they had to distinguish between image changes resulting from changes in the viewing angle of an object and those resulting from changes in both the viewing angle and the identity of the object. The intertrial interval was 1.5 s after correct responses and 2.5 s after error responses. The monkey had to maintain eye fixation with an accuracy of $\pm 2.5^\circ$ until the last stimulus appeared. The probability of an object being changed was one-third for each of the second, third, and fourth presentations. If the trial advanced to the fifth presentation, a view of a second object was always presented. During the tests, the change in the viewing angle during successive stimulus presentations within a trial was 0°, 30°, 60°, or 90° in the ordinary Object task. In the first-experience test that was used to measure view-invariant object recognition ability immediately after the preparation experience, the angle change was fixed at 30°. With the exception of the above constraints, all stimuli were randomly selected from the set.

Preparatory tasks for behavioral experiments. Stimuli in a new set were presented in one of two types of preparatory tasks before testing the monkey's view-invariant recognition performance in the Object task. The time sequence and structure of the preparatory tasks were essentially the same as those used for the Object task. However, an identical view of the first object was repeated one to four times before a view of a second object appeared. Therefore, the task did not require view-invariant object recognition. In one of the two preparatory tasks (Within-set Image task), the second object was one of the other three objects in the same set (Fig. 2*b*, middle). The viewing angle of the second object image was identical to that of the first object image. Therefore, the monkeys discriminated one object from the others within the same viewing angle. Different views of the same object never appeared within a single trial, but appeared in different trials. Therefore, there was no chance for the monkey to associate different views of the same object in this preparatory task. The other preparatory task (Fig. 2*b*, bottom, Across-set Image task) was the same as the Within-set Image task, except that the second stimulus was an image of an object belonging to a different object set. Therefore, no discrimination of the small differences among objects within a set was required. In both preparatory tasks, the first object was randomly selected from the set.

Figure 1. Object image sets used in the experiments. **a**, Examples of prototype objects. **b**, An example of a stimulus set. **c**, Results of human psychophysics for the six stimulus sets used for cell-recording experiments. The task used in the human psychophysics experiments was a delayed matching to object. In each trial, a view of an object was shown as a sample for 200 ms, followed by a delay of 750 ms, during which a mask pattern was presented. After the delay, another view of the sample object was presented for 500 ms together with a view of another object in the set, followed by a mask for 500 ms. The subject had to determine which of the two test stimuli was a view of the sample object. Five subjects participated in the experiments. The percentages of correct responses are shown separately for 30, 60, and 90° differences in viewing angle between the sample and test views of the object. The error bars indicate SEM across subjects. The results show that the difficulty of object recognition across changes in viewing angle against other objects within the set was comparable among the six sets. All the stimulus sets used in the behavioral experiments were also adjusted in the human psychophysical experiments on four to six subjects to obtain ~80% correct responses. **d**, Dissimilarity between images for within-object and across-object pairs. The Euclidean distance between coefficients of wavelet transformations of the images was calculated (after the translation and rotation that

Each new stimulus set was presented to the monkey for at least 20 d using a preparation task before testing view-invariant recognition ability on the set with the Object task. On each day, the new set was presented for 200 trials in the preparation task, while another 200 trials were performed with the Object task using a fixed familiar object set to maintain the monkeys' familiarity with the Object task. Trials of the preparation task were randomly intermingled with those of the Object task. Although the Across-set Image task was very easy, we expected that the monkeys would maintain general attention in trials of the Across-set Image task as the trials of the Across-set Image task were randomly intermingled with those of the Object task.

Analyses of behavioral performance. We focused on the monkeys' performance on the second stimulus presentation in each trial, as we were able to uniquely define the change in the viewing angle from the previous to the current object images in this case. We calculated the proportion of trials with correct releases at the second stimulus presentation among the trials in which the object identity changed from the first to the second stimulus presentation (hit rate). We then compared it with the proportion of the trials with incorrect, or false, releases at the second stimulus presentation among the trials in which the object did not change from the first to the second stimulus presentation (false alarm rate) using a χ^2 test. Note that the hit and false alarm rates are independent measures for different types of trials. For the Object task, hit and false alarm rates were calculated separately for trials in which the viewing angle changed by 0, 30, 60, and 90° from the first to the second stimulus presentation. When the hit rate was significantly larger than the false alarm rate in trials with nonzero changes in viewing angle in the Object task, this indicated that the monkey was able to discriminate object-identity changes from pure view changes of an object. No difference between the two rates indicates that the monkey did not make this discrimination. We did not truncate the trial after the second stimulus presentation to keep the task more difficult and to reduce the chance of obtaining a reward by responding randomly.

Preparatory experience before single-cell recordings. Each of the two monkeys experienced six sets of object images in three tasks: first two sets in the Object task, another two sets in the Within-set Image task, and the remaining two sets in the Across-set Image task. The combinations of set and task were changed between the two monkeys so that discriminability differences between sets, which might have remained after the adjustments, should not have influenced the final results. Sets A and B were combined with the Across-set Image task, Sets C and D with the Within-set Image task, and Sets E and F with the Object task in Monkey 1, whereas Sets C and F were combined with the Across-set Image task, Sets A and E with the Within-set Image task, and Sets B and D with the Object task in Monkey 3. The six sets were first introduced one by one in the combined task, while the monkey performed the Object task with a fixed familiar object set. Trials with stimuli in the new set were randomly intermingled with those with the Object task and stimuli in the fixed set. In the Within-set Image task and in the Object task, the hit rate became larger than the false alarm with a difference between the two of >0.5 within 20 d (Monkey 1) or 10 d (Monkey 3) after the introduction. In the Across-set Image task, the performance was close to perfect (with hit rate of ~1.0 and false alarm rate of ~0.0) from the beginning. After the monkey had learned all six sets, they were presented together each day for a few more months, with the same set-task combinations as before. During this period, the frequency of trials from each of the six sets was adjusted, so that the total number of times each image was presented across all six sets became equal. Trials with different set-task combinations were intermingled.

