The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

1860-10

Borsellino College 2007. Spike Trains to Actions: Brain Basis of Behavior

3 - 14 September 2007

Visual Perception
(Object category structure in monkey IT cortex)

Hossein ESTEKY
School of Medicine, Shaheed Beheshti Univ. and School of Cognitive Sciences, IPM, Tehran, Iran

Articles in PresS. J Neurophysiol (April 11, 2007). doi:10.1152/jn.00024.2007

Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex

Roozbeh Kiani[1,3], **Hossein Esteky**[1,2], **Koorosh Mirpour**[2] **& Keiji Tanaka**[4,5]

[1]Research Group for Brain and Cognitive Sciences, School of Medicine, Shaheed Beheshti University, Tehran, Iran

[2]School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics, Niavaran, Tehran, Iran

[3]Department of Neurobiology and Behavior, University of Washington, Seattle, Washington, USA

[4]Cognitive Brain Mapping Laboratory, RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan

[5]Graduate School of Science and Engineering, Saitama University, Saitama, Saitama 338-8570, Japan

Running head

Object category structure in monkey IT cortex

Correspondence should be addressed to:

Hossein Esteky, Research Group for Brain and Cognitive Sciences, School of Medicine,
Shaheed Beheshti University, P.O. Box 19835-181, Tehran, Iran
E-mail: esteky@ipm.ir
Tel: (+98) 21 241 4164
Fax: (+98) 21 228 0352

Abstract

Our mental representation of object categories is hierarchically organized, and our rapid and seemingly effortless categorization ability is crucial for our daily behavior. Here, we examine responses of a large number (>600) of neurons in monkey inferior temporal (IT) cortex with a large number (>1000) of natural and artificial object images. During the recordings the monkeys performed a passive fixation task. We found that the categorical structure of objects is represented by the pattern of activity distributed over the cell population. Animate and inanimate objects created distinguishable clusters in the population code. The global category of animate objects was divided into bodies, hands and faces. Faces were divided into primate and non-primate faces, and the primate-face group was divided into human and monkey faces. Bodies of human, birds, and four-limb animals clustered together, while lower animals such as fish, reptile and insects made another cluster. Thus, the cluster analysis showed that IT population responses reconstruct a large part of our intuitive category structure, including the global division into animate and inanimate objects, and further hierarchical subdivisions of animate objects. The representation of categories was distributed in several respects, e.g., the similarity of response patterns to stimuli within a category was maintained by both the cells that maximally responded to the category and the cells that responded weakly to the category. These results advance our understanding of the nature of the IT neural code, suggesting an inherently categorical representation that comprises a range of categories including the amply investigated face category.

Keywords

Ventral visual pathway, area TE, distributed representation, multidimensional scaling, cluster analysis

Introduction

The mental representation of object categories has been a source of general and perpetual interest in cognitive neuroscience. Several imaging studies have investigated this representation in humans (e.g., Aguirre et al. 1998; Allison et al. 1994; Chao et al. 1999; Epstein and Kanwisher 1998; Gauthier et al. 1999; Gauthier et al. 2000; Haxby et al. 2001; Kanwisher et al. 1997; Martin et al. 1996; McCarthy et al. 1997) and suggest that several object classes such as faces, houses, animals, and tools are represented in human temporal cortex. However, the use of a limited (and potentially biased) stimulus set and posing a presumed category structure limits the scope of many of these studies. Studies at the level of single cells in nonhuman primates share these shortcomings. Although the existence of face-selective cells in the inferior temporal (IT) cortex of naive monkeys is well established (Bruce et al. 1981; Desimone et al. 1984; Kiani et al. 2005; Perrett et al. 1982; Rolls and Tovee 1995; Tsao et al. 2006), the generality of the finding to the representation of other object categories is unknown.

In monkeys trained to categorize stimuli into a few arbitrary groups, some single cells in the prefrontal cortex show responses covering most stimuli in one of the learned categories (Freedman et al. 2001, 2002). In humans, some cells in medial temporal lobe structures such as the hippocampus respond categorically (Kreiman et al. 2000; Quiroga et al. 2005). Both prefrontal cortex and medial temporal lobe structures receive visual input about objects from IT cortex. Although the stimulus selectivity of IT cells is affected by training for visual categorization (Baker et al. 2002; Sigala et al. 2002), responses of single IT cells appear to represent individual stimuli rather than the learned categories (Freedman et al. 2003; Vogels 1999). Taken together, these studies suggest that object categories could be represented in the polymodal association cortices downstream of the IT cortex. Although some types of categorization may rely on the prefrontal cortex and medial temporal structures, it is conceivable that certain classes of visual categories would be represented in the IT cortex. Given our rapid and seemingly effortless ability for categorization of natural objects (Li et al. 2002; Thorpe et al. 1996), natural categories provide plausible candidates.

A possibility, which has not been tested extensively, is that object categories are represented by response patterns over a large population of IT cells. According to this hypothesis objects that belong to the same category would tend to elicit similar population response patterns. We therefore asked whether, for a large (>1,000) set of natural object images rather than arbitrary and limited sets of object images, responses of a population of IT cells (rather than single units) represent the categorical structure of objects. Responses to the stimuli were examined during a fixation task, to investigate object categories independent of artificially-imposed or task-dependent requirements. Furthermore, the use of a large stimulus set with many categories allowed us to examine the representation of object categories in a data-driven fashion without prior assumption of any particular category structure. We found that response patterns distributed over the IT cell population represented many animate categories, as well as the structure among them.

Methods

Three macaque monkeys (*M. mulatta*) were used, two for single cell recordings and one for behavioral experiments. All experimental procedures complied with the guidelines of the National Institutes of Health and the Iranian Society for Physiology. The monkeys were raised in human houses and then in zoos before they were brought to the laboratory. Therefore, it is likely that they had encountered many animate and inanimate objects in the course of their life.

Recordings and stimuli

In a preparatory aseptic surgery, a block for head fixation and recording chamber were anchored to the dorsal surface of the skull. The position of the recording chamber was determined stereotaxically referring to the magnetic resonance images (MRIs) acquired before the surgery. Action potentials of single cells were recorded extracellularly with tungsten electrodes (FHC, ME) from the IT cortex, on the right side for one monkey (Monkey 1) and on the left side for the other monkey (Monkey 2), while the monkeys were performing a fixation task. The electrode was advanced with an oil-driven manipulator (Narishige, Japan) from the dorsal surface of the brain through a stainless steel guide tube inserted into the brain down to 10-15 mm above the recording sites. Recording positions were evenly distributed at anterior 15-20 mm (Monkey 1) or 13–20 mm (Monkey 2) over the ventral bank of the superior temporal sulcus and the ventral convexity up to the medial bank of the anterior middle temporal sulcus with 1-mm track intervals (Fig. 1). The recording was not biased by response properties. The action potentials from a single neuron were isolated in real time by a template matching algorithm (Worgotter et al. 1986).

Responses of each cell were tested with 1124±71 (mean±SD; median, 1084) stimuli presented in a pseudorandom order. The stimulus set was repeated 9±2 (mean±SD; median, 10) times for each recording site. The sequence of stimuli changed randomly in each repetition, and also for different recording sites, to avoid any consistent interaction between successively presented stimuli. The stimuli were colorful photographs of natural and artificial objects isolated on a gray background. The size of the larger dimension (vertical or horizontal) of each stimulus was ~7 degrees of visual angle.

The monkey had to maintain fixation within ±2 deg of a 0.5 deg fixation spot presented at the center of the display. The eye position was measured by an infra-red eye-tracking system (i_rec, http://staff.aist.go.jp/k.matsuda/eye/), which allowed a precision of 1 deg or less for the measurement of eye position. The presentation of stimulus sequence started after the monkey maintained fixation for 300 ms. Each stimulus lasted for 105 ms and was followed by another stimulus without intervening gap. The sequence stopped when 60 stimuli were presented or when the monkey broke the gaze fixation. The monkey was rewarded with a drop of juice every 1.5–2 seconds during the fixation. It has been shown that cells in the monkey IT cortex preserve their stimulus selectivity in rapid serial presentations as fast as 14–28 ms/stimulus (Edwards et al. 2003; Foldiak et al. 2004; Keysers et al. 2001). Also, backward masking has a minimal effect on the initial part of neuronal responses when the stimulus onset asynchrony is longer than 80 ms (Kovacs et al. 1995; Rolls and Tovee 1994).

Data analyses

The data set consisted of all the cells with reliable unit isolation throughout the stimulus presentation, regardless of the cell's stimulus selectivity (n=674). The spontaneous activity was measured in a 200-ms window immediately before the sequence of stimulus presentation initiated, and its standard deviation was calculated across different sequences. We measured the neural activity for each stimulus presentation in a 140-ms window starting 71 ms and ending 210 ms after the onset of the stimulus. Responses to the last two stimuli in each sequence did not enter the analysis. To minimize the contamination of neural activity measurement by responses to the previous stimuli, we excluded presentations with large activity (exceeding the spontaneous activity by 2×s.d.) in the 50-ms period immediately after the stimulus onset. This resulted in exclusion of 15% of the presentations. However, neither the contamination correction nor the exact size of the window was crucial to the results. Due to our rapid stimulus presentation paradigm we could not assess how our results might have changed by taking into account very late responses of the cells.

Similarity of response patterns measured with degree of correlation

Responses elicited in the population of cells were used to calculate a measure of similarity between stimuli (Fig. 2). First, the mean responses of a cell to the set of stimuli were arranged in a vector, and were normalized by subtracting the mean response of the cell from the vector and then dividing it by its Euclidean length. The response normalization for single cells canceled the bias due to different baseline activity and different ranges of firing rates in different cells. Changes in the method of normalization did not change the basic results. Second, for any pair of stimuli, we calculated Pearson's correlation coefficient (r) between patterns of responses evoked by the stimuli in the cell population. The distance (or dissimilarity) between two stimuli was quantified by $1–r$ (neural distance). Two stimuli with similar response patterns in the cell population, therefore, have a small distance. The use of correlation coefficient as a measure of distance has the advantage of focusing on the population response pattern and discounting effects that nonspecifically change the firing rate of the IT population (e.g. contrast or luminance of the stimulus). Nevertheless, similar results were obtained with other distance metrics, such as Euclidean distance, which do not specifically measure the pattern similarity.

Multidimensional scaling and cluster analyses

Multidimensional scaling (MDS) (Young and Hammer 1987) was used to visualize the distribution of stimuli based on the neural distances (Fig. 4). Both classic MDS and nonlinear dimensionality reduction methods (Tenenbaum et al. 2000) showed segregation of categories in a low-dimensional space. Results of nonlinear MDS are shown in Fig. 4.

We also applied agglomerative cluster analysis (Johnson 1967) to the neural distances (Figs. 5 and S2, Tables 1 and 2, and Supplementary Program). The results shown here were obtained by measuring the distance between nodes by averaging distances of all pairs of stimuli under the two nodes. Varying the method of distance calculation (average, largest, shortest, and others), however, did not change the basic properties of the tree structure.

We listed 23 intuitive object categories, based on human convention, with at least 12 category members in the stimulus set (see Table 1). For each of them and also for higher categories made of them, we examined whether there was a corresponding node in the tree. Two indices were defined for this purpose:

Ratio 1 = (number of category members under the node) / (total members of the category)

Ratio 2 = (number of category members under the node) / (total stimuli under the node)

The average of the two ratios was used as a score for the match between the category and the node. We searched for the node with the maximum score for each category. To determine the mean value and variation of the score expected by chance clustering of the stimuli we repeated the same procedure for a group of randomly selected stimuli of the same size as the category (Monte Carlo method). The match of the higher categories (see Table 2) with the nodes was examined by the same method.

