

Representation of Regular and Irregular Shapes in Macaque Inferotemporal Cortex

Greet Kayaert¹, Irving Biederman² and Rufin Vogels¹

¹Laboratorium voor Neuro- en Psychofysiologie, KU Leuven Medical School, Leuven, Belgium and ²Department of Psychology and Neuroscience Program, University of Southern California, Los Angeles, CA, USA

We determined the degree to which the response modulation of macaque inferior temporal (IT) neurons corresponds to perceptual versus physical shape similarities. IT neurons were tested with four groups of shapes. One group consisted of variations of simple, symmetrical (i.e. regular) shapes that differed in nonaccidental properties (NAPs, i.e. viewpoint-invariant), such as curved versus straight contours. The second and third groups were composed of, respectively, simple and complex asymmetrical (i.e. irregular) shapes, all with curved contours. A fourth group consisted of simple, asymmetrical shapes, but with straight (corners) instead of curved contours. The neural modulations were greater for the shapes differing in NAPs than for the shapes differing in the configuration of the convexities and concavities. Multidimensional scaling showed that a population code of the neural activity could readily distinguish the four shape groups. This pattern of neural modulation was strongly manifested in the results of a sorting task by human subjects but could not be predicted using current image-based models (i.e. pixel energies, V1-like Gabor-jet filtering and HMAX). The representation of shape in IT thus exceeds a mere faithful representation of physical reality, by emphasizing perceptually salient features relevant for essential categorizations.

Keywords: categorization, object recognition, shapes, single cell, symmetry

Introduction

Inferior temporal (IT) cortex is generally believed to be involved in object recognition and categorization (Gross *et al.*, 1972; Dean, 1976; Tanaka, 1996). The shape selectivity of IT units may thus reflect the subject's recognition and categorization demands, particularly in the challenges posed in differentiating objects, and object categories at arbitrary orientations in depth (Vogels, 1999; Sigala and Logothetis, 2002). Indeed, there is greater IT neuronal modulation, and hence sensitivity, to a change in a nonaccidental property (NAP) than a change in a metric property (MP) when these differences are equated according to physical (namely, pixel energy and wavelet) differences (Vogels *et al.*, 2001; Kayaert *et al.*, 2003). NAPs are stimulus properties that are relatively invariant with orientation in depth, such as whether a contour is straight or curved or whether a pair of edges is parallel or not, or whether a part is present or not. These properties can allow efficient object recognition at different orientations in depth not previously experienced, despite drastic image changes at the new orientations (Biederman, 1987; Biederman and Gerhardstein, 1993; Logothetis *et al.*, 1994; Biederman and Bar, 1999). MPs are properties that vary with orientation in depth, such as aspect ratio or the degree of curvature.

Differences in symmetry and simplicity may also furnish highly salient attributes when classifying real world entities,

such as when we readily distinguish a naturally growing, highly irregular bush from a regular one shaped, through topiary, into a cube. Such variation in symmetry and simplicity was addressed by the Gestalt concept of 'Good Figure' (Koffka, 1935). Cooper *et al.* (1995; described in Biederman, 1995) had human subjects judge whether a sequentially presented pair of object images belonged to the same basic-level class, e.g. both hats or both trucks. On the trials when the objects did belong to the same class, they could differ in a NAP of a regular part, such as a base of a lamp being a vertical brick or cylinder, or the shape of that part was irregular and the object could differ in the configuration of that irregular part (Fig. 1). Cooper *et al.* (1995) found a greater sensitivity to the NAP differences in the regular parts compared to the configurational differences in an irregular part, and also a high sensitivity to the regularity of a part, *per se*.

The general aim of this investigation was to assess the degree to which the shape tuning of IT neurons reflects recognition and categorization demands rather than global image similarity measures, such as pixel or wavelet energy.

We measured single-cell modulation in IT to four groups of shapes that included regular and irregular shapes, analogous to those used by Cooper *et al.* (1995). We will adopt, in this investigation, Cooper *et al.*'s definition of regularity in which differences between regular shapes [simple smooth shapes with at least one (curved) axis of symmetry, and a low number of features] covaried with NAP differences. It is clear, however, that simplicity, axis structure and NAPs can be separated operationally. We will also defer for the present investigation whether the ultimate underlying attribute(s) distinguishing 'regular' shapes from 'irregular' shapes, are, in fact, just symmetry and simplicity or some other variable confounded with the manipulations of symmetry and simplicity in the present study, such as smoothness, size and number of features.