We started single-cell recordings from the inferotemporal cortex (Fig. 3) after the preparation experience was completed. During cell recordings, images from all sets were presented within the Across-set Image

← realized the minimum distance), normalized by the maximum distance in each set, and plotted against the difference in viewing angle between the two images. Broken lines represent pairs within each object and solid lines represent pairs across objects. The set names correspond to those noted for the prototypes shown in **a**. Error bars represent SD across pairs.

task. There were no cell recordings for 1 d each week, and on these days the monkeys experienced the images in the three tasks with the same set-task combinations as in the initial preparatory period. The monkeys maintained a good performance on these no-recording days: the overall mean difference between hit and false alarm rates was 0.97 (ranging from 0.94 to 1.00) in the Across-set Image task, 0.74 (0.64–0.78) in the Within-set Image task, and 0.83 (0.77–0.88) for no-rotation trials, 0.84 (0.76–0.89) for 30° rotation trials, 0.81 (0.77–0.89) for 60° rotation trials, and 0.61 (0.34–0.84) for 90° rotation trials in the Object task.

Single-cell recordings. Before the start of cell recordings, a recording chamber was implanted in the skull in an aseptic surgery similar to the first preparatory surgery. Recordings were conducted with tungsten electrodes (FHC), which passed through a guide tube, and were advanced by a micromanipulator (Narishige). The positions of recorded cells were determined with reference to MRIs taken before the first preparatory surgery. Cells were recorded from the ventrolateral region of inferotemporal cortex, lateral to the anterior middle temporal sulcus, in the posterior/anterior range between 18 and 26 mm anterior to the ear bar position for Monkey 1 and between 16 and 19 mm for Monkey 3 (Fig. 3). Action potentials of single cells were recorded and isolated using Cheetah Data Acquisition System (Neuralynx). All cell recordings were conducted while the monkey was performing the Across-set Image task.

Once action potentials from a single neuron were isolated, the cell's responses were first examined with 24 images of six objects. The six objects were selected from all of the six object sets, and were changed every recording day. The most effective image was determined by examining response histograms, and then the object set, which included the image that was used for further examination of the cell's responses. When the recorded cell did not show clear responses to any of the 24 images, we moved the electrode to another neuron.

Analyses of neuronal data. We focused on neuronal responses to the first stimulus presentation in each trial to avoid the effects of preceding stimuli and of the monkey's decision. Only responses for correct trials were included. The magnitude of the responses was determined as mean firing rate during a 500 ms time window starting 60 ms after the stimulus onset minus the spontaneous firing rate during the 500 ms period immediately before the stimulus onset.

The statistical tests were all conducted by nonparametric methods. This was because we found that for any of the statistical tests there was a group of data or a pair of data groups that violate the assumptions of parametric tests (the normality of distribution for *t* test and ANOVA and the equal variances for ANOVA). The normality of distribution was tested by applying Kolmogorov–Smirnov test to the difference between the distribution of measured data and the normal distribution of equivalent mean and SD, and the assumption of equal variances by Levene's test.

Results

Altogether we used 22 sets of object images (16 sets for the purely behavioral experiments and 6 sets for the cell-recording experiments), each of which was composed of 16 object images, four views each (at 30° intervals) of four similar objects. The differences between sets were much larger than the differences within a set (compare Fig. 1*a, b*), which were adjusted, using human psychophysics, to be comparable across sets (Fig. 1*c*). The differences between images, measured by the Euclidean distance between coefficients of their wavelet transformations, were made comparable between within-object and across-object pairs of images with the same viewing angle difference of 30° or larger (Fig. 1*d*). The last adjustment was made to assure that monkeys could not achieve view-invariant object recognition within a set based on image similarity in lower level feature space.

Three macaque monkeys were trained on a task that required view-invariant object recognition (hereafter referred to as the Object task; Fig. 2*a, b*, top). In each trial, images of objects selected from a stimulus set were presented sequentially while the monkey held a lever pressed and maintained gaze fixation at a point on the

Figure 2. Behavioral tasks. *a*, The basic time sequence of events was common to all three tasks used in the present study. While the monkey kept a lever pressed and maintained fixation on a point, two to five object images were presented sequentially in each trial: one to four presentations of a first object were followed by one view of a second object. The monkey had to release the lever when the second object appeared. *b*, The viewing angle of object images changed from presentation to presentation in the Object task, but not in the Within-set and Across-set Image tasks. The second object was selected from the same set as that of the first object in the Object and Within-set Image tasks, whereas an object belonging to a separate set was used as the second object in the Across-set Image task.

screen. Different views of one object were presented one to four times, after which a view of a second object appeared. The monkey had to release the lever when the second object appeared. The monkeys learned this Object task using multiple sets of object images different from those used in the main experiments described in this paper.

We used two other tasks for preparatory exposure of stimulus sets to the monkeys before testing using the Object task. Their time courses were similar to that of the Object task, and the monkey's task was also similar in that it had to release the lever when the stimulus object changed. In a trial from one preparatory task (Within-set Image task), an identical image was repeated one to four times followed by the image of a second object at the same viewing angle (Fig. 2*b*, middle). Thus, there was no view change within each trial. However, all the different views of the objects were covered by different trials. The first and second objects belonged to the same set in the Within-set Image task, as in the Object task. The other task (Across-set Image task) was similar to the Within-set Image task, but the second stimulus was one of four views of an object belonging to a different stimulus set (Fig. 2*b*, bottom). Thus, discrimination of small differences in object shapes within a set was necessary only in the Within-set Image

Figure 3. Recording sites. The recording sites are indicated by shading on the lateral views of the hemispheres (top and middle) and on representative frontal (coronal) sections (bottom), which were reconstructed from structural MRI images. The recordings were limited to the ventrolateral surface of the brain lateral to the anterior middle temporal sulcus (AMTS). STS, superior temporal sulcus. Scale bars: in the frontal sections, 10 mm.

task. Only detection of a big change was required in the Across-set Image task.

Behavioral results

We first summarize the results of purely behavioral experiments. Parts of the results had been published (Wang et al., 2005; Yamashita et al., 2010). Two of the three monkeys participated in the behavioral experiments (Monkeys 1 and 2). We let the monkeys experience a new set of object images in one of the preparation tasks for ~1 month, after which the monkeys' capability of view-invariant object recognition among the objects within the set was examined in the Object task. We focused on the monkey's behavior at the second stimulus presentation in each trial, because at this position in the sequence we can uniquely determine the change in viewing angle across which the monkey made object recognition. The decision at the third, and later, stimulus presentation could be influenced by the view difference between the first and current stimuli as well as that between the second and current stimuli.

When a new set was suddenly introduced to the Object task without any previous preparatory experience of the set, the monkeys could not perform well (Fig. 4a). The rate of hit (correct bar release at object change) was comparable to that of false alarm (erroneous bar release at pure view change; $p > 0.3$, by χ^2 test), which means a complete failure in view-invariant object recognition/discrimination.