Cluster analysis on stimulus similarity in low-level features

We also applied cluster analysis to the physical similarity between stimulus images (Fig. 6A). Physical similarity was measured by 1) sum of absolute differences in red, green, and blue values over the pixels of two images, 2) sum of absolute difference in intensity over all pixels, and 3) sum of absolute differences in coefficients of Wavelet transformation of stimulus images. We used a biorthogonal wavelet (Daubechies 1992) from the Wavelet Toolbox of Matlab (bior 5.5 with seven levels of decomposition), but similar results were obtained with other wavelets as well.

Similarity of the stimuli was also measured by 4) outputs of a population of modeled V1 simple cells, and 5) modeled V1 complex cells (Fig. 6A). The receptive fields of simple cells were simulated by Gabor filters of different orientations (0, 90, -45 and 45 degrees), sizes (0.3 to 1.2 degrees in steps of 0.1 degree), and contrast selectivity (preferring lightness or darkness at the center of the receptive field) for cells without color selectivity. To introduce color information we replaced the intensity contrast with red-green, or blue-yellow). Cells were distributed over the stimulus image with 0.04-deg intervals between the receptive field centers of adjacent cells. Negative values in outputs were rectified to zero. Each of the images was presented to the model separately. The receptive fields of complex cells were modeled by MAX operation (Lampl et al. 2004; Riesenhuber and Poggio 1999) on outputs of neighboring simple cells with similar orientation selectivity. Simple cells were divided into four groups based on their receptive field size (0.3-0.4, 0.4-0.6, 0.6-0.9, 0.9-1.2 degree), and each complex cell pooled responses of neighboring simple cells in one of these groups. The spatial range of pooling varied in the four groups (4×4, 6×6, 9×9 and 12×12 for the four groups, respectively) (Riesenhuber and Poggio 1999). Similarity of responses evoked by two stimuli in the population of modeled cells was calculated either by the same method used for the responses in the IT cell population or by calculating the absolute value of difference in outputs of individual cells and summing over all cells. The results obtained by the latter method are shown in Fig. 6A, but similar results were obtained by the former method as well.

Cluster analysis on responses of model units tuned to randomly selected complex features

To examine whether the representation of object categories was a feature that could emerge without category knowledge in a population of units selective for complex features, we created model units with complex feature selectivity based on the hierarchical object recognition model of Riesenhuber and Poggio (1999) (Fig. 6B and the rightmost set of bars in Fig. 6A). This model effectively combines several experimental findings and consists of a hierarchy of units with increasingly complex stimulus selectivity and invariance. Units either perform template matching on their input to develop more complex pattern-specificity from simpler features (S units), or they perform a nonlinear operation (MAX) to develop invariance by pooling over units tuned to the same feature but at different positions or scales (C units). A hierarchy of units with these operations leads to C2 units, which are tuned to partially-complex features and are invariant to changes in position and scale (roughly similar to V4 neurons). The model was implemented with 256 C2 units as described in Riesenhuber and Poggio (1999), except that, due to the difference of image sizes in the two studies, S1 and C1 units' receptive field sizes were similar to the simple and complex V1 cells described above. The final stage of the hierarchy consisted of shape-tuned units (STUs) which were selective to the images in our stimulus set. Each STU received inputs from 32 C2 units that were most strongly activated by its preferred stimulus. We randomly selected 674 images from our stimulus set, and tuned the STUs to these images. The tuning width of the STUs was adjusted so that their response sparseness and response distribution matched the average of the recorded IT cells. The exact tuning width of STUs or the number of C2 units connected to each STU was not crucial for our basic results.

Selectivity of single cells for object categories

The selectivity of single cells for object categories was examined based on the 13 categories located at the lowest level in the tree of Fig. 5. We will refer to these 13 categories as "the lowest-level categories." Responses to individual presentations of all the stimulus members within each category were pooled for the analysis. A cell was regarded category-selective if responses to the best category were significantly larger than responses to any of the other categories (Newman-Keuls post-hoc, $p<0.05$) (similar to the two cells in Figs. 7A and B). A cell was also regarded selective to a combination of categories if responses to any category within the combination were significantly ($p<0.05$) larger than responses to any of the categories outside of the combination (similar to the cell in Fig. 7C).

To visualize the overlap of response magnitude distributions between categories, the mean responses to individual stimuli were plotted against the normalized stimulus rank for each category (Fig. 7, right column). The stimuli of each lowest-level category were ranked according to the magnitudes of mean responses, and the rank was normalized by the number of stimuli within the category. A normalized rank of one represents the stimulus that evoked the largest response within the category. To average these magnitude-rank profiles across the cells (Fig. 8, left column), the mean responses to individual stimuli were first normalized by the maximum response in each cell, and then averaged across cells. In Fig. 8, the stimuli were divided into two groups, those in the preferred category (or preferred category combination)

and those in the remaining categories.

In the right panel of Fig. 8A, the normalized mean responses to the individual stimuli were averaged over 10–20 cells preferring the same category in each monkey. The magnitude-rank profiles were then averaged across categories and monkeys**.** In the right panel of Fig. 8B, the normalized mean responses to individual stimuli were averaged among 11 or 20 cells selective to human faces in each monkey. The magnitude-rank profiles were then averaged for the two monkeys. Monkey and non-primate faces were excluded in Fig. 8B to allow comparison with previous studies.

We measured the significance of differences between responses to sub-optimal categories (Fig. 9) by comparing responses of a cell to different categories. For each cell, the lowest-level categories were ranked based on their average response magnitude, responses to individual presentations were pooled across stimuli belonging to each category, and the significance of difference in response magnitude was calculated for each pair of category rank (Wilcoxon test, significance defined as $p<0.05$). The proportion of cells showing a significant difference for each category-rank pair is presented for the 255 category-selective cells (Fig. 9A) and for the other 419 cells (Fig. 9B).

Correlation of mean response patterns between categories

For further evaluation of the contribution of cells without maximal responses to a category to the discrimination of the category (Fig. 10), we used a correlation analysis similar to the one employed by Haxby et al. (2001). The analysis was done in two stages. First, the members in each of the lowest-level categories were randomly divided into two equally sized groups, and the mean responses of each cell to each half-category group was calculated by averaging the normalized mean responses to individual stimuli. Mean responses of 674 cells to each half-category formed a response pattern. Second, the pairwise correlation of these response patterns was calculated for all possible pairs of categories (n=91). The procedure was repeated 1000 times with different random divisions of the categories, and the mean value of the correlation coefficient was obtained. Note that the correlation was calculated for responses to categories in this analysis, while the neural distances that were used for the previous analyses (e.g., the clustering analysis) were based on the correlations of response patterns to individual stimuli.

We performed three versions of the described correlation analysis. 1) The correlation was calculated for the responses of all 674 cells (white bars in Fig. 10). 2) The correlation was calculated after removing the cells that maximally responded to either of the two categories involved in each correlation calculation (gray bars in Fig. 10). 3) Finally, the responses of the cells that did not respond maximally to either of the paired categories were de-correlated by shuffling the mean response values across cells. The cells maximally responding to either of the paired categories were not shuffled. The correlation was calculated for the combination of shuffled and non-shuffled data points (black bars in Fig. 10).

Spatial distribution of cells with similar categorical selectivity

To examine whether there was clustering of cells with similar category selectivity, we compared responses of cell pairs with different spatial distances in the recording region. The position of the recording sites were estimated based on the location of the guide tube, the reading of the manipulator, the

estimated distribution and location of gray matter, and the depth of ventral brain surface (detected by a characteristic noise upon arrival of the electrode tip to the ventral cortical surface). Cell pairs were grouped according to the distance of recording sites in the three-dimensional space: same recording site, < 0.5 mm, 0.5 ~ 1 mm, 1 ~ 1.5 mm, and so on. Cell pairs were excluded from the analysis when one of the cells was located in the lower bank of STS and the other one in the convexity of IT.

The similarity of the cells' responses was quantified by the coefficient of correlation between mean responses to individual stimuli (stimulus correlation, Fig. 11, left column) or between mean responses to the lowest-level categories (category correlation, Fig 11, right column). The mean response to a category was obtained by averaging mean responses to individual stimuli in the category.

Behavioral experiment

A third monkey was used for a preliminary behavioral examination of the monkey's perceptual categories (Figs. 12 and S4). Prior to the experiment, the monkey was trained extensively for more than a year on a serial delayed matching-to-sample task with stimuli that were not included in our stimulus set. For the behavioral test, we selected 44 stimuli from the 1084 in our set. The monkey, sitting comfortably in a monkey chair without head–restraint, started each trial by pressing a lever. A fixation point appeared for 400 ms at the center of the display, followed by a sample stimulus that was shown for 300-500 ms. After a 1000-1300 ms delay, a test stimulus appeared for 300-500 ms. If the test stimulus was identical to the sample the monkey had to release the lever, within 700 ms of the test stimulus onset, to obtain reward. If the test stimulus did not match the sample, the monkey had to keep pressing the lever until the appearance of a third stimulus, which was always identical to the sample. The monkey was rewarded only based on the identity of the stimuli, not their membership in a particular category. We gradually introduced the 44 stimuli to the monkey and started the data collection when the mean performance reached 80%. During the data collection, we adjusted the length of the stimulus presentation and delay periods, within the specified ranges, to keep the monkey's performance at 80-85%. Because the focus of our analysis was on non-match trials (see below), such trials were presented slightly more frequently (50%-65%, average: 55%).

We calculated the probability of correct discrimination for each non-match stimulus pair, as an estimate of their perceptual distance for the monkey (Sands et. 1982). There were 13-48 repetitions (mean, 29.7) for each stimulus pair in the data set. Kruskal's non-metric multidimensional scaling was performed on the resulting discrimination matrix (Fig. S4); it is equivalent to running MDS on the confusion matrix (Sands et al. 1982). Note that, our choice of a small subset (n=44) of the stimuli is dictated by practical limitations. Testing all possible non-match pairs of the 1084 stimuli would require several years of data collection.

Results

We recorded activity from 674 neurons, in multiple data-collection sessions, in the anterior IT cortex of two macaque monkeys (Fig. 1) while the monkeys performed a fixation task. Responses of each neuron were examined with more than 1000 colorful photographs and paintings of natural and artificial objects. The monkeys had not been trained for any categorization task previously.

We will first explain that response patterns distributed over the IT cell population represented our intuitive category structure. Then, the distributed nature of the category representation will be shown.

IT response patterns form category clusters

Different cells in IT cortex increased or decreased their firing rate to different stimuli, so that each stimulus elicited a particular pattern of response over the population of recorded cells. The response pattern is defined by the set of response magnitudes in the 674 cells. Stimuli that are closer to each other in our hierarchical category structure elicited more similar response patterns. This tendency is illustrated by the scatter plots in Fig. 2 for three exemplar stimulus pairs. For each pair of stimuli, the similarity of population response patterns was measured by Pearson's correlation coefficient (r) of normalized response patterns evoked by the two stimuli (see Materials and Methods). The coefficient was 0.35, 0.20 and –0.20 for the pairs shown in Figs. 2A, B and C, respectively. The correlation coefficients varied from -0.31 to 0.54 across the stimulus set (Fig. 3). Generally, animate and inanimate objects evoked negatively correlated responses while animate objects evoked positively correlated responses. Within the group of animate objects, the highest correlations belonged to stimuli in the same intuitive category.