The first group of stimuli was composed of regular geometric shapes (shown in the first two columns of Fig. 2 and denoted as R for Regular) that all have at least one axis of symmetry. These shapes are simple, i.e. have low medial axis complexity (Kimia, 2003), and can be regarded as two-dimensional geons or projections of a surface of a three-dimensional geon. Geons are relatively simple generalized cylinders that differ in NAPs and are hypothesized to be the primitives corresponding to the representation of object parts (Biederman, 1987). The stimulus pairs in each row of these two columns differ in a NAP, insofar as a contour in one shape might be curved and in the other shape straight, or a pair of contours might be parallel in one shape and non-parallel in the other.

The three other groups are all termed 'Irregular'. They are generated from Fourier descriptors and differ from the Regular shapes in that they do not have a single axis of symmetry. The

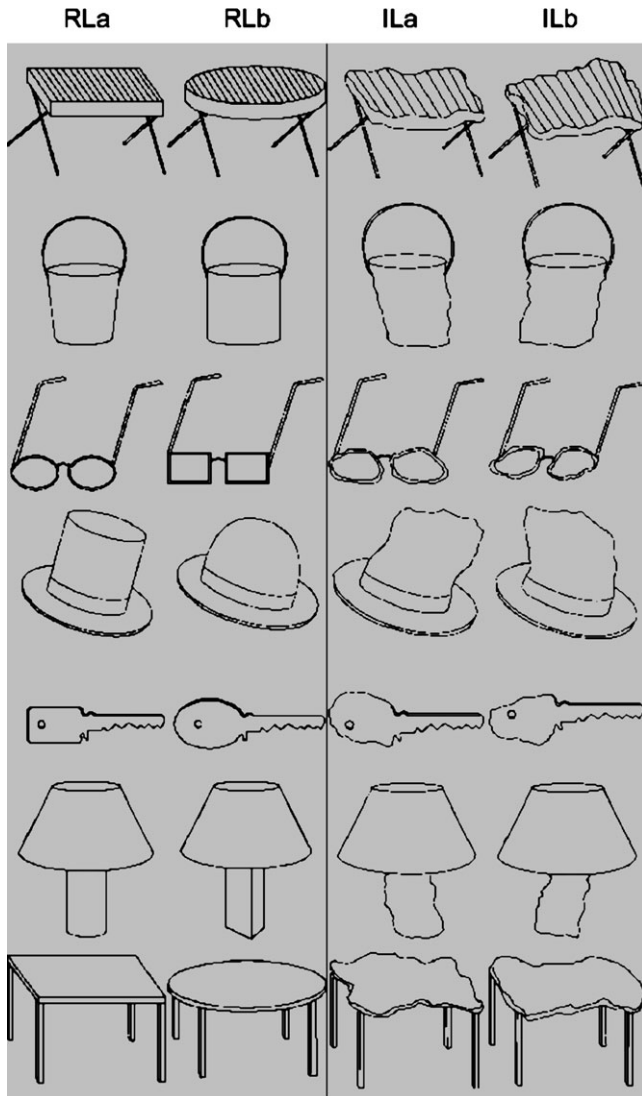


Figure 1. Line drawings of irregularly and regularly shaped objects. The stimuli are adapted from Cooper *et al.* (1995). The seven rows each consist of two wavelet-calibrated (Lades *et al.*, 1993) pairs of objects. RL stands for a line drawing with a regular part, IL for a line drawing with an irregular part. One pair (RLa versus RLb) has only regular parts; the other pair (ILa versus ILb) has one irregular part that differs between the members of a pair.

two shapes in each row of the three Irregular sets differ in the spatial configuration of their concavities and convexities or corners. The shapes in the irregular simple curved (ISC) set all have curved contours. The irregular simple straight (ISS) shapes are derived from the ISC shapes by replacing the curved contours with straight lines. Thus the corresponding stimuli in these two groups, ISC and ISS, differ in a NAP (curved versus straight). Last, the irregular complex group (IC) is more complex in that the shapes in that group have a greater number of contours.