Results for the Object task after performing one of the preparatory tasks was quite different depending on which of the two

preparatory tasks was used. When a new set was introduced for the Object task after a preparatory experience of the set in the Within-set Image task, the hit rate was significantly higher than the false alarm rate (Fig. 4b); that was always true for 30° view changes ($p < 0.0001$, by χ^2 test) and also in many of the tests for 60° view changes (four with $p < 0.001$ and one with $p < 0.01$; $p > 0.3$ for the other three). These latter findings indicated that the experience of the set within the Within-set Image task, which did not include view-invariant recognition/discrimination by itself, was enough to develop the monkey's capability of view-invariant recognition/discrimination within the set with viewing angle changes up to 60°. On the other hand, when the preparatory experience was given in the Across-set Image task, there was no sign of view-invariant recognition/discrimination when the set was introduced to the Object task (Fig. 4c; $p > 0.5$). This last finding indicated that passive viewing was not enough for development of view-invariant recognition capability. The discrimination of small differences between the target and distractor objects, at each of the multiple viewing angles, was necessary.

Parts of the behavioral results described above were obtained in experiments with a counterbalanced design. i.e., effects of two types of preparatory experience were compared between two monkeys by using the same set, and also between two stimulus sets on the same monkey (Wang et al., 2005, their Fig. 5; Yamashita et al., 2010, their Fig. 5). Because the performance of the monkeys in the after-preparation tests clearly depended on the types of preparatory experience even with the same set or in the same monkey, the differences were not due to differences in stimulus sets or monkeys, but were due to differences in preparatory experience.

Responses of single cells in the inferotemporal cortex

To find neural bases of the monkeys' performance, we conducted cell-recording experiments. Two of the three monkeys were used in the cell-recording experiments (Monkeys 1 and 3). Monkey 1 had participated in the behavioral experiments, whereas Monkey 3 was newly trained in the Object task before the preparation for cell-recording experiments started. As preparation for single-cell recordings, each monkey experienced six sets of object images in three tasks: first two sets in the Object task, then another two sets in the Within-set Image task, and finally the remaining two sets in the Across-set Image task. We started single-cell recordings from the inferotemporal cortex (Fig. 3) after the preparation experience was completed. During cell recordings, images from all sets were presented within the Across-set Image task, to avoid possible effects of the ongoing task on neuronal responses. Once action potentials from a single neuron were isolated, the cell's responses were first examined with 24 images of six objects, selected from all of the six object sets. The most effective image was determined by examining response histograms, and then the object set that included that image was used for further examination of the cell's responses.

A total of 353 cells was recorded (140 and 213 cells in Monkeys 1 and 3, respectively). We focused on responses evoked by the first stimulus presentation in each trial, to avoid the effects of preceding stimuli and of the monkey's decision. The magnitude of response was quantified as the mean firing rate in a 500 ms window starting at 60 ms after the stimulus onset subtracted by the mean spontaneous firing rate in a 500 ms window immediately before the stimulus onset. The results described below were obtained on 179 cells in which at least 1 of the 16 images in the set evoked statistically significant responses (activities in the 500 ms window starting at 60 ms after the stimulus onset were compared

Figure 4. Monkeys' performance for view-invariant object recognition in the Object task after three different types of preparatory experience. In the main scatter plots, the performance in each experiment is plotted by a symbol with its y-position representing the hit rate (correct release to an object change) and x-position the false alarm rate (erroneous release to a pure view change). Circles indicate the results for the first experience of the pair of images, and crosses indicate those accumulated over the first 50 trials during the first day when the set was introduced to the Object task. The color of symbols indicates the monkey with which the data were obtained (red for Monkey 1 and blue for Monkey 2). The bar graph plots d' values of the individual experiments calculated from the combinations of hit and false alarm rates. The left parts show the results when the change in viewing angle was 30°, and the right parts those for 60° changes. We focused on the monkeys' behavior at the second stimulus presentation. **a**, The performance when a new set was introduced to the Object task without any prior experience. **b**, The performance after prior experience of the set in the Within-set Image task. **c**, The performance after prior experience of the set in the Across-set Image task. Some of the data points were replotted from Wang et al., 2005 and Yamashita et al., 2010.

with those in the 500 ms window immediately before the stimulus onset, $p < 0.05$ with Bonferroni correction, by Wilcoxon signed-rank test). As results of the experimental procedure described above, responses were examined with one of the two sets that the monkey had experienced in the Object task for 53 cells, with one of the two sets that the monkey had experienced in the Within-set Image task for 77 cells, and with one of the two sets that the monkey had experienced in the Across-set Image task for 49 cells. As responses of a cell were recorded only with one object set, a set of a cell and responses in it will be referred to as cell/responses.

There were cells/responses that showed larger selectivity for objects than for viewing angle (Fig. 5, Cell 1) and cells/responses with larger selectivity for the viewing angle than for objects (Fig. 5, Cell 3). Other cells/responses showed selectivity for both object and viewing angle (Fig. 5, Cell 2). Each of the three groups of

cells/responses defined by the type of preparatory task was quite heterogeneous, including all these types of cells. Therefore, we quantified a few aspects of response tunings and compared them among the three groups of cells/responses. In each cell, we first determined the most effective stimulus that evoked the largest response among the 16 stimuli in the selected set, and then normalized responses to individual stimuli by the largest response. Responses were then labeled by the viewing angle difference of each stimulus from the most effective, or preferred, stimulus and the rank order of the objects determined at the viewing angle of the most effective stimulus; i.e., 0°, 30°, 60°, or 90° for the viewing angle difference and ranks 1–4 for object rank. Responses to the images of the rank 2–4 objects were averaged over the three objects to obtain averaged responses to images of nonpreferred objects. Rank 1 object is also referred to as the preferred object. As expected from the way to make four objects in each set, i.e., the distances between any pairs of four objects were made equivalent by human psychophysics, the rank 2, 3, and 4 objects varied among the cells that maximally responded to images of an object. Therefore, the comparison between responses to the preferred objects and the averaged responses to the nonpreferred objects is more relevant in our case than the comparison between the responses to the preferred and rank 2 objects when we discuss averaged responses. When the preferred stimulus was a 30° or 60° view, responses at 30° difference were averaged between +30° and –30° from the preferred angle. Also, in this case, there was no value at 90° difference. We thus obtained, for each cell, tuning curves of responses plotted against the viewing angle relative to the preferred angle separately for the preferred and nonpreferred objects (Fig. 5, right). The tuning curves were then averaged across cells within each group of cells/responses (Fig. 6a–c). The tuning curves averaged over cells recorded from each monkey are also shown in Figure 7, and values, in individual cells, of the normalized responses to the preferred and nonpreferred objects at 30° difference, are shown in Figure 8.