To visualize the relationship between the similarity of stimuli in our intuitive category structure and the similarity of the neural response patterns, we first used a multi-dimensional scaling analysis (MDS). MDS has been used to infer the internal representation of stimuli based on neuronal responses or behavioral data (Cutzu and Edelman 1998; Hasselmo et al. 1989; Op de Beeck et al. 2001; Sugihara et al. 1998). MDS allows us to generate a low-dimensional layout of the stimuli based on the similarity of response patterns. We used $1- r$ as a measure of distance between two stimuli (neural distance). Using a correlation coefficient has the advantage of focusing on the population patterns of responses rather than nonspecific response changes. The 1084 stimuli were plotted in a low-dimensional space with inter-stimulus distances approximating the original neural distances.

The stimuli were roughly divided into four category clusters — faces, bodies, hands, and inanimate objects — which can be appreciated even in a two-dimensional (2D) projection of the space (Fig. 4A). Each of these clusters was further divided into smaller groups in other projections: faces were divided into human, monkey, and non-primate animal faces (Fig. 4B), and bodies were also divided into several subgroups (Fig. 4C). The Scree plot (the percentage of unexplained variance plotted against the number of dimensions) indicates that 2D projections of the space can explain only 35% or less of the variance in the data (Fig. S1). Therefore, it is important to note that each 2D map captured only a small part of the structure that appeared in a higher dimensional space.

IT categories resemble human- intuitive categories

To better understand the organization of the objects in the high dimensional space and to further examine the category structure reconstructed from the neural distances, we conducted an agglomerative cluster analysis of the neural distances. The analysis started with 1084 nodes corresponding to 1084 stimuli that were consistently used in the experiments. The nodes were connected to each other step by step to make larger nodes. In each step, the two nodes with the smallest distance were connected to make a new node, and all the stimuli were connected to a single node after 1083 steps. The whole reconstructed tree is shown as a Supplementary Program. A one-dimensional alignment of the stimuli based on the tree is also shown in Fig. S2. The tree shows several levels of organization. In the first branching, the stimulus set was divided into animate and inanimate object groups. The animate object group was further divided into several meaningful categories, as expected from the MDS analysis.

To objectively determine which categories appeared in the tree, we listed 23 intuitive categories, based on human convention, (see Table 1) that had at least 12 category members in the stimulus set and examined whether they and their combinations had significantly corresponding nodes in the tree. For each category, we calculated two ratios for each node in the tree: the fraction of category members that were located under the node (Ratio 1), and the fraction of stimuli under the node that were members of the category (Ratio 2). We then selected the node that gave the maximum averaged value of the two ratios. The match was regarded as significant if (a) the value exceeded by 4 s.d. the chance value calculated by a Monte Carlo method for randomly selected stimulus groups of the same size as the category, and (b) more than half of the category members were under the node (Ratio 1 > 0.5).

Many animate object categories at several hierarchical levels had significantly matching nodes (Fig. 5 and Tables 1 and 2). The tree also reconstructed positional relations among the animate object categories in our intuitive category structure: the global category of animate objects was divided into bodies, hands and faces, and the categories of bodies and faces were divided into meaningful subcategories. Faces were divided into primate and non-primate faces, and the primate face group was divided into human and monkey faces. For bodies, human, birds and four-limb animals clustered together, while lower animals such as fish, reptile and insects made their own cluster. Thus, the cluster analysis formally showed that the similarity of population response patterns reconstructed a large part of our intuitive category structure, including the global division into animate and inanimate objects, as well as further hierarchical subdivisions of animate objects. However, with the exception of cars, there were no nodes matching categories of inanimate objects (Table 1). Importantly, unlike imaging and psychological studies which indicate the representation of manmade object classes such as tools in the temporal cortex of humans (Chao et al. 1999; Martin et al. 1996; Moore and Price 1999; Tranel et al. 1997), such categories were not represented in the monkey's IT. The lack of representation for inanimate categories is consistent with the lack of relevance of such categories for the monkey, and magnifies the significance of the represented animate categories.

The tree in Fig. 5 was reconstructed based on responses of all 674 cells recorded from the two monkeys. The trees that were constructed for individual monkeys showed all the basic properties seen here: the first division into animate and inanimate objects, subdivision of animate objects into faces and bodies, and further subdivision of faces and bodies into subcategories. The scores for categories, listed in

Table 1 for the combined data, were well correlated between the two monkeys ($r = 0.8$, $p<10^{-6}$). Hereafter, the 13 categories located at the lowest-level in the tree of Fig. 5 will be referred to as "the lowest-level categories."

Low-level features cannot account for the categorical structure

To test whether the reconstruction of category structure was due to similarity of stimuli in low-level features, we applied the cluster analysis to low-level physical similarity (e.g. color) of the stimuli or similarity in responses of modeled V1 cell population. The modeled V1 cells had size and orientation selectivity, and some of them were color selective.

The trees that were built based on these similarity measures failed to show the categories (Fig. 6A). Therefore, the representation of category structure in the population responses of IT cells appears to be a result of visual information processing after V1.

Randomly selected complex features cannot account for the categorical structure

We then examined whether the representation of category structure emerges trivially in a population of units tuned to various, but randomly selected, complex features. We tested this possibility by creating 674 shape-tuned units (STUs) based on the hierarchical object recognition model of Riesenhuber and Poggio (1999). Each STU was tuned to a randomly selected stimulus, from our stimulus set, through adjusting its input connections from V4-like units (C2 units) in the model. The broadness of tuning of STUs was adjusted to match that of actual IT cells. Hence, for individual STUs, the sparseness of responses and the information about the identity of stimuli were comparable to those of actual IT cells. We did not necessarily regard this model as the best model to simulate the real monkey IT; we used it only to create 674 model units tuned to complex images.

The population response patterns of STUs did not show any meaningful grouping of stimuli, as demonstrated by the mixed distribution of categories in a 2D stimulus projection based on the MDS analysis (Fig. 6B). The tree reconstructed from the responses of the STUs also failed to represent the categories (Fig. 6A, rightmost set of bars). Similar results were obtained for the C2 population in the model. Because of the position and size invariance of C2 units and STUs the failure cannot be attributed to the lack of invariance. These results suggest that the reconstructed object category structure based on the responses of actual IT cells reflects something about the monkey IT cortex beyond the representation of a randomly selected set of complex features.

Category membership can be read out by means of a linear classifier

Different categories elicited easily separable population response patterns in IT cortex. Full classification of the stimuli into different categories in the tree can be performed reliably, even by a linear classifier. We used a simple two-layer perceptron network (Duda et al. 2001) with 674 cells in the input layer and the significant lowest-level categories in the output layer. The magnitudes of mean responses to each stimulus were introduced to the input units, and the output unit yielding the largest value was taken as the classification result. The network was trained by adjusting its connection weights until it achieved a

perfect classification for a training data set.

After training with a random selection of 50% of the stimuli, the network correctly classified 86±3% (mean±s.d.) of the remaining stimuli. This performance is not simply the result of a large degree of freedom in the parameters of linear classifier or a result of the high dimensionality of the response space. When stimuli were randomly assigned to ten groups of the same sizes as the ten categories, the performance of the linear classifier decreased to 50±3% (the 50% chance performance is because one group corresponding to the "other inanimate objects" included about a half of the stimuli (Table 1)).

Single-cell responses are less clearly categorical than population responses

To examine properties of the category representation in IT cell population, we determined the selectivity of individual cells to the lowest-level categories or their combinations. Ten of the lowest-level categories were significant in the tree of Fig. 5 and the rest were added to indicate the category combinations significantly matching the higher nodes.

Some cells (184/674) discriminated stimuli in one of the categories from those in any of the other categories: responses to the category were significantly larger than responses to any other category (Newman-Keuls post-hoc, $p<0.05$) (Table 3). Figure 7 illustrates responses of a cell that responded selectively to human bodies (Fig. 7A), and another cell that responded selectively to bodies of four-limb animals (Fig. 7B).

In addition to selectivity for the lowest-level categories, many cells discriminated a combination of categories from others (145 cells, 186 combinations): responses to any category within the combination were significantly larger than responses to any of the categories outside the combination ($p<0.05$). Many of these combinations (67/186) matched significant higher-level nodes in the tree shown in Fig. 5 (Table 3), although these nodes were only a very small fraction (1.3%) of all possible combinations of the lowest-level nodes. Figure 7C shows responses of a cell that selectively responded to bodies of humans, birds and four-limb animals.

In summary, a total of 255 cells (38% of 674 cells) were selective to a category or a combination of categories. The number is smaller than the sum of the two abovementioned category-selective groups (184+145) because 74 cells showed selectivity to both a single category and a combination of categories (as did the cell in Fig. 7B).

The categorical selectivity of single cells was imperfect compared with that of the cell population as a whole. For single cells, the distribution of response magnitudes for the preferred category (or category combination) largely overlapped with the distribution of responses to other categories (Fig. 7, rightmost column for the example cells, and Fig. 8A, left). This was true even for the cells selective to the face categories (Fig. 8B, left), although a previous study (Tsao et al. 2006) has shown that there may be a cluster of cells in more posterior parts of monkey IT cortex with near perfect face-selectivity. Unlike this overlap in single cell responses, when responses to individual stimuli were averaged over 10–20 categorical cells preferring the same category in each monkey (tested for monkey faces, human faces, human bodies, or hands), the overlap largely disappeared (Fig. 8, right column). These results suggest that

partial deficits in categorical selectivity of individual cells can be compensated by averaging responses over a small population of cells preferring the same category.

Responses to suboptimal categories contribute to the category representation

Another aspect of the distributed nature of category representation in IT cortex is significant difference of neural responses even for categories that a neuron is not maximally responsive to. For each cell, we compared the magnitude of responses (Wilcoxon test) for all pairs of the lowest-level categories after ranking the categories according to the cell's mean category responses. The proportion of cells with significant difference ($p<0.05$) for each rank pair is shown for the 255 category-selective cells and for the other cells in Fig. 9. Significant differences were widely distributed over different rank combinations. For example, more than 50% of cells showed significantly different responses for a rank difference of five between categories among the category-selective cells (Fig. 9A). This high probability of significant difference implies that not only responses to the best categories but also those to other categories, ranging from suboptimal to the worst, carry information. For the cells without sharp category selectivity, a rank difference of eight was enough to achieve significant difference in 50% of cells (Fig. 9B).

The presence of categorical information in responses to suboptimal categories was also demonstrated by examining the correlations in the population response patterns to each category. In this analysis each lowest-level category was randomly divided into two groups of equal size, responses of each cell to individual stimuli were averaged across stimuli in each half-category, and the correlation of averaged responses were calculated for all possible pairs of categories across the cell population. The response to a half-category was more similar to, or more correlated with, responses to the other half of the same category than another category (white bars in Fig. 10). This was true even after removing the cells that maximally responded to either of the categories in the pair (gray bars in Fig. 10), meaning that a category can be potentially discriminated from another one even in the absence of maximally responsive cells. For example, images of four-limb animal bodies were discriminated from images of fish even without the cells selective to either category. As a complementary test, when we shuffled the mean responses of cells with suboptimal responses to the two categories without shuffling the responses of maximally-responsive cells, the strength of correlation was significantly reduced ($p=0.0002$, black bars in Fig. 10).