As we will show, human subjects are highly sensitive to the differences among these four groups of stimuli in a sorting task. Would a differentiation between these groups also be manifested in the modulation of IT cells? That is, would there be greater changes in the activity of IT cells when presented with stimuli from different groups than from the same group? Such a separation of these groups would be based not on general

similarity, but on the extraction of salient shape properties such as complexity and straightness of a contour.

We also predict that IT modulation will be relatively larger for some shape difference than for others. On the basis of geon theory (Biederman, 1987) and the results of Cooper *et al.* (1995), larger IT response modulation to these regular shapes differing in a NAP would be expected than for the same degree of configuration change of the irregular shapes. A comparison of the irregular simple and complex shapes allowed us to assess the effect of complexity independent of NAP differences.

Given the theoretical and empirical importance of NAPs (e.g. curved versus straight), the neuronal response modulation *between* the curved versus straight ISC-ISS groups would be expected to be larger than the modulation among the shapes *within* each of the irregular simple groups (Fig. 2). We should note that testing the effect of changes in a NAP using irregular shapes fills an important gap in our understanding of NAP sensitivity as former single cell and psychophysical studies of the sensitivity to nonaccidental variations employed relatively simple, geon-like shapes (Biederman and Bar, 1999; Vogels *et al.*, 2001; Kayaert *et al.*, 2003).

Figure 3 summarizes the different stimulus groups and their shape properties, as well as the nature of the differences between their shapes.

To determine whether the observed neural modulations to these shapes reflect mere physical stimulus similarities, we compared the neural modulations to a pixel-based measure of shape similarity, which reflects the input to the system. In addition, we computed the co-relation of the neural modulations and similarities among the stimuli using two models of visual processing. The first, wavelet-based model (Lades *et al.*, 1993) attempts to mimic the spatial filtering of V1-like complex receptive fields and predicts rather well the reaction times and error rates for discriminating between a pair of irregular shapes, similar to those used here, at least as long as the shapes do not differ in a NAP (Biederman and Subramaniam, 1997). The second model was Riesenhuber and Poggio's (Riesenhuber and Poggio, 1999) HMAX model of object recognition and categorization. This hierarchical, feature-based model consists of five layers. The units of the first four layers show a greater degree of position and size invariance and respond to increasingly more complex features the higher the layer in the hierarchy. The fifth layer consists of units that are made to respond optimally to particular views of particular objects. We computed the similarity between the shapes using the output of the fourth, C2 layer, which has an overcomplete position- and size-invariant representation of moderately complex features.

Materials and Methods

Single Cell Recordings

Subjects

Two male rhesus monkeys served as subjects. Before conducting the experiments, a head post for head fixation was implanted in each of them, under isoflurane anesthesia and strict aseptic conditions. In addition, we implanted a scleral search coil in one monkey. After training in the fixation task, we stereotactically implanted a plastic recording chamber using coordinates based on anatomical magnetic resonance images (MRIs) of each monkey. The recording chambers were positioned dorsal to IT, allowing a vertical approach, as described by Janssen *et al.* (2000). During the course of the recordings, we obtained anatomical MRI scans and computerized tomography (CT) scans with the guiding tube *in situ* in each monkey. This, together with