The preferred object versus nonpreferred object difference was significant at 30° difference both after preparation in the Object task (Figs. 6a, 8a; $p < 0.0001$, by sign test) and after preparation in the Within-set Image task (Figs. 6b, 8b, $p < 0.0001$), but not after preparation in the Across-set Image task (Figs. 6c, 8c, $p = 0.39$). There were also significant differences at 60° difference both after experience in the Object task ($p = 0.00080$) and after experience in the Within-set Image task ($p = 0.012$), but not after experience in the Across-set Image task ($p = 0.25$). There was a significant difference at 90° difference only after experience in the Object task ($p = 0.0021$, 0.43, and 0.60 after experience in the Object, Within-set, and Across-set Image tasks, respectively). The main points of the results were common to the cells recorded from Monkey 1 (Fig. 7, left) and those recorded from Monkey 3 (Fig. 7, right). Also, similar results were obtained even when responses to rank 2 objects were compared with responses to the preferred objects. The differences at 30° difference were significant both after preparation in the Object task ($p < 0.0001$, by sign test) and after preparation in the Within-set Image task ($p = 0.0013$), but not after preparation in the Across-set Image task ($p = 0.57$), and those at 60° difference were significant after preparation in the Object task ($p = 0.013$), but not after preparation in the Within-set Image task ($p = 0.25$) and after preparation in the Across-set Image task ($p = 0.11$).

When the difference between the normalized response magnitudes to the preferred and nonpreferred objects was compared among the three groups defined by the type of preparatory experience (Fig. 6d), a significant dependence on the group was found

Figure 5. Responses of three example cells. Cells 1, 2, and 3 were examined with the object sets that the monkeys had experienced in the Object, Within-set, and Across-set Image tasks, respectively. Raster plots and peristimulus time histograms in the middle are aligned in the same matrix as that for stimuli shown in the left. The maximum responses are circumscribed by squares. The preferred object and viewing angle of a cell are defined as those by which the maximum response was evoked in the cell. In the tuning plots in the right, solid lines indicate responses to the preferred object, and the broken lines indicate those to the rank 2–4 (nonpreferred) objects. The diamonds without connecting lines indicate the averaged responses to the nonpreferred objects.

at 30° difference ($p < 0.0001$, by Kruskal–Wallis test). The *post hoc* test indicated that the differences after experience in the Object task and after experience in the Within-set Image task were significantly larger than the difference after experience in the Across-set Image task ($p < 0.0001$ for both Object vs Across-set Image and Within-set Image vs Across-set Image, by Steel–Dwass test), whereas there was no significant difference between the Object task and Within-set Image task preparation groups ($p = 0.29$). There were similar numerical tendencies at 0, 60, and 90° differences, but they were not statistically significant ($p = 0.25$ at 0° difference, $p = 0.25$ at 60° difference, and $p = 0.083$ at 90° difference, by Kruskal–Wallis test). These results suggest that the strong object selectivity at the preferred viewing angle more spread to the neighboring viewing angles after the Object task and Within-set Image task experience compared with the Across-set Image task experience.

To directly test the experience dependency of object selectivity spread from 0° to other angles, we subtracted preferred/nonpreferred differences at 30°, 60°, or 90° from those at 0° in individual

cells (Fig. 6e). The decrease in the preferred/nonpreferred difference from 0 to 30° showed a significant dependence on the group ($p = 0.0083$, by Kruskal–Wallis test). The *post hoc* test indicated significant differences between the Object task and Across-set Image task preparation groups ($p = 0.025$) and between the Within-set Image task and Across-set Image task preparation groups ($p = 0.013$), but not between the Object task and Within-set Image task preparation groups ($p = 0.98$). The decrease in the preferred/nonpreferred difference from 0 to 60° or from 0 to 90° did not show any dependence on the group (both $ps > 0.85$).

Parts of the cellular results described above were obtained in experiments with a counterbalanced design. i.e., effects of two types of preparatory experience were compared between two monkeys by using the same set, and also between two stimulus sets on the same monkey. Because the object selectivity of cells at 30° from the preferred viewing angle showed a clear difference, even with the same sets, between the cells recorded from the monkeys that had experienced the sets in the Within-set Image task and those recorded from the monkeys that had experienced the sets in the Across-set Image task ($p = 0.0032$, by Mann–Whitney *U* test), the differences were not due to differences in preparatory experience (Fig. 9). We also implemented a set of counterbalanced experiments for the comparison between preparatory experience in the Object task and that in the Within-set Image task: Set D had been shown in the Within-set Image task for Monkey 1 and in the Object task for Monkey 3, whereas Set E had been shown in the Object task for Monkey 1 and in the Within-set Image

task for Monkey 3. There was no significant difference in the preferred versus nonpreferred differences at 30° between the cells recorded from the monkeys that had experienced the sets in the Within-set Image task and the cells recorded from the monkeys that had experienced the sets in the Object task ($p = 0.66$, by Mann–Whitney *U* test).

The differences among the three groups were not due to differences in general effectiveness of stimuli. There were no significant differences in the magnitude of the largest responses to the preferred stimulus in each cell among the groups (mean $\pm$ SEM, 23.6 ± 2.0 , 21.6 ± 1.5 , and 24.7 ± 2.1 spikes/s after experience in the Object, Within-set, and Across-set Image tasks, respectively; $p = 0.53$, by Kruskal–Wallis test). The differences in each monkey were also not significant (26.6 ± 2.8 , 18.8 ± 2.1 , and 22.1 ± 3.1 spikes/s after experience in the Object, Within-set, and Across-set Image tasks, respectively, and $p = 0.052$ in Monkey 1; 22.0 ± 2.7 , 23.8 ± 2.1 , and 27.1 ± 2.8 spikes/s after experience in the Object, Within-set, and Across-set Image tasks, respectively, and $p = 0.15$ in Monkey 3).