Cells with similar category selectivity make multiple small clusters in IT cortex

Cells located close to each other in the IT cortex tended to have similar stimulus and category selectivity. This similarity can be quantified by the correlation of responses of the two cells. We measured the correlation either for mean responses to individual stimuli (stimulus correlation) or for mean responses to the lowest-level categories (category correlation). The average magnitude of stimulus correlation was higher for pairs with < 1 mm distance (Fig. 11, left), consistent with the previous finding of local clustering of cells with similar stimulus selectivity (Fujita et al. 1992; Tamura et al. 2005; Tsunoda et al. 2001; Wang et al. 1996, 1998; Yamane et al. 2006). The category correlation had a similar tendency, but showed stronger correlation values (Fig. 11, right) compared to stimulus correlations.

Cells with similar category selectivity were usually found in the same penetration, and clusters of cells in neighboring penetration sites (1 mm interval on our grid system) were rarely selective to the same category. This is reflected in the small change in the category correlations beyond 1 mm in Fig. 11. Instead of creating a big spatial cluster, cells with similar category selectivity appeared in multiple small clusters distributed over the recorded region, as shown for the global face category and global body category in Fig. S3. This may be another aspect of the distributed representation of object categories in the IT cortex. However, the number of cells recorded in each hemisphere (322 cells in the first monkey and 352 cells in the second one) was not large enough to let us draw strong conclusions about the topography of categorical representation in IT. It is also important to note that our analysis in Fig. 11 measured the spatial extent of clusters by making the assumption that such clusters were spherical. It is possible that the topography of IT consists of non-spherical neural clusters with similar category selectivity that extends in one spatial dimension for distances larger than 1mm. A finer and more extensive sampling of IT cortex is required to test this possibility.

Behavioral object confusion reflects IT category structure

The monkey's perceptual categories were examined by analyzing the probability of confusion between stimuli for a third monkey in a delayed matching-to-sample task. The task required discrimination of stimulus pairs based on their identity. We chose 44 stimuli from the stimulus set that was used in the recording experiments. The frequency of the erroneous responses (confusion) in this task can be influenced by the categorical similarity between stimuli (Sands et al. 1982): stimuli belonging to the same category or to closely related categories can be more often confused with each other.

The pattern of confusions between the stimuli corroborates the results based on IT responses. The 44 stimuli included 11 categories at the lowest level of the tree in Fig. 5; human faces, monkey faces, non-primate animal faces, four-limb-animal bodies, bird bodies, fish, insects, reptiles, and other inanimate objects. Each category had 4 stimuli except the category of other inanimate objects, which had 8 stimuli. The small number of stimuli reflects practical limitations in the length of experiments (see Materials and Methods). We collected 51,806 trials in 27 sessions of data collection. Figure 12 shows the neural distances between stimulus pairs plotted against the probability of correct discrimination in non-match trials. Note that the probability of correct discrimination equals 1-(probability of confusion). The probability of a correct discrimination was larger for stimulus pairs that elicited more distinct response patterns in IT (r=0.44, $p<10^{-6}$).

Using the MDS analysis, we projected the 44 stimuli on a two-dimensional space in Fig. S4 according to the correct discrimination rates. In the figure, the faces are clustered in the left bottom corner, the inanimate objects in the left upper corner, and the bodies in the remaining region. The stimuli in each of the lowest-level categories occupied a smaller region compared to the region occupied by the higher-level face or body categories.

Discussion

Representation of categories and category structure

We found that the similarity of population response patterns evoked by object images in the monkey IT cortex was correlated with the distance between their categories in the intuitive category structure. The images of objects selected from the same category tended to evoke similar response patterns, whereas those of objects belonging to more distant categories evoked disparate patterns. The response pattern here is defined by the distribution of magnitudes of responses over the cell population. The correlation between the category distance and the response pattern similarity was strong enough for us to find the category structures in the distribution of stimuli plotted in a low-dimensional space according to the degree of similarity of the response patterns (Fig. 4). The reconstruction of category structure was verified by an agglomerative clustering analysis (Figs. 5 and S2, and Supplementary Program). The match between the category and stimuli under the node was as large as 0.86 on average for the significant lowest-level categories in Fig. 5 (Table 1). The information about categories can be easily read out from the activity of IT cells. For example, when a linear classifier, in which the 674 IT cells were connected to ten output units with different connection weights, was trained with a half of the stimuli, it classified the remaining object images with nearly 90% accuracy.

Hung et al (2005) have recently shown that membership of stimuli in a set of predefined categories can be extracted from responses of a population of IT cells. However, the small size and stimulus homogeneity of the categories, and more importantly, the pose of a predefined set of categories leave it unclear whether an inherently categorical representation of objects exists in IT. We provide the evidence for such a categorical representation for animate objects and show that responses of a population of IT cells represent both the individual categories and the intuitive relationship of the categories.

It is important to note that the MDS and cluster analyses revealing the grouping and cluster tree were data-driven in that they did not require prior specification of any category structure. The intuitive category structure was only used post-hoc, for quantitative assessment of the similarity between the IT stimulus clusters and our intuitive categories. The processes to make the MDS maps and the tree were completely data-driven. When we view the arrangements, we immediately notice that the clustering of objects appear to match with our intuitive categories. To quantify this, we listed 23 intuitive categories existing in the stimulus set with at least 12 category members, and examined the match of the categories and their higher categories with the nodes in the tree. Categories at several hierarchical levels, especially those of animate objects, had significantly matching nodes. We conclude that response patterns over a large population of the monkey IT cortex reconstruct intuitive object categories and their structure. We do not intend to assume that the monkeys had all of the categories identified in the tree. It might be unlikely that the monkeys had the category of cars, for example. However, Sands et al. (Sands et al. 1982) showed evidence suggesting that monkeys have the category of human faces and that of monkey faces, discriminating them from each other and also combining them as a higher category and discriminating human and monkey faces from other objects. Our own preliminary behavioral test performed on a monkey with a subset of the stimuli (n=44) demonstrates that the distance of the stimulus pairs in the response

space of IT is correlated significantly with the probability of confusing the stimuli in a delayed match to sample task (Fig. 12). The behavioral test also suggests that monkeys have a category structure similar to that of humans for animate objects (Fig. S4). These results indicate that the category structure of monkeys correlate, at least partly, with that of humans.

Distributed nature of the representation

Although about 40% of IT cells showed significantly larger responses to stimuli in one category (or a combination of categories) than to stimuli in any of the remaining categories, the distribution of responses of each cell to the preferred category and other categories overlapped substantially (Fig. 8, left). The information that each cell carried about object categories was limited. The overlap in the magnitude of responses largely disappeared when responses to individual stimuli were averaged over 10–20 cells preferring the same category (Fig. 8, right). The averaging was effective because the mismatch between the cells' stimulus selectivity profile and the preferred category was different in different cells. In other words, cells that were selective to a category complemented each other. Pooling of the responses to individual stimuli among cells with similar category selectivity, therefore, increased the information about the category membership of the stimuli (Vogels 1999). Because cells with similar category selectivity also clustered locally in the cortex (Fig. 11), the increase of information could be achieved by pooling the neural responses based on cortical position of the cells.

The results in the present study also indicate that the information about categories is largely distributed over the cell population. Single cells showed significantly different magnitudes of responses between many pairs of non-preferred categories (Fig. 9). For example, some cells that maximally responded to human faces significantly discriminated bird bodies from inanimate objects. Correspondingly, the similarity of response patterns to stimuli in a category was maintained not only by the cells that maximally responded to the category but also by other cells that responded to the category with medium and weak responses (Fig. 10). Such suboptimal category selectivity helped the tree in Fig. 5 capture the appropriate relative positions of categories, e.g., fish were farther from monkey faces than four-limb animal bodies. The suboptimal category selectivity may, thus, underlie the perceptual structure of our hierarchical category system. The suboptimal category selectivity may also help the simultaneous classification of an individual stimulus at multiple category levels. For instance, while a human face is classified into the human face category by the strongest responses in human face cells, it can be classified into the global face category based on submaximal responses in cells tuned to other face categories. Moreover, responses distributed over cells that are tuned to various animate categories can be used to classify the stimulus into the animate object category.

How does the category structure emerge from feature selectivity?

Previous studies have shown that individual IT cells respond to moderately complex features of object images (Desimone et al. 1984; Fujita et al. 1992; Brincat and Connor 2004; Ito et al. 1994; Ito et al. 1995; Kobatake and Tanaka 1994; Tanaka et al. 1991). An important question that arises is whether

responding to a set of moderately complex features by IT cells would automatically result in the representation of the category structure. To examine this possibility, we tuned a population of shape-tuned model units (STUs), which simulate monkey IT cells (Riesenhuber and Poggio 1999), to a set of randomly selected images from the stimulus set used in the present study. Both the MDS and clustering analysis failed to reconstruct the category structure from outputs of the STUs (the rightmost set of bars in Fig. 6A and the distribution in Fig. 6B). This result suggests that the monkey's IT cortex does something more than just respond to a random selection of moderately complex features. There are a huge number of such features that IT cells could potentially be tuned for. However, IT cells do not randomly select their favorite features. Instead, they may select the features that are useful for the purposes of the monkey's behavior. Categorical discrimination of object images may be one of the factors that dictate what features the cells should be tuned for.

The images of objects belonging to the same category or close categories in Fig. 5 may appear more similar to each other than those of objects belonging to distant categories. This intuitive impression, however, has to be more formally defined. The failure of reconstructing the category structure from the similarity of images in low-level features indicates that the similarity of stimuli in low-level features did not underlie the success of reconstruction based on IT response patterns. The failure with the outputs of STUs tuned to randomly selected object images also suggests that the similarity of stimulus images in terms of randomly selected complex features did not underlie the categorical clustering of IT response patterns. Our results suggest that the monkey IT specifically adapts complex features for the purpose of object categorization. The visual system may have found these features, through post-natal experience and possibly through evolutionary processes, and have implemented them in the selectivity of neurons in the IT cortex and its afferent stages (Baker et al. 2002; Sigala and Logothetis 2002; Kobatake et al. 1998; Logothetis et al. 1995; Miyashita 1988; Sakai and Miyashita 1994).

Acknowledgements

This research was supported by a collaborative research grant from RIKEN BSI and a Grant-in-Aid for Scientific Research on Priority Areas - Higher-Order Brain Functions - from MEXT. We thank Mohammad Noorbakhsh for his technical assistance and Michael N. Shadlen, Bharathi Jagadeesh, Adrienne L. Fairhall, Nancy Kanwisher, Barry J. Wark, and Timothy D. Hanks for helpful discussions and comments on earlier versions of the manuscript.

References

Aguirre GK, Zarahn E, and D'Esposito M. An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron* 21: 373-383, 1998.

Allison T, McCarthy G, Nobre A, Puce A, and Belger A. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb Cortex* 4: 544-554, 1994.

Baker CI, Behrmann M, and Olson CR. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neurosci* 5: 1210-1216, 2002.

Brincat SL and Connor CE. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neurosci* 7: 880-886, 2004.

Bruce C, Desimone R, and Gross CG. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46: 369-384, 1981.