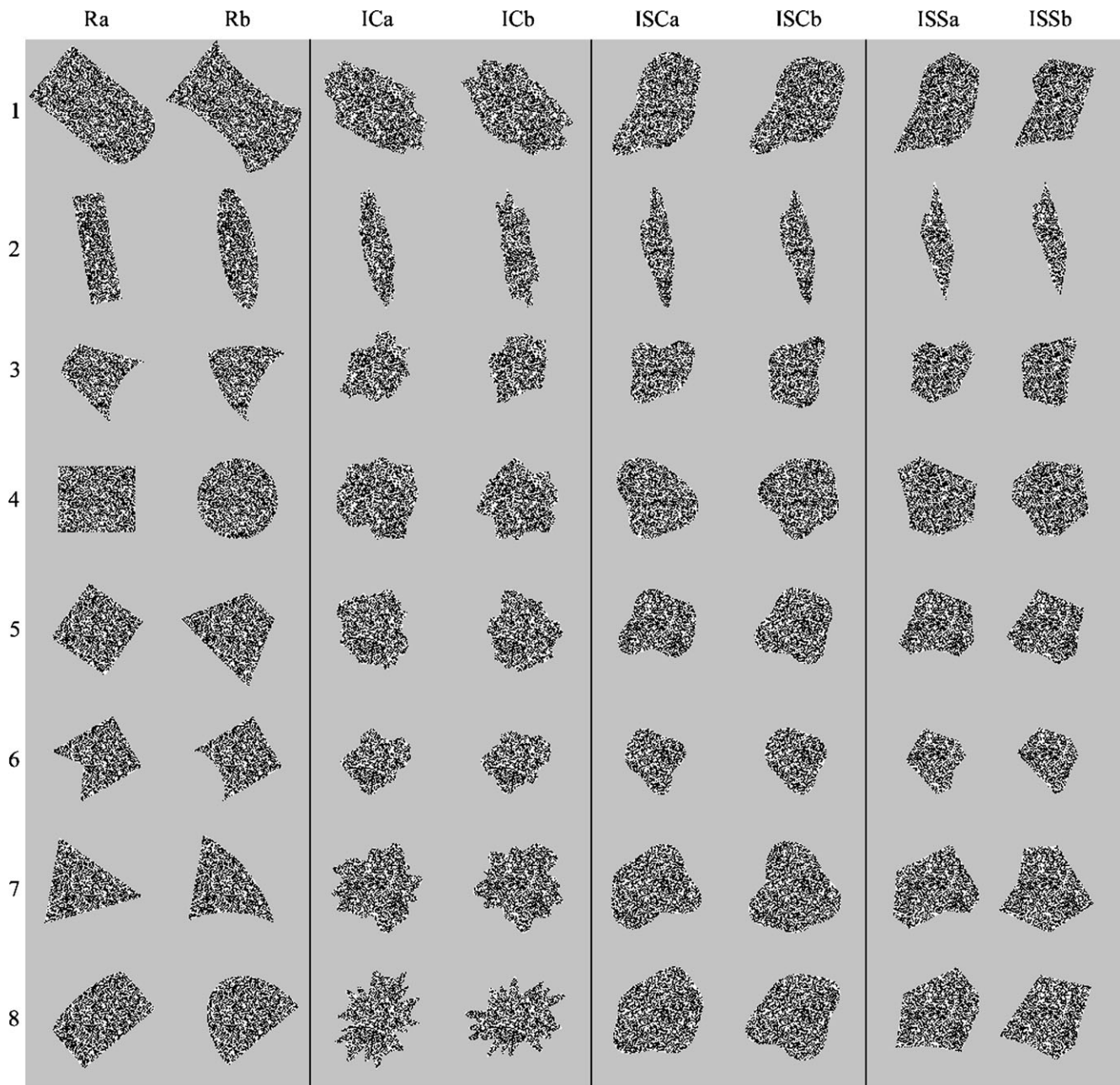


Figure 2. Regular and irregular shapes used in the main study. Regular geometrical shapes (R) and three sorts of irregular shapes (I): complex shapes (IC), simple curved shapes (ISC) and simple straight shapes (ISS). These four groups correspond to successive pairs of columns. Each of the eight rows contains four pixel-calibrated within-group pairs (column 'a' versus column 'b' for each of the four groups). The correspondence between ISS and ISC results in a fifth group of differences: ISC versus ISS, containing differences in the nature of the contour (straight versus curved) with a minimum of shape variation (ISSa versus ISCa and ISSb versus ISCb).

depth ratings of the white and gray matter transitions and of the skull basis during the recordings, allowed reconstruction of the recording positions. All surgical procedures and animal care was in accordance with the guidelines of NIH and of the KU Leuven Medical School.