Note that the object selectivity over multiple viewing angles found in the present study is different from the perfectly view-invariant object selectivity. In the tunings of normalized responses to the stimulus sets experienced in the Within-set Image task (Fig. 6*b*), the mean responses to the preferred object at 60 and 90° from the preferred viewing angle were comparable or smaller than the mean response to the nonpreferred objects at the preferred viewing angle ($p = 0.068$ for 60° and $p = 0.018$ for 90°, by sign test). Even in the tunings of normalized responses to the stimulus sets experienced in the Object task (Fig. 6*a*), the mean response to the preferred object at 90° from the preferred viewing angle was smaller than the mean response to the nonpreferred objects at the preferred viewing angle ($p = 0.007$; whereas there was no significant difference at 60°, $p = 0.784$). None of the 179 cells that we recorded in the present study individually showed perfect view invariance. The best cell (Fig. 5, Cell 1), which was examined with a stimulus set that the monkey had experienced in the Object task, showed comparably strong responses to a nonpreferred object at the preferred viewing angle and the preferred object at the worst viewing angle (both 0.34 of the maximum response to the preferred object at the preferred viewing angle). In the second best cell, which was also examined with a stimulus set that the monkey had experienced in the Object task, responses to views of the preferred object at 60 and 90° from the preferred viewing angle were smaller than the responses to two nonpreferred objects at the preferred viewing angle.

In summary, for the stimulus sets that the monkey had experienced in the Object or Within-set Image task, the greater response to the preferred object relative to nonpreferred objects occurred not only at the preferred viewing angle, but spread out to other nearby viewing angles. No significant response generalization occurred when objects were experienced under the Across-set Image task.

Response patterns in model cell populations

We have so far focused on the responses of single cells. Not only the cells that maximally responded to the sample stimulus but also other cells tuned to other stimuli in the stimulus set may contribute to the discrimination, and then a population of cells may perform better than single cells in view-invariant object recognition. Therefore, we analyzed response patterns distributed over a cell population. Because we did not record enough cells for each stimulus set, we decided to examine the performance of model cell populations.

Figure 6. Averaged normalized magnitudes of the responses to images of the preferred (solid lines) and nonpreferred (broken lines) objects plotted against viewing angle differences of the stimuli from the preferred viewing angle. Responses were first normalized by the maximum response to the preferred view of the preferred object in each cell, and then averaged across cells within each group of cells. The three types of broken lines indicate magnitude of averaged responses for objects of ranks 2, 3, and 4. The diamonds without connecting lines indicate magnitude of responses averaged over the three nonpreferred objects. *a*, Averaged normalized responses to the stimulus sets that the monkeys had experienced in the Object task. *b*, Those to the stimulus sets that the monkeys had experienced in the Within-set Image task. *c*, Those to the stimulus sets that the monkeys had experienced in the Across-set Image task. The asterisks in *a–c* indicate the statistical significance of the differences between the averaged responses to the images of the preferred and nonpreferred objects at each angle difference from the preferred viewing angle; $**p < 0.001$; $*p < 0.05$; no mark, $p \geq 0.05$. The numbers in parentheses under x-axes in *a–c* indicate the number of cells contributed to the calculation at the angle difference. *d*, The differences between normalized magnitudes of the responses evoked by the images of the preferred and nonpreferred objects at the preferred viewing angle and angles 30, 60, and 90° different from the preferred angle. *e*, The decrease in the preferred–nonpreferred difference plotted in *d* from 0 to 30°, from 0 to 60°, and from 0 to 90°. The responses averaged over the three nonpreferred objects were used in *d* and *e*. Blue, red, and green bars indicate the values for the responses evoked by the stimuli that the monkeys had experienced in the Object, Within-set, and Across-set Image tasks, respectively. The asterisks in *d* and *e* indicate the statistical significance of the differences between the groups defined by the type of preparatory experience; $**p < 0.01$; $*p < 0.05$; no mark, $p > 0.05$. Error bars in *a–e* represent SEM across cells.

We made three cell populations by pooling cells recorded with the stimulus sets that the monkeys had experienced in the same task to separately analyze the responses to the stimuli that the monkeys had experienced in the Object, Within-set Image, and Across-set Image tasks. We assumed that the rank order of objects, including the preferred object, and the preferred viewing angle were merely results of random sampling and only the tun-

Figure 7. Averaged normalized magnitudes of the responses to images of the preferred (solid lines) and nonpreferred (broken lines) objects for cells recorded in Monkey 1 (left) and Monkey 3 (right). The notations are the same as those in Figure 6a–c.

ing over viewing angles and between the preferred and nonpreferred objects had relevance, as discussed in the previous section. Then, we duplicated the recorded cells by placing the most effective stimulus at all the objects in the set and all the viewing angles, except that the cells that showed the largest responses at view 30° or 60° were duplicated only at view 30° or view 60° (with the largest responses at view 30° or 60°; Fig. 10). Cells were also duplicated by shuffling the objects of rank 2–4. Needless to say, cells were not duplicated across different preparatory tasks: cells recorded with the sets that the monkeys had experienced in a task were used only to form the cell population used to analyze effects of preparatory experience in the task. Effects of the ongoing task on neuronal responses, e.g., repetition suppression and feature-based attention, were not considered.

We first examined changes in response pattern by calculating correlation between the two response patterns. How much did the response pattern change by pure rotations or by combined object-identity changes and rotations? The amount of change in response pattern can be quantified by 1 minus the coefficient of correlation between the two response patterns. As for view change of 30°, we focus on those between view 0° and view 30° and between view 30° and view 60°. Other pairs of views with a 30° difference had a set of correlation coefficients the same as the set in either of the two pairs. For the stimulus sets that the monkey had experienced in the Object or Within-set Image task, the

population response pattern changed only moderately for pure rotations of an object (with correlation coefficient of 0.53 and 0.44 in view 0° vs view 30° and view 30° vs view 60° comparisons, respectively, for the sets experienced in the Object task and 0.52 and 0.49 in view 0° vs view 30° and view 30° vs view 60° comparisons, respectively, for the sets experienced in the Within-set Image task), whereas there were larger changes for combined object and view changes (with correlation coefficients of 0.17 and -0.00040 view 0° vs view 30° and view 30° vs view 60° comparisons, respectively, for the sets experienced in the Object task and 0.31 and 0.26 in view 0° vs view 30° and view 30° vs view 60° comparisons, respectively, for the sets experienced in the Within-set Image task). By detecting a big change in the response pattern, the system could discriminate changes including object changes from pure rotations of an object. However, for the stimulus sets that the monkey had experienced in the Across-set Image task, the changes in response pattern were similar in magnitude for both pure rotations of an object (with correlation coefficients of 0.35 and 0.26 for view 0° vs view 30° and view 30° vs view 60° comparisons, respectively) and combined object and view changes (with correlation coefficients of 0.28 and 0.16 for view 0° vs view 30° and view 30° vs view 60° comparisons, respectively). It was therefore more difficult for the system to selectively detect object changes after preparatory experience in the Across-set Image task.