Chao LL, Haxby JV, and Martin A. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neurosci* 2: 913-919, 1999.

Cutzu F and Edelman S. Representation of object similarity in human vision: psychophysics and a computational model. *Vis Res* 38: 2229-2257, 1998.

Desimone R, Albright TD, Gross CG, and Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. J Neuroscience 4: 2051-2062, 1984.

Epstein R and Kanwisher N. A cortical representation of the local visual environment. *Nature* 392: 598-601, 1998.

Daubechies I. Ten Lectures on Wavelets. Philadelphia: SIAM, 1992.

Duda RO, Hart PE, and Stork DG. Pattern Classification New York: Wiley, 2001.

Edwards R, Xiao D, Keysers C, Foldiak P, and Perrett D. Color sensitivity of cells responsive to complex stimuli in the temporal cortex. *J Neurophysiol* 90: 1245-1256, 2003.

Epstein R and Kanwisher N. A cortical representation of the local visual environment. *Nature* 392: 598-601, 1998.

Foldiak P, Xiao D, Keysers C, Edwards R, and Perrett DI. Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Prog Brain Res* 144: 107-116, 2004.

Freedman DJ, Riesenhuber M, Poggio T, and Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291: 312-31, 2001.

Freedman DJ, Riesenhuber M, Poggio T, and Miller EK. Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. Journal of Neurophysiology 88: 929-941, 2002.

Freedman DJ, Riesenhuber M, Poggio T, and Miller EK. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23: 5235-5246, 2003.

Fujita I, Tanaka K, Ito M, and Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360: 343-346, 1992.

Gauthier I, Skudlarski P, Gore JC and Anderson AW. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience* 3: 191-197, 2000.

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, and Gore JC. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neurosci* 2: 568-573, 1999.

Hasselmo ME, Rolls ET and Baylis GC. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res* 32: 203-218, 1989.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, and Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293: 2425-2430, 2001.

Hung CP, Kreiman G, Poggio T, and DiCarlo JJ. (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310: 863-866, 2005.

Ito M, Fujita I, Tamura H, and Tanaka K. Processing of contrast polarity of visual images in inferotemporal cortex of the macaque monkey. *Cereb Cortex* 4: 499-508, 1994.

Ito M, Tamura H, Fujita I, and Tanaka K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73: 218-226, 1995.

Johnson SC. Hierarchical clustering schemes. *Psychometrika* 2: 241-254, 1967.
Kanwisher N, McDermott J, and Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302-4311, 1997.
Keysers C, Xiao DK, Foldiak P, and Perrett DI. The speed of sight. *J Cog Neurosci* 13: 90-101, 2001.
Kiani R, Esteky H, and Tanaka K. Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J Neurophysiol* 94: 1587-1596, 2005.
Kovacs G, Vogels R, and Orban GA. Cortical correlate of pattern backward masking. *Proc Nat Acad Sci USA* 92: 5587-5591, 1995.
Kobatake E and Tanaka K. (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856-867, 1994.
Kobatake E, Wang G, and Tanaka K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80: 324-330, 1998.
Kreiman G, Koch C, and Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neurosci* 3: 946-953, 2000.
Lampl I, Ferster D, Poggio T, and Riesenhuber M. Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J Neurophysiol* 92: 2704-2713, 2004.
Li FF, VanRullen R, Koch C, and Perona P. Rapid natural scene categorization in the near absence of attention. *Proc Nat Acad Sci USA* 99: 9596-9601, 2002.
Logothetis NK, Pauls J, and Poggio T. Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5: 552-563, 1995.
Martin A, Wiggs CL, Ungerleider LG, and Haxby JV. Neural correlates of category-specific knowledge. *Nature* 379: 649-652, 1996.
McCarthy G, Puce A, Gore JC, and Allison T. Face-specific processing in the human fusiform gyrus. *J Cog Neurosci* 9: 605-610, 1997.
Miyashita Y. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335: 817-820, 1988.
Moore CJ and Price CJ. A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* 122: 943-962, 1999.
Op de Beeck H, Wagemans J and Vogels R. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neurosci* 4: 1244-1252, 2001.
Perrett DI, Rolls ET, and Caan W. Visual neurones responsive to faces in the monkey temporal cortex. *Expl Brain Re* 47: 329-342, 1982.
Quiroga RQ, Reddy L, Kreiman G, Koch C, and Fried I. Invariant visual representation by single neurons in the human brain. *Nature* 435: 1102-1107, 2005.
Riesenhuber M and Poggio T. Hierarchical models of object recognition in cortex. Nature Neuroscience 2: 1019-1025, 1999.
Rolls ET and Tovee MJ. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc Biol Sci* 257: 9-15, 1994.
Rolls ET and Tovee MJ. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. Journal of Neurophysiology 73: 713-726, 1995.
Sakai K and Miyashita Y. Neuronal tuning to learned complex forms in vision. *Neuroreport* 5: 829-832, 1994.
Sands SF, Lincoln CE, and Wright AA. Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *J Exp Psychol: General* 111: 369-389, 1982.
Sigala N and Logothetis NK. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415: 318-320, 2002.
Sugihara T, Edelman S and Tanaka K. Representation of objective similarity among three-dimensional shapes in the monkey. *Biol Cyber* 78: 1-7, 1998.
Tamura H, Kaneko H, and Fujita I. Quantitative analysis of functional clustering of neurons in the macaque inferior temporal cortex. *Neurosci Res* 52: 311-322, 2005.

Tanaka K, Saito H, Fukada Y, and Moriya M. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66: 170-189, 1991.

Tenenbaum JB, de Silva V, and Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 290: 2319-2323, 2000.

Thorpe S, Fize D, and Marlot C. Speed of processing in the human visual system. *Nature* 381: 520-522, 1996.

Tranel D, Damasio H, and Damasio AR. A neural basis for the retrieval of conceptual knowledge. Neuropsychol 35: 1319-1327, 1997.

Tsao DY, Freiwald WA, Tootell RB, and Livingstone MS. A cortical region consisting entirely of face-selective cells. *Science* 311: 670-674, 2006.

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat. Neurosci. 4: 832-838, 2001.

Vogels R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J Neurosci 11: 1239-1255, 1999.

Wang G, Tanaka K, and Tanifuji M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272: 1665-1668, 1996.

Wang G, Tanifuji M, and Tanaka K. Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neurosci Res* 32: 33-46, 1998.

Worgotter F, Daunicht WJ, and Eckmiller R. An on-line spike form discriminator for extracellular recordings based on analog correlation technique. *J Neurosci Methods* 17: 141-151, 1986.

Yamane Y, Tsunoda K, Mastumoto M, Phillips AN, and Tanifuji M. Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *J. Neurophysiol* 96: 3147-3156, 2006.

Young FW and Hammer RM. Scaling History, Theory and Applications. New York: Erlbaum, 1987.

Figure Legends

Fig. 1. Positions of recording sites in two monkeys. **Left**, lateral views of the recorded hemispheres. Vertical lines indicate the anterior-posterior extent of the recording sites. **Right**, representative coronal sections. Recorded regions are indicated by gray. Recording sites were evenly distributed. ls: lateral sulcus, sts: superior temporal sulcus, amts: anterior middle temporal sulcus, rs: rhinal sulcus.

Fig. 2. Examples of correlation between response patterns evoked in the 674 cells by three pairs of stimuli. Each dot corresponds to one of the cells, and the x- and y-values of each dot represent the normalized responses of the cell to the stimulus pair. The three pairs share a common stimulus which is shown at the left. The Pearson's correlation coefficient (r) was 0.35, 0.20, and –0.20 for A, B and C, respectively. All three correlations are significant.

Fig. 3. Distribution of the correlation coefficients for the population response patterns. For each pair of the 1084 stimuli, the correlation was calculated for the response patterns evoked by the two stimuli across the recorded cells.

Fig. 4. Arrangement of the stimuli in a low-dimensional space based on multidimensional scaling (MDS) on the neural distances ($1 - r$) of the stimuli. Each point represents one of the stimuli. **A**, **B**, and **C** represent three different projections of the space, as denoted by the axis labels. All 1084 stimuli are shown in **A**, while only faces and bodies are shown in **B** and **C**, respectively. The categories that are labeled here were found to have significantly matching nodes in the tree shown in Fig. 5.

Fig. 5. The tree reconstructed based on the neural distances. Red circles indicate the nodes significantly matching the categories. Blue circles indicate the nodes that had scores (see Materials and Methods) significantly larger than chance score, but included fewer than half of the category members. The blue nodes were added to indicate category combinations significantly matching the higher nodes. Five examples of category members are shown for each of the lowest-level categories, except for "other inanimate objects" (the rightmost node). The thirteen categories located at the lowest level are referred to as "the lowest-level categories" throughout this paper.

Fig. 6. Object categories were represented by the responses of IT cell population but not by low-level image similarity or by model unit responses. **A**: Average match of the intuitive categories with the best representative nodes of the trees formed by different distance measures (gray bars). The match was quantified by the node score ((Ratio 1 + Ratio 2)/2), and averaged over all the significant categories of Fig. 5. The white bars show the expected scores for chance-level clustering of stimuli. Error bars represent s.e.m. STUs: shape-tuned units in the HMAX model. These units were tuned to 674 stimuli randomly selected from the stimulus set. **B**: Arrangement of the stimuli in the stimuli set, according to MDS analysis on the response patterns of STUs did not replicate the clustering of stimuli based on the real IT cell population (compare with Fig. 4A).

Fig. 7. Three examples of category-selective cells. Example responses to individual members of the preferred category (left panel), the averaged responses to the lowest-level categories (middle panel), and the magnitude of responses to individual stimuli of the lowest-level categories plotted against the

normalized stimulus rank within each category (right panel) for a cell preferring human bodies (**A**), four-limb animal bodies (**B**), and the combination of human bodies, four-limb animal bodies and bird bodies (**C**). In the left and middle panels, horizontal bars indicate the stimulus presentation period. In the middle and right panels, the best categories are shown in red. The categories that evoked responses significantly smaller than the best category but significantly larger than other categories (bird and reptile in B; reptile, fish, and other insects in C) are shown in blue. Gray indicates other categories. Normalized rank of 1 indicates the stimulus that evoked the largest response within the category.

Fig. 8. The overlap of average response magnitudes of stimuli in the preferred category (black lines) with responses to other stimuli (gray lines) for all the categorical cells (A), and for cells selective to human faces (B). For the left panel of A and B, mean responses to individual stimuli were normalized by the maximum mean response in each cell, and then responses to stimuli of the same normalized rank were averaged across cells. Normalized rank of 1 indicates the largest response in the stimulus group. For the right panel of A, normalized mean responses to the same stimulus were averaged over 10–20 cells preferring the same category in each monkey (performed for monkey faces, human faces, human bodies, or hands). The resulting magnitude-rank curves were then averaged across categories and monkeys for the figure. Right panel of B is similar to A but the normalized mean responses to individual stimuli were averaged among 11 or 20 cells selective to human faces in each monkey. Monkey faces and non-primate faces were excluded in B to allow comparison with previous studies.

Fig. 9. Many IT cells showed significant differences in their responses to suboptimal categories. For each cell, the lowest-level categories were ranked based on the average response magnitude, and the significance of difference in response magnitudes was calculated for each pair of category ranks (Wilcoxon test, significance defined as $p<0.05$). Individual trial responses pooled for all the stimuli belonging to each category were used for the comparison. The proportion of cells that showed a significant difference for each category-rank pair is color-coded for the 255 category-selective cells (**A**) and the remaining 419 cells (**B**).