Apparatus

The apparatus was similar to that described by Kayaert *et al.* (2003). The animal was seated in a primate chair, facing a computer monitor (Panasonic PanaSync/ProP110i, 21 in. display) on which the stimuli were displayed. The head of the animal was fixed and eye movements were recorded using either the magnetic search coil technique or an ISCAN eye position measurement device. Stimulus presentation and the behavioral task were under control of a computer, which also displayed the eye movements. A Narishige microdrive, which was mounted firmly

on the recording chamber, lowered a tungsten microelectrode (1–3 M Ω measured *in situ*; Frederick Hair) through a guiding tube. The latter tube was guided using a Crist grid that was attached to the microdrive. The signals of the electrode were amplified and filtered, using standard single-cell recording equipment. Single units were isolated on line using template-matching software (SPS). The timing of the single units and the stimulus and behavioral events were stored with 1 ms resolution by a PC for later offline analysis. The PC also showed raster displays and histograms of the spikes and other events that were sorted by stimulus.

Fixation Task and Recording

Trials started with the onset of a small fixation target at the display's center on which the monkey was required to fixate. After a fixation period of 300 ms, the fixation target was replaced by a stimulus for 200 ms. If the

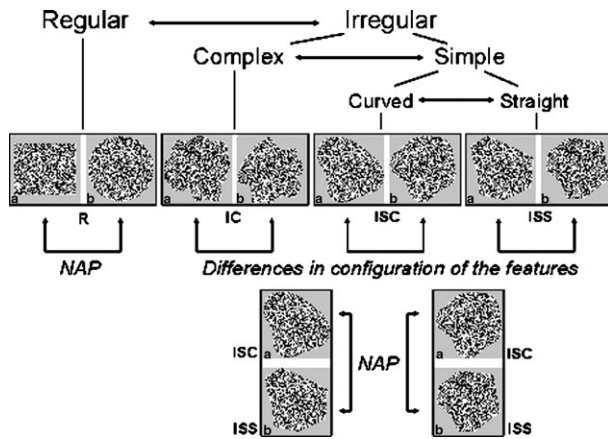


Figure 3. Overview of the relationships between the different stimulus groups illustrated by two examples of each group. The properties of the four different stimulus groups (e.g. straight versus curved) and the comparisons they enable are shown above the images; the specific variations in shape between (*a* and *b*) members of a single stimulus pair within a group are labeled below the images. The panel on the lower right illustrates the NAP difference corresponding to ISC and ISS pairs that have the same configuration (i.e. same row in Figure 3)

monkey's gaze remained within a 1.5°–2° fixation window, the stimulus was replaced again by the fixation spot, and the monkey was rewarded with a drop of apple juice. When the monkey failed to maintain fixation, the trial was aborted and the stimulus was presented during one of the subsequent fixation periods. As long as the monkey was fixating, stimuli were presented with an interval of ~1 s. Fixation breaks were followed by a 1 s time-out period before the fixation target was shown again.

All stimuli of one of the two stimulus sets (see below) were shown randomly interleaved. Responsive neurons were searched while presenting the stimuli of one of the two stimulus sets. The minimum number of interleaved presentations of a given shape was 5; the median was 10.

Stimuli

We employed two different stimulus sets. The stimuli of the first set are shown in Figure 2. They can be divided into four groups: regular geometrical shapes (R) and three sorts of irregular shapes: complex shapes (IC), simple curved shapes (ISC) and simple straight shapes (ISS). The R were created manually, using Studio MAX, release 2.5. The IC and ISC were created by means of different Fourier boundary descriptors, using Matlab, release 5. The ISS were created by replacing the curves of the ISC by straight lines, thereby taking care to preserve the general shape of the stimuli. This was done in Photoshop, release 5.5. The increase in complexity of IC compared to ISC stimuli was produced by increasing the number and frequency of the Fourier boundary descriptors composing the ISC.

Each group contains eight pairs of stimuli (one stimulus in column 'a' and one in column 'b' in Fig. 2). The rows in Figure 2 comprise a set of four pairs (one for each group) that are matched in overall size and aspect ratio, both within and between groups. The pixel-based gray level differences between the members of the pairs were equated. The members of the pairs within R differ in a NAP, such as parallel versus nonparallel sides, or straight versus curved contours. The differences among the members of an irregular pair are created by varying the phase, frequency and amplitude of the Fourier boundary descriptors (ICa versus ICb), or by varying only the phase of the Fourier boundary descriptors (ISCa versus ISCb). The differences within ISS are derived directly from the differences within ISC. The correspondence between ISS and ISC results in a fifth group of differences: ISC versus ISS, containing the nonaccidental contrast of straight versus curved contours with a minimum of global shape variation. There are 16 possible ISC versus ISS pairs, but in order to facilitate comparisons with the eight pairs within ISC and ISS, we used only one pair for each row. We selected these pairs to equate, as much as possible, the pixel gray level similarities of the within-group differences to the between-group differ-