However, the above-described algorithm, which was based on general similarity (correlation coefficients) of response patterns, had difficulty explaining the monkeys' occasional success with 60° rotation. For pure rotations of 60°, the correlation coefficients were 0.32 and 0.35 for the sets that the monkey had experienced in the Object and Within-set Image tasks, respectively. Note that the two pairs of viewing angles with 60° difference (view 0° vs view 60°, and view 30° vs view 90°) produced the same set of correlation coefficients. Although these changes in response pattern evoked by pure 60° rotations were smaller than those evoked by combinations of object changes and 60° rotation (with correlation coefficients of -0.01 and 0.14 for the Object and Within-set Image tasks, respectively), the monkey, or the system, had to lower the correlation threshold for discriminating a pure rotation from combined (rotation + object) changes compared with a 30° rotation. This indicates that a general model would have to base discriminations on a more complex criterion than a fixed threshold for the amount of change in population activity. For example, when a test stimulus is presented, the system has to first determine the amount of rotation from the sample to test stimuli, select an appropriate threshold for the rotation, and then apply the threshold to the calculated similarity. For the stimulus set that the monkey had experienced in the Across-set Image task with a 60° rotation, the correlation coefficients were 0.14 and 0.0072 for the pure rotation of an object and combined (rotation + object) changes, respectively.

Second, we took a linear classification approach to analyze response patterns. Analyzing response patterns distributed over a cell population is equivalent to analyzing population responses in a response space spanned by responses of the cells constituting the population. A response pattern evoked by a stimulus over the cell population is represented by a position in the multidimensional space. We here determined a hyperplane that optimally divided responses evoked by view 0° of object 1 from responses evoked by views 0° of objects 2, 3, and 4 with the longest distances by using a support vector machine (SVM) procedure (SVM^{light}; <http://svmlight.joachims.org/>) with a linear kernel, the C-SVM algorithm, and cost (C) set to 0.5. We assumed here that the monkey, or the system, had developed this hyperplane through

the preparatory experience to discriminate view 0° of object 1 from views 0° of objects 2, 3, and 4 in the Within-set Image task. Finally, we examined the positions of responses evoked by other views in relation to the hyperplane. The results obtained with all the cells recorded from the two monkeys are shown in Figure 11, in which the responses evoked by the 16 stimuli are plotted in a 2D plane spanned by the axis perpendicular to the hyperplane and the axis of the first principal component of the remaining variability in positions. The same side of the space relative to the hyperplane as the view 0° of the object will be referred to as the correct side and the opposite side as the wrong side. Responses evoked by views 30° and 60° of object 1 were located far from views 30° and 60° of objects 2, 3, and 4 and both views 30° and 60° of object 1 and those of objects 2, 3, and 4 were located at the correct sides of the hyperplane (indicated by the broken line) for the stimulus sets that the monkey had experienced in the Within-set Image task (Fig. 11, top). However, views 30° and 60° of object 1 were located close to views 30° and 60° of objects 2, 3, and 4 and at the wrong side for the stimulus sets that the monkey had experienced in the Across-set Image task (Fig. 11, bottom). These differences may explain why the monkeys could perform the view-invariant recognition with 30 and 60° rotations after the preparatory experience of the set in the Within-set Image task, but not after the preparatory experience of the set in the Across-set Image task.

The results of the SVM analyses for each of the two monkeys, i.e., those obtained with cells recorded from each monkey, are shown in Figure 12. The data of Monkey 1 are very similar to those of the pooled data shown in Figure 11. In Monkey 3, views 30° and 60° of object 1 were located close to the hyperplane after the preparatory experience in the Within-set Image task, although they were located far right from the positions of corresponding views of objects 2, 3, and 4. The object identity changed with a probability of 1/3 at each stimulus presentation during the tasks, and the monkeys generally committed more miss errors (failing in responding to object changes) than false alarm errors (falsely responding when the object identity did not change) during the preparation in the Within-set Image task. Specifically, the ratios of miss and false alarm that Monkey 3 committed during preparation in the Within-set Image task were 0.25 and 0.05, respectively, for Set A and 0.06 and 0.02, respectively, for Set E. These facts suggest that the hyperplane was placed at a position closer to views 0° of the other objects (objects 2, 3, and 4) than the first one (object 1). Then, views 30° and 60° of object 1 were located well in the correct side far from the

Figure 8. Comparison of normalized magnitudes of the responses evoked in individual cells by the preferred object image and nonpreferred object images at 30° from the preferred viewing angle. *a*, Cells/responses examined with the object sets that the monkeys had experienced in the Object task. *b*, Cells/responses examined with the object sets that the monkeys had experienced in the Within-set Image task. *c*, Cells/responses examined with the object sets that the monkeys had experienced in the Across-set Image task. In the scatter plots, the x- and y-positions indicate the normalized magnitude of responses to the preferred object and nonpreferred object images, respectively, at 30° from the preferred viewing angle. One dot corresponds to one cell. Dark circular disks indicate cells recorded from Monkey 1, and light triangles indicate cells recorded from Monkey 3. Note that responses to nonpreferred object images were averaged over the three objects. The bar graphs on the right plot the number of cells that had differences between the normalized magnitudes of responses evoked by the preferred object and nonpreferred object images (preferred–nonpreferred) at 30° from the preferred viewing angle in the bin. Arrows indicate the median values.

hyperplane, and the results for Monkey 3 still well explain its successful performance in the tests of view-invariant recognition after the preparation in the Within-set Image task.

Discussion

We found that, once monkeys learned to discriminate among a new set of objects in each of four different viewing angles at 30° intervals, they could immediately recognize the objects across a certain range of viewing angles without any further learning. Correspondingly, neurons recorded from the inferotemporal cortex of the monkeys showed object selectivity spreading over multiple viewing angles for the object images.