Fig. 10. Within-category and between-category correlations of population response patterns. Each lowest-level category was randomly divided into two halves, and mean responses of every cell to each half were calculated. The correlation of the mean responses across the population of cells was calculated for all possible pairs of categories (n=91). The procedure was repeated 1000 times with different random divisions of the categories, and the mean value of correlation coefficient was obtained for each of the category pairs. White bars indicate the correlations calculated over the 674 cells. Gray bars indicate the correlation coefficient calculated after excluding the cells maximally responding to either of the paired categories. Black bars indicate the correlation coefficient for all the cells but after shuffling of the responses for the cells that did not respond maximally to either of the two categories. Error bars represent 95% confidence interval. The figure is symmetrical around the diagonal; all the bars are shown for convenient visual comparison. HF: human face, MF: monkey face, AF: animal face, HD: hand, HB: human body, 4L: four-limb animal, BD: bird, RP: reptile, FS: fish, BF: butterfly insects, IS: other insects, CA: car, OB: other inanimate objects.

Fig. 11. Response correlations for pairs of cells with various distances from each other. The Pearson's correlation coefficients were calculated based on mean responses to the 1084 individual stimuli (left), or based on averaged mean responses to the lowest-level categories. Distances between recording sites were divided into 11 bins, and the correlation coefficient was averaged over cell pairs within each distance bin. The averaging was performed separately in each monkey. Error bars represent s.e.m. Note that unlike the neural distance, which was based on the response correlations for stimulus pairs across the neural population, this analysis is based on the response correlation for cell pairs across the stimulus set.

Fig. 12. Probability of correct discrimination of stimulus pairs in a delayed matching–to-sample task plotted against the neural distance of stimulus pairs. A third monkey performed the task with 44 stimuli selected from the stimulus set.

Table 1. Degrees of match between intuitive object categories and nodes in the tree reconstructed from responses of IT cells

	Ratio1	Ratio2	Score	Chance	# stimuli
Animal face	0.38	0.94	0.66*	0.52	42
Monkey face*	0.97	0.81	0.89*	0.52	39
Human face*	0.97	0.98	0.98*	0.53	64
Hand*	0.93	1.00	0.96*	0.52	27
Bird body	0.16	1.00	0.58*	0.53	56
4-limb animal body*	0.57	0.88	0.73*	0.55	103
Human body*	0.95	0.93	0.94*	0.52	40
Butterfly*	0.53	1.00	0.76*	0.51	17
Other insects	0.19	1.00	0.59*	0.52	27
Reptile*	0.84	0.41	0.63*	0.52	19
Fish*	0.87	1.00	0.93*	0.51	15
Car*	0.87	0.83	0.85*	0.52	23
Tree	0.15	1.00	0.58*	0.51	13
Leaf	1.00	0.02	0.51	0.51	12
Flower	0.14	1.00	0.57	0.52	22
Fruit	0.12	1.00	0.56	0.51	17
Vegetable	0.11	1.00	0.56	0.51	18
Food	0.07	1.00	0.53*	0.52	44
Furniture	0.16	1.00	0.58*	0.52	25
Lamp	0.19	1.00	0.60*	0.51	21
Common tool	0.11	1.00	0.56	0.52	27
Kitchen utensil	1.00	0.03	0.52	0.51	19
Home appliance	0.15	1.00	0.58*	0.51	13

Ratio1 = (number of category members under the node)/(number of all members in the category)

Ratio2 = (number of category members under the node)/(number of all stimuli under the node)

Score = (Ratio1+Ratio2)/2,

with the asterisk indicating that the value is significantly larger than the chance value

Chance = chance score expected from random clustering of stimuli, estimated by Monte-Carlo simulation

stimuli=number of stimuli belonging to the category

The asterisk after category name indicates that the match is satisfying both of the following criteria: a) the score is significantly larger than the chance value, and b) Ratio 1 > 0.5.

Table 2. Degrees of match between combined categories and nodes in the tree reconstructed from responses of IT cells

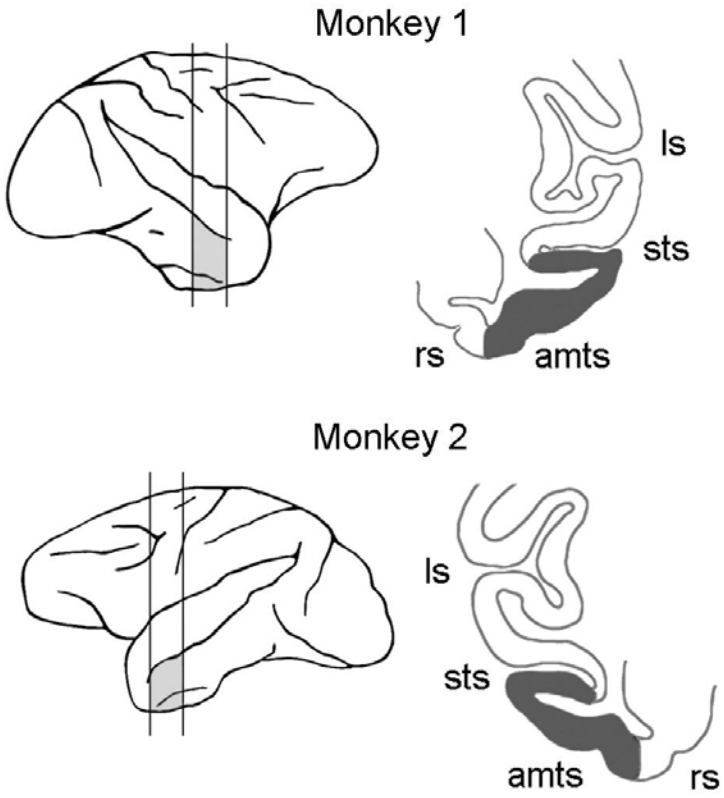
	Ratio1	Ratio2	Score	Chance
Primate face*	0.97	0.91	0.94*	0.55
Face*	0.86	0.89	0.87*	0.57
Face+Hand*	0.91	0.75	0.83*	0.58
4-limb animal+Bird body*	0.83	0.83	0.83*	0.57
4-limb animal+Bird+Human body*	0.87	0.85	0.86*	0.59
Insect	0.41	0.90	0.65*	0.52
Fish+Reptile*	0.85	0.74	0.80*	0.52
Insect+Fish+Reptile*	0.71	0.85	0.78*	0.54
Body*	0.86	0.90	0.88*	0.64
Animate*	0.94	0.90	0.92*	0.72
Inanimate*	0.94	0.91	0.92*	0.76

Legends are similar to Table 1.

Table 3. Number and response properties of cells selectively responding to the categories identified in the tree of Fig. 5.

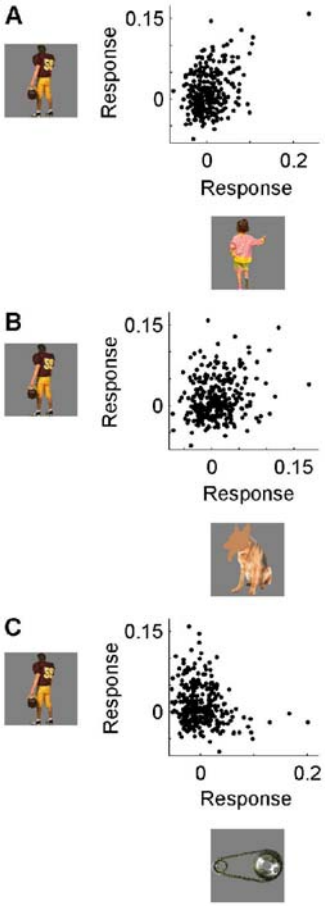
Category	Number of cells	Averaged maximum response (spikes/s)*	Average spontaneous response (spikes/s)
Animal face	13	37.3	6.3
Monkey face	19	32.4	6.3
Human face	54	39.5	7.6
Hand	26	38.8	5.5
Bird	2	51.0	7.8
Four-limb animal	5	32.8	4.7
Human body	36	28.3	4.6
Butterfly	10	53.7	11.5
Other insects	3	49.1	2.8
Reptile	3	40.6	6.1
Fish	5	52.4	6.4
Car	8	28.2	5.6
Other inanimate objects	0		
Primate face	10	40.8	6.1
Face	37	39.3	6.7
Face + Hand	4	62.5	12.1
Four-limb + Bird body	0		
Four-limb + Bird + Human body	5	35.8	8.9
Insect	2	22.9	1.2
Fish + Reptile body	3	41.3	5.0
Insect + Fish + Reptile body	0		
Body	0		
Animate (face+body+hand)	6	49.9	10.3
Inanimate	0		

*The mean firing rate in the 140-ms response window to the best stimulus for individual cells was averaged over the cells that responded selectively to the category.



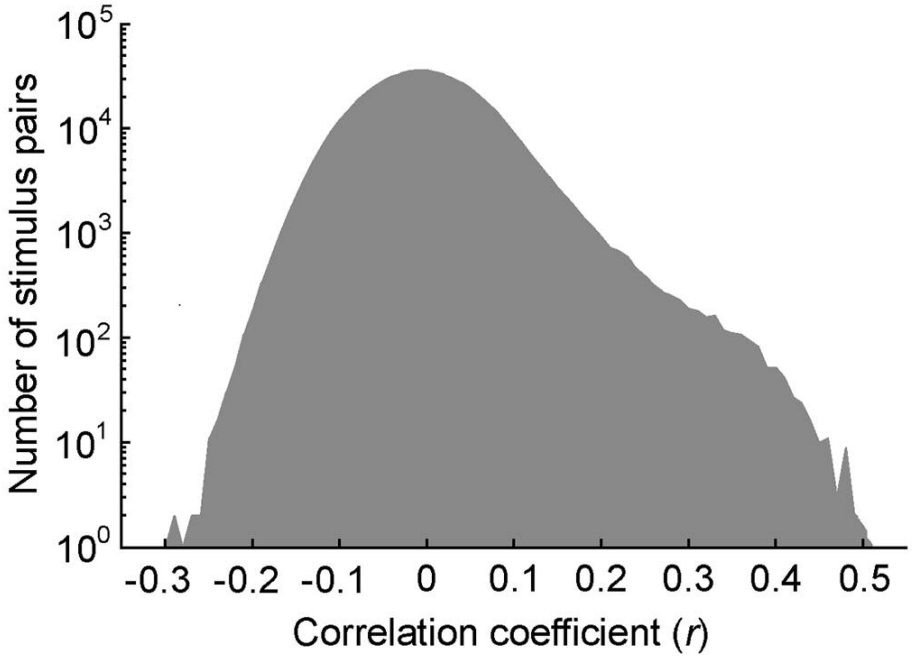
Kiani et al, Fig. 1

Fig. 1. Positions of recording sites in two monkeys. Left, lateral views of the recorded hemispheres. Vertical lines indicate the anterior-posterior extent of the recording sites. Right, representative coronal sections. Recorded regions are indicated by gray. Recording sites were evenly distributed. ls: lateral sulcus, sts: superior temporal sulcus, amts: anterior middle temporal sulcus, rs: rhinal sulcus.