ences. If the stimulus differences within each of the ISC and ISS pairs were smaller than the between-group differences (i.e. ISC versus ISS) we selected the ISC versus ISS pair with the smallest difference, and if both within-group pairs were larger than the between-group pairs, we used the ISC versus ISS pair with the largest difference. If one between-group pair was larger than the within-group pairs and one was smaller, we selected the smaller ISC versus ISS pair.

All stimuli were filled with the same random dot noise-pattern, consisting of black and white dots, as in the study of Op de Beeck *et al.* (2001). We incorporated the restriction that the number of black and white dots should be equal for 2 × 2 squares of pixels in the texture, so the textures were highly uniform. Stimuli were presented on a gray background and extended ~7°. The background had a mean luminance of 6.4 cd/m² and the black and white dots had luminance values of, respectively, 0 and 20 cd/m². Stimuli were shown at the center of the screen.

The second stimulus set is presented in Figure 1. It contains seven groups of line drawings of objects, each consisting of a pair with only regular parts in which one of the parts is of a different geon and a pair where that part is irregular but the configuration of that part differs between the members of the irregular pair. The stimuli were a subset of those used by Cooper *et al.* (1995). They were also presented at the center of the screen, on a gray background. The lines were in black and extended ~1.3' arc.

Image-similarity Measures

Pixel-based Differences. This calibration was done on the silhouettes of the stimuli, i.e. without the noise texture filling. We computed the Euclidean distance between the gray-level values of the pixels of the images, and this for each of 99 × 99 relative positions, as follows:

$$((\sum_i^n (G_i^1 - G_i^2)^2) / n)^{1/2}$$

where G^1 and G^2 are the gray levels of pixel i for picture 1 and 2 and n is the number of pixels. The similarity of a stimulus pair was defined as the smallest value over the 99 × 99 positions. This pixelwise gray-level analysis compares image similarities as present in the input, without making any commitment to differential sensitivities to higher-order features. Thus, any difference between the neural-based representation and this pixel-based physical representation must originate within the visual system.

Lades Model Outputs. The Lades (Lades *et al.*, 1993) model produces wavelet-based image measures. It filters the image using a grid of Gabor jets and then computes the difference between the outputs of the different Gabor jets for the images of a pair. In the present implementation, the square grid consisted of 10 × 10 nodes with a node approximately every 10 pixels, horizontally and vertically, as the images are rescaled to be squares of 128 pixels (instead of the original 350 pixels). A Gabor jet consisting of 40 Gabor filters at five spatial frequencies and eight orientations was positioned at each node. The five spatial frequencies were logarithmically scaled with the lowest spatial frequency filter covering a quarter of the image and the highest-frequency filter covering about four pixels on each side. At each node, each jet can be described as a vector of 40 Gabor filter outputs. The similarity of a pair of images is computed as an average of local similarities between corresponding Gabor jets. The local similarity is given by the cosine of the two vectors, thereby discarding phase parts. The grids were positioned on the stimuli in such a way as to maximize the similarity between pairs of images. Throughout the Results section, we will use Lades model dissimilarities, which are computed by subtracting the Lades model similarities from 1.00.

HMAX Outputs. We computed the Euclidean distances between the outputs of C2-units of the HMAX model (described by Riesenhuber and Poggio, 1999) and downloaded from <http://www.ai.mit.edu/people/max/> on 9 July 2003). These units are designed to extract moderately complex features from objects, irrespective of size, position and their relative geometry in the image (see Riesenhuber and Poggio, 1999, for more details). Image similarity was computed as the Euclidean distance between the output of the 256 C2 units.

Analyses

The response of a neuron was defined as the number of spikes during a time interval of 250 ms, starting from 50 to 150 ms. The starting point of the time interval was chosen independently for each neuron to best capture its response, by inspecting the peristimulus time histograms, but was