Figure 9. Cell responses obtained in experiments with a counterbalanced design, i.e., effects of two types of preparatory experience were compared between two monkeys with the same sets. During the preparation before cell recordings, Set A was shown in the Across-set Image task to Monkey 1 and in the Within-set Image task to Monkey 3. On the contrary, Set C was shown in the Within-set Image task to Monkey 1 and in the Across-set Image task to Monkey 3. We recorded responses with Set A and Set C from 20 cells (11 cells with Set A and 9 cells with Set C) in the monkeys that experienced these sets in the Across-set Image task, and from 31 cells (14 cells with Set A and 17 cells with Set C) in the monkeys that experienced these sets in the Within-set Image task. There were significant differences between the normalized responses to the preferred objects and those to the nonpreferred objects at 30° from the preferred viewing angle in the cells recorded from the monkeys that had experienced the sets in the Within-set Image task ($p < 0.0002$, by sign test), whereas the differences were not significant in the cells recorded from the monkeys that had experienced the sets in the Across-set Image task ($p = 0.9999$). The preferred versus nonpreferred differences (preferred–nonpreferred) at 30 deg were significantly larger in the cells recorded from the monkeys that had experienced the sets in the Within-set image task ($p = 0.0032$, by Mann–Whitney U test); $**p < 0.001$. Error bars depict the SEs. Other notations are the same as those in Figure 6a–c.

These behavioral and cellular results suggest that the experience of discriminating images of new objects in each of several viewing angles results in a development of partially view-generalized object selectivity distributed over many neurons in the inferotemporal cortex, which in turn bases the mon-

Figure 10. The procedure to make a model cell population from experimental data. For the cells with the largest responses at view 0° or view 90°, one new cell was first created by shifting the original responses along the axis of view by 30° and filling the column of view 0° by assuming a symmetrical tuning (Copy 1, top). Then, two more new cells were created by mirroring the original responses and Copy 1 (Copy 2 and Copy 3, top). Note that the view arrangement was preserved (either the original or mirror symmetrical). For the cells with the largest responses at view 30° or view 60°, only one new cell was generated by mirroring the experimental data for views (Copy 1, bottom). Finally, for both groups of cells, objects were shuffled to create 23 new cells from each cell. Thus, 95 new cells were created from a cell with the largest response at view 0° or view 90°, and 47 new cells were created from a cell with the largest response at view 30° or view 60°. A cell population was formed separately for the Object task, Within-set Image task, and Across-set Image task. Cells recorded with the stimulus sets that the monkey had experienced in a task were used only to form the cell population used for examining the effects of prior experience in that task.

keys' emergent capability to discriminate the objects across changes in viewing angle.

For the stimulus sets that the monkey had experienced in the Within-set Image task (that required only within-view discrimination), strong responses to the preferred view of the preferred object selectively spread to nearby views of the preferred object, but not to views of nonpreferred objects, and thereby there were significant differences between responses to the preferred and nonpreferred objects not only at the preferred viewing angle but also at angles 30 and 60° different from the preferred angle (Fig. 6). This spread of object-selective responses from the preferred view to nearby views seen after the experience in the Within-set Image task was as large as that seen after the experience in the Object task (which required view-invariant discrimination), whereas there was no such spread after the experience in the Across-set Image task (which did not require within-set discrimination). By finding the cells that were maximally activated by the sample stimulus and checking whether their responses to the following test stimulus were larger than a certain threshold, the system can determine, for rotations up to 60°, whether the test stimulus was a view of the same object as that of the sample stimulus or a view of another object.

Not only the cells that maximally responded to the sample stimulus but also other cells tuned to other stimuli in the stimulus set may contribute to the discrimination. The algorithm known as linear classifier is useful when we discuss mechanisms of view-invariant object recognition by a cell population. By linearly combining activities of cells tuned to different features with appropriate weights and setting an appropriate threshold, the output cell may be able to commonly discriminate views of a particular object from views of different objects. This is equivalent to setting a linear hyperplane that separates views of an object

After experience in Within-set Image task

After experience in Across-set Image task

Figure 11. Positions of object images relative to the hyperplane that optimally divided the view 0° of object 1 from view 0° of objects 2, 3, and 4. The plots were calculated separately for the responses to the stimulus sets that the monkeys had experienced in the Within-set Image task (top) and Across-set Image task (bottom). The horizontal axis represents the axis orthogonal to the hyperplane. The vertical axis represents the axis of the first principal component of the variances of the object images' positions projected onto the hyperplane. The unit of the coordinates is arbitrary but common to the vertical and horizontal axes and different graphs. The images of objects 2, 3, and 4 are located at the same positions on the projection plane.

from views of other objects in the response space spanned by the magnitudes of individual cells' responses (DiCarlo et al., 2012). We can use this framework to explain the view-invariant recognition capability that the monkeys showed immediately after experience of the objects in the Within-set Image task. Because the monkeys were trained only for the discrimination of objects within each viewing angle, the training should have developed multiple hyperplanes, each of which was specific to discrimination at each viewing angle. Nevertheless, for the stimulus sets that the monkey had experienced in the Within-set Image task, the hyperplane formed for one viewing angle was useful for recognition of object images at other nearby viewing angles. The views of the preferred and nonpreferred objects at 30 and 60° differences were located at the appropriate sides of the hyperplane (Fig. 11, top). These were not the case for the stimulus sets that the monkey had

Figure 12. Positions of object images determined by using only cells recorded from Monkey 1 (left) and those determined by using only cells recorded from Monkey 3 (right). The notations are the same as those in Figure 11.

experienced in the Across-set Image task (Fig. 11, bottom). In fact, after a month of experience in the Across-set Image task, 2 d of learning the within-view discrimination, which was enough to achieve very good discrimination performance within each viewing angle (and then to form appropriate hyperplanes for within-view discriminations), did not improve the recognition at 30° or 60° difference (Yamashita et al., 2010, their Fig. 7).

Common to all the explanations described above, the difference in the monkey's behavior after the preparatory experience in the Within-set Image task and after the preparatory experience in the Across-set Image task was due to the presence or absence of specific spreading of strong responses from the preferred view of the preferred object to the nearby views of the preferred object. By "specific" we note that the spreading did not occur to views of the nonpreferred objects. Due to that specific spreading, views of each object were located closer to one another in the response space and were aligned more parallel to the hyperplanes determined for discrimination at each viewing angle after the experience in the Within-set Image task than after the experience in the Across-set Image task.