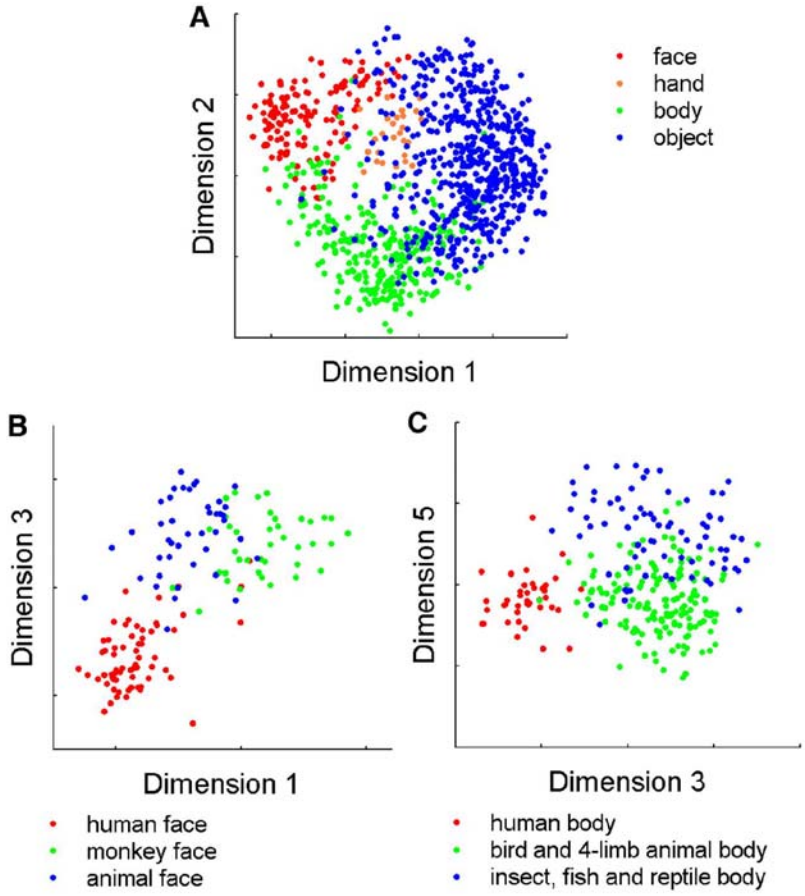
Kiani et al, Fig. 2

Fig. 2. Examples of correlation between response patterns evoked in the 674 cells by three pairs of stimuli. Each dot corresponds to one of the cells, and the x- and y-values of each dot represent the normalized responses of the cell to the stimulus pair. The three pairs share a common stimulus which is shown at the left. The Pearson's correlation coefficient (r) was 0.35, 0.20, and -0.20 for A, B and C, respectively. All three correlations are significant.



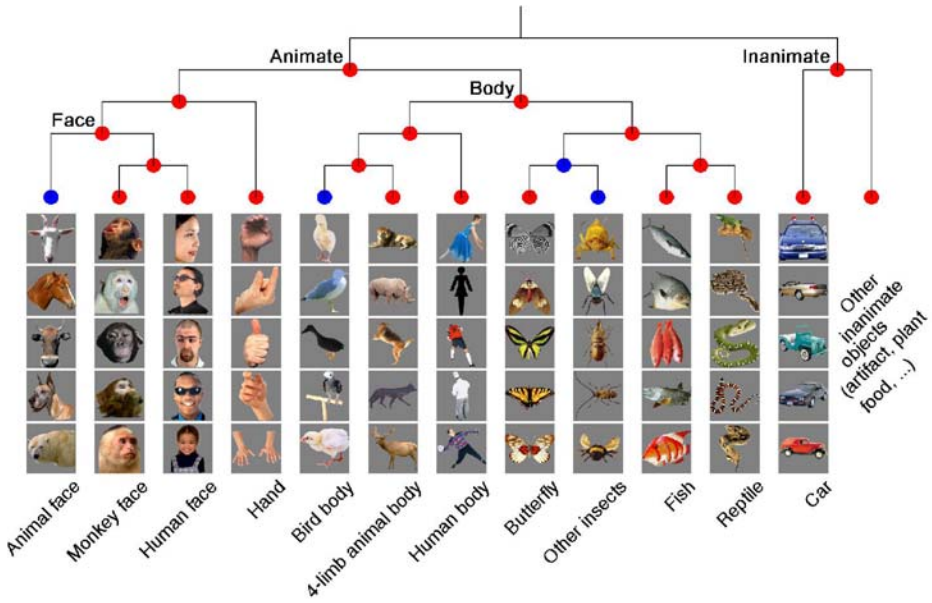
Kiani et al, Fig. 3

Fig. 3. Distribution of the correlation coefficients for the population response patterns. For each pair of the 1084 stimuli, the correlation was calculated for the response patterns evoked by the two stimuli across the recorded cells.



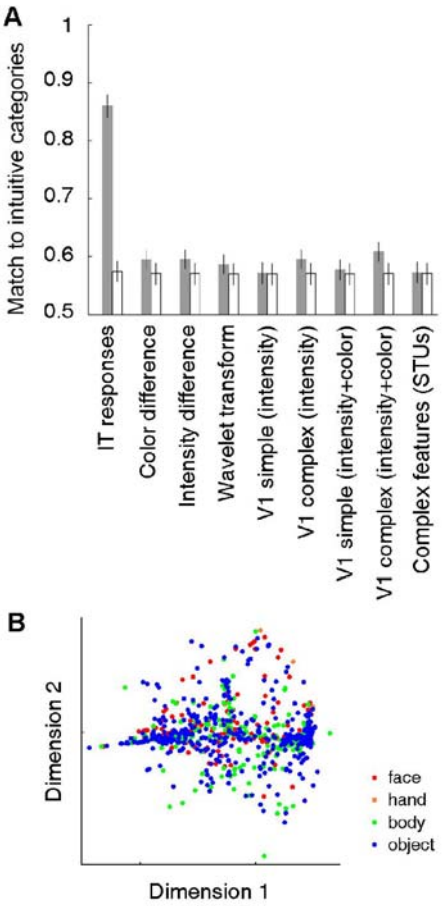
Kiani et al, Fig. 4

Fig. 4. Arrangement of the stimuli in a low-dimensional space based on multidimensional scaling (MDS) on the neural distances (1 - r) of the stimuli. Each point represents one of the stimuli. A, B, and C represent three different projections of the space, as denoted by the axis labels. All 1084 stimuli are shown in A, while only faces and bodies are shown in B and C, respectively. The categories that are labeled here were found to have significantly matching nodes in the tree shown in Fig. 5.



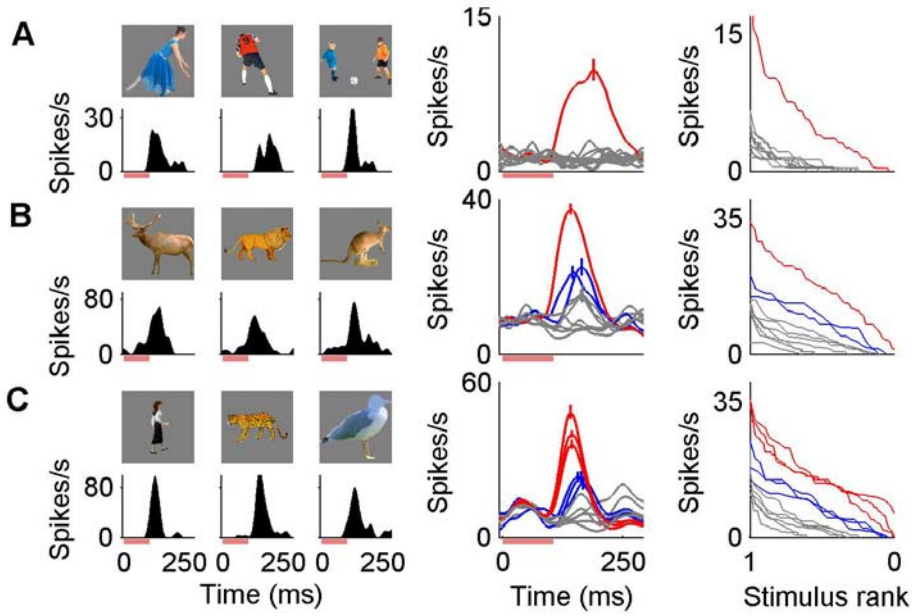
Kiani et al, Fig. 5

Fig. 5. The tree reconstructed based on the neural distances. Red circles indicate the nodes significantly matching the categories. Blue circles indicate the nodes that had scores (see Materials and Methods) significantly larger than chance score, but included fewer than half of the category members. The blue nodes were added to indicate category combinations significantly matching the higher nodes. Five examples of category members are shown for each of the lowest-level categories, except for "other inanimate objects" (the rightmost node). The thirteen categories located at the lowest level are referred to as "the lowest-level categories" throughout this paper.



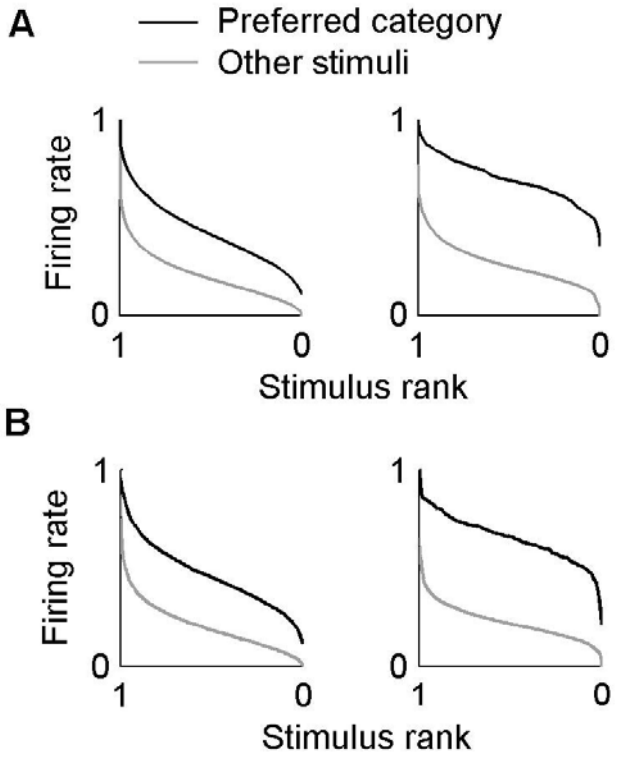
Kiani et al, Fig. 6

Fig. 6. Object categories were represented by the responses of IT cell population but not by low-level image similarity or by model unit responses. A: Average match of the intuitive categories with the best representative nodes of the trees formed by different distance measures (gray bars). The match was quantified by the node score ((Ratio 1 + Ratio 2)/2), and averaged over all the significant categories of Fig. 5. The white bars show the expected scores for chance-level clustering of stimuli. Error bars represent s.e.m. STUs: shape-tuned units in the HMAX model. These units were tuned to 674 stimuli randomly selected from the stimulus set. B: Arrangement of the stimuli in the stimuli set, according to MDS analysis on the response patterns of STUs did not replicate the clustering of stimuli based on the real IT cell population (compare with Fig. 4A).



Kiani et al, Fig. 7

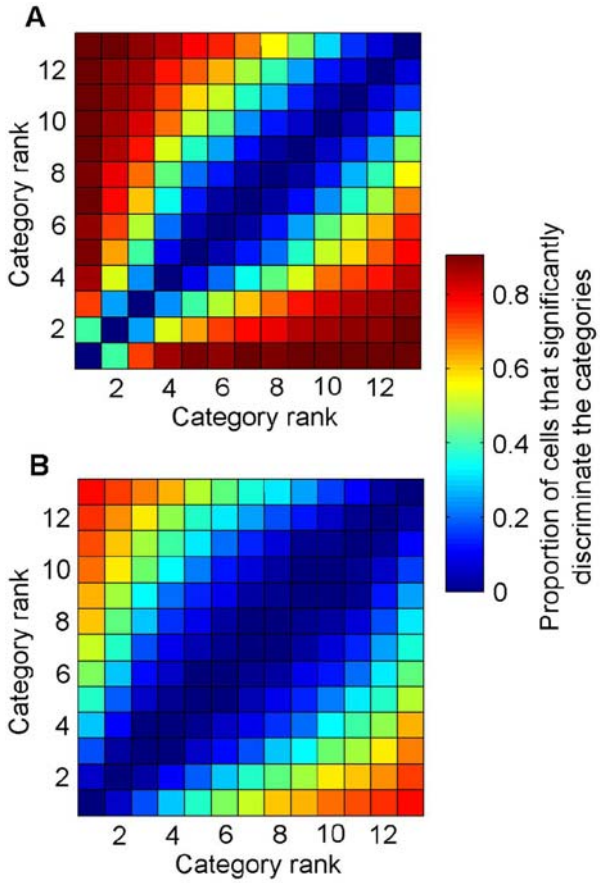
Fig. 7. Three examples of category-selective cells. Example responses to individual members of the preferred category (left panel), the averaged responses to the lowest-level categories (middle panel), and the magnitude of responses to individual stimuli of the lowest-level categories plotted against the normalized stimulus rank within each category (right panel) for a cell preferring human bodies (A), four-limb animal bodies (B), and the combination of human bodies, four-limb animal bodies and bird bodies (C). In the left and middle panels, horizontal bars indicate the stimulus presentation period. In the middle and right panels, the best categories are shown in red. The categories that evoked responses significantly smaller than the best category but significantly larger than other categories (bird and reptile in B; reptile, fish, and other insects in C) are shown in blue. Gray indicates other categories. Normalized rank of 1 indicates the stimulus that evoked the largest response within the category.



Kiani et al, Fig. 8

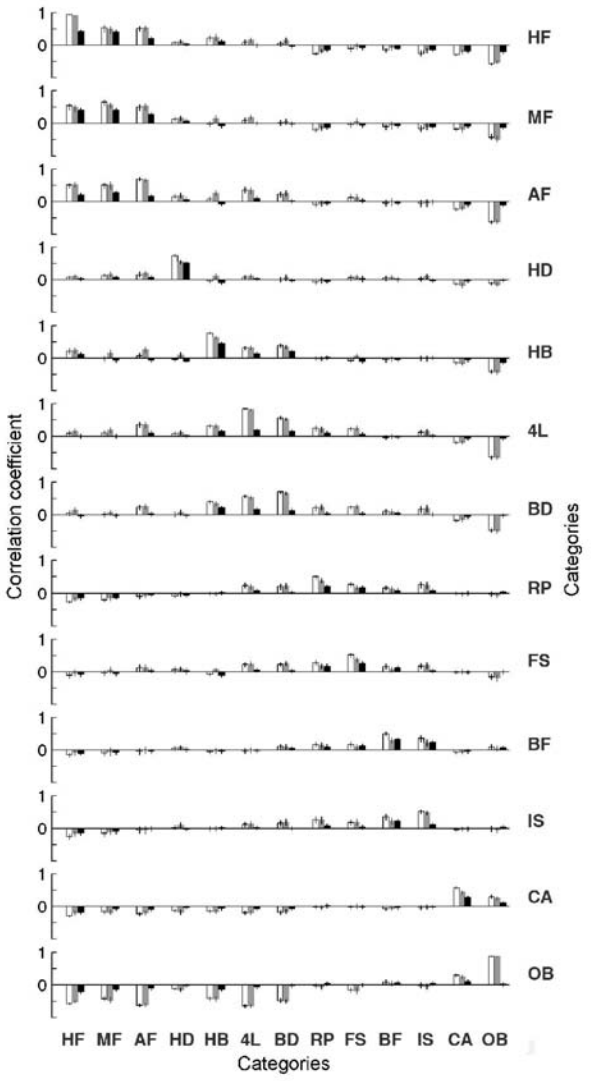
Fig. 8. The overlap of average response magnitudes of stimuli in the preferred category (black lines) with responses to other stimuli (gray lines) for all the categorical cells (A), and for cells selective to human faces (B). For the left panel of A and B, mean responses to individual stimuli were normalized by the maximum mean response in each cell, and then responses to stimuli of the same normalized rank were averaged across cells. Normalized rank of 1 indicates the largest response in the stimulus group. For the right panel of A, normalized mean responses to the same stimulus were averaged over 10-20 cells preferring the same category in each monkey (performed for monkey faces, human faces, human bodies, or hands). The resulting magnitude-rank curves were then averaged across categories and monkeys for the figure. Right panel of B is similar to A but the normalized mean responses to individual stimuli were averaged among 11 or 20 cells selective to human faces in each monkey. Monkey faces and non-primate faces were

excluded in B to allow comparison with previous studies.



Kiani et al, Fig. 9

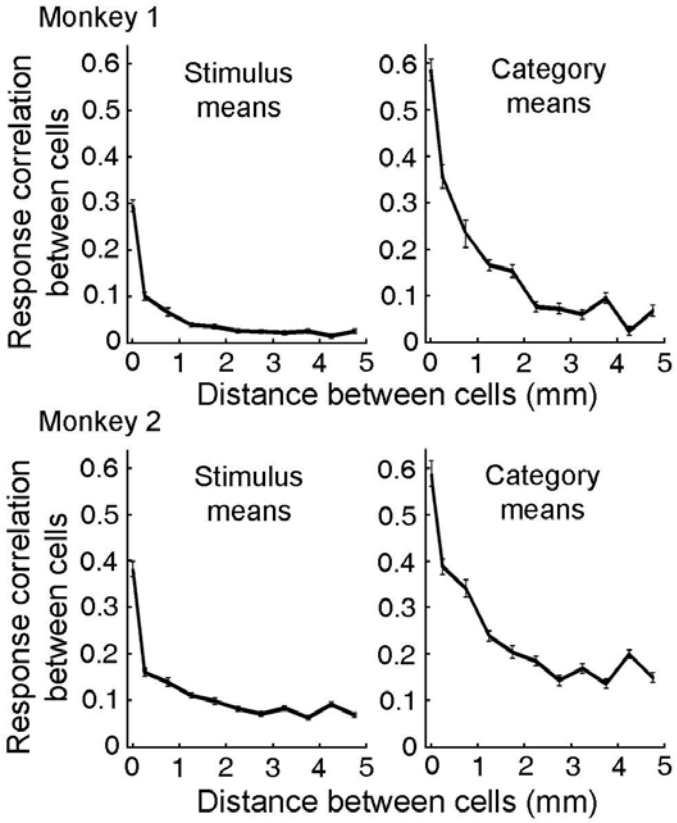
Fig. 9. Many IT cells showed significant differences in their responses to suboptimal categories. For each cell, the lowest-level categories were ranked based on the average response magnitude, and the significance of difference in response magnitudes was calculated for each pair of category ranks (Wilcoxon test, significance defined as $p<0.05$). Individual trial responses pooled for all the stimuli belonging to each category were used for the comparison. The proportion of cells that showed a significant difference for each category-rank pair is color-coded for the 255 category-selective cells (A) and the remaining 419 cells (B).



Kiani et al, Fig. 10

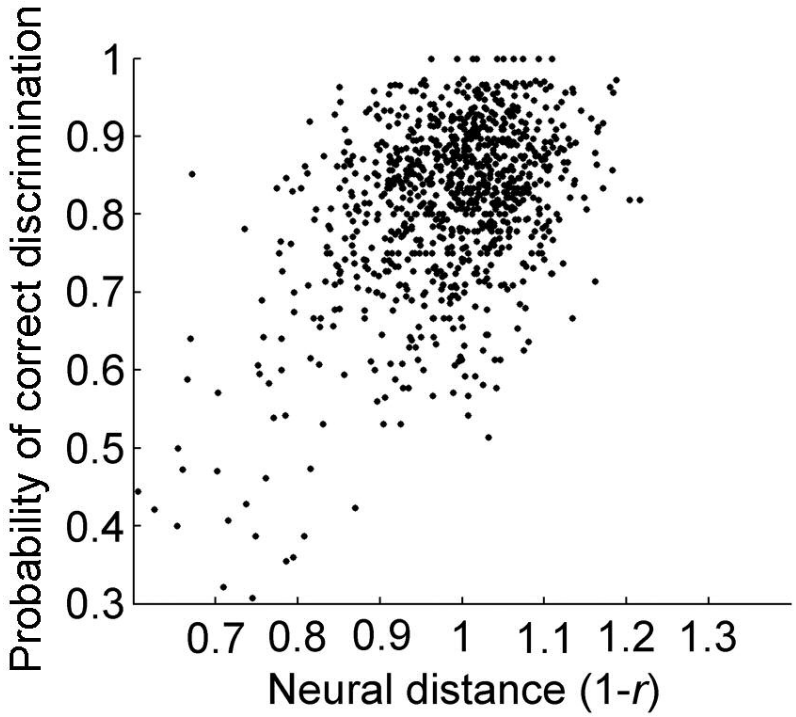
Fig. 10. Within-category and between-category correlations of population response patterns. Each lowest-level category was randomly divided into two halves, and mean responses of every cell to each half were calculated. The correlation of the mean responses across the population of cells was calculated for all possible pairs of categories (n=91). The procedure was repeated 1000 times with different random divisions of the categories, and the mean value of correlation coefficient was obtained for each of the category pairs. White bars indicate the correlations calculated over the 674 cells. Gray bars indicate the correlation coefficient calculated after excluding the cells maximally responding to either of the paired categories. Black bars indicate the correlation coefficient for all the cells but after shuffling of the responses for the cells that did not respond maximally to either of the two categories. Error bars represent 95% confidence interval. The figure is symmetrical around the diagonal; all the bars are shown for

convenient visual comparison. HF: human face, MF: monkey face, AF: animal face, HD: hand, HB: human body, 4L: four-limb animal, BD: bird, RP: reptile, FS: fish, BF: butterfly insects, IS: other insects, CA: car, OB: other inanimate objects.



Kiani et al, Fig. 11

Fig. 11. Response correlations for pairs of cells with various distances from each other. The Pearson's correlation coefficients were calculated based on mean responses to the 1084 individual stimuli (left), or based on averaged mean responses to the lowest-level categories. Distances between recording sites were divided into 11 bins, and the correlation coefficient was averaged over cell pairs within each distance bin. The averaging was performed separately in each monkey. Error bars represent s.e.m. Note that unlike the neural distance, which was based on the response correlations for stimulus pairs across the neural population, this analysis is based on the response correlation for cell pairs across the stimulus set.



Kiani et al, Fig. 12

Fig. 12. Probability of correct discrimination of stimulus pairs in a delayed matching-to-sample task plotted against the neural distance of stimulus pairs. A third monkey performed the task with 44 stimuli selected from the stimulus set.