How did this specific spreading of strong responses within each object develop during the preparatory experience in the Within-set Image task? Based on previous findings of experience-based changes in responsiveness or selectivity of inferotemporal cells (Logothetis et al., 1995; Kobatake et al., 1998; Baker et al., 2002; Sigala and Logothetis, 2002), we assume that, as the monkeys learned to discriminate among the objects at each of several viewing angles, the inferotemporal cells that were useful for the discriminations were reinforced. An interaction likely occurred between nearby viewing angles in this reinforcement. If there had been cells that had some degree of responsiveness to multiple nearby views of an object, their responses would have been more frequently reinforced than other cells that were exclusively re-

sponsive to a single view of an object. Thus, common responses to nearby views would have specifically developed via the discrimination learning in the preparatory experience. Because the spreading of strong responses occurred only within each object, the supposed interaction between nearby viewing angles should have occurred in the domain of a specific group of higher-order features, with which the similarity between different views of the same objects was larger than the similarity between different views of different objects. One possibility is that, through previous experience with many natural objects that had occurred from birth, cells in the inferotemporal cortex and earlier stages of the ventral visual pathway had become more tuned to features less sensitive to rotation of objects (Vogels et al., 2001; Kayaert et al., 2003) before the discrimination learning in the present experiments started. This pre-existing bias was enhanced by the preparatory experience in the Within-set Image task, but not by the preparatory experience in the Across-set Image task. Also, because the differences between responses to the preferred and nonpreferred objects were not significantly larger after the preparatory experience in the Within-set Image task than after the preparatory experience in the Across-set Image task, the supposed reinforcement may have generally enhanced responses of the cell without sharpening the selectivity. Although there is a good consensus that the proportion of inferotemporal cells useful for discrimination increases by discrimination training, it is controversial whether discrimination training sharpens the selectivity tuning of inferotemporal cells (De Baene et al., 2008).

In natural life, objects do rotate. In this sense, the preparatory experience in the Within-set Image task, in which objects never rotate, was artificial. However, that does not mean that the mechanism revealed by the preparatory experience in the Within-set Image task is unlikely to work in natural life. While objects rotate, there are also many opportunities to discriminate among objects at each of multiple viewing angles. Both the interaction between discrimination training in different viewing angles and the association of different views based on time contingency should work in daily life. We have to note a few qualifications to the above-described conclusions. First, view-invariant recognition of a new object is immediate even without any prior experience of the object, when its shape is very different from those of distractor objects (Biederman, 1987; Logothetis et al., 1994; Hummel, 2001). We have previously confirmed this phenomenon in our experimental setting (Wang et al., 2005, their Fig. 3). In this range of object-distractor differences, view-invariant recognition may depend on the detection of unique features common to any image of the object regardless of the viewing angle but not in any image of the distractors (Biederman, 1987; Hummel, 2001) or mental rotation of the representation (Tarr, 1995). Second, the discrimination preparation for each of several viewing angles produced view-invariant recognition across viewing-angle changes only up to 60° (Fig. 4; Wang et al., 2005, their Fig. 4). Active learning of the pairing between different views of the same object and/or experiencing successive appearance of them may be necessary for recognition across larger changes in viewing angle.

To conclude, we found that once monkeys learned to discriminate among a set of objects at each of a few discrete viewing angles, individual neurons in the inferotemporal cortex developed object selectivity distributed over multiple viewing angles. Based on population modeling incorporating the properties of individual cells measured here, we suggest that this spread of object selectivity across views underlies the monkeys' immediate capability of view-invariant recognition on new stimulus sets after the preparatory experience.

References

- Baker CI, Behrmann M, Olson CR (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci* 5:1210–1216. [CrossRef Medline](#)
- Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115–147. [CrossRef Medline](#)
- Booth MC, Rolls ET (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* 8:510–523. [CrossRef Medline](#)
- Bülhoff HH, Edelman S (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci U S A* 89:60–64. [CrossRef Medline](#)
- Cox DD, Meier P, Oertel N, DiCarlo JJ (2005) 'Breaking' position-invariant object recognition. *Nat Neurosci* 8:1145–1147. [CrossRef Medline](#)
- Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans Inf Theory* 36:961–1005. [CrossRef](#)
- De Baene W, Ons B, Wagemans J, Vogels R (2008) Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learn Mem* 15:717–727. [CrossRef Medline](#)
- DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73:415–434. [CrossRef Medline](#)
- Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845–851. [CrossRef Medline](#)
- Földiák P (1991) Learning invariance from transformation sequences. *Neural Comput* 3:194–200. [CrossRef](#)
- Hummel JE (2001) Complementary solutions to the binding problem in vision: implications for shape perception and object recognition. *Vis Cogn* 8:489–517. [CrossRef](#)
- Humphrey GK, Khan SC (1992) Recognizing novel views of three-dimensional objects. *Can J Psychol* 46:170–190. [CrossRef Medline](#)
- Kayaert G, Biederman I, Vogels R (2003) Shape tuning in macaque inferior temporal cortex. *J Neurosci* 23:3016–3027. [Medline](#)
- Kobatake E, Wang G, Tanaka K (1998) Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80:324–330. [Medline](#)
- Logothetis NK, Pauls J, Bülhoff HH, Poggio T (1994) View-dependent object recognition by monkeys. *Curr Biol* 4:401–414. [CrossRef Medline](#)
- Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5:552–563. [CrossRef Medline](#)
- Masquelier T, Thorpe SJ (2007) Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput Biol* 3:e31. [CrossRef Medline](#)
- Rock I, DiVita J (1987) A case of viewer-centered perception. *Cogn Psychol* 19:280–293. [CrossRef Medline](#)
- Sakai K, Miyashita Y (1994) Neuronal tuning to learned complex forms in vision. *Neuroreport* 5:829–832. [CrossRef Medline](#)
- Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415:318–320. [CrossRef Medline](#)
- Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109–139. [CrossRef Medline](#)
- Tanaka K, Saito H, Fukada Y, Morioka M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189. [Medline](#)
- Tarr MJ (1995) Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychon Bull Rev* 2:55–82. [CrossRef Medline](#)
- Vogels R, Biederman I, Bar M, Lorincz A (2001) Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *J Cogn Neurosci* 13:444–453. [CrossRef Medline](#)
- Wallis G, Bülhoff HH (2001) Effects of temporal association on recognition memory. *Proc Natl Acad Sci U S A* 98:4800–4804. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Wiskott L, Sejnowski TJ (2002) Slow feature analysis: unsupervised learning of invariances. *Neural Comput* 14:715–770. [CrossRef Medline](#)
- Wyss R, König P, Verschure PF (2006) A model of the ventral visual system based on temporal stability and local memory. *PLoS Biol* 4:e120. [CrossRef Medline](#)
- Yamashita W, Wang G, Tanaka K (2010) View-invariant object recognition ability develops after discrimination, not mere exposure, at several viewing angles. *Eur J Neurosci* 31:327–335. [CrossRef Medline](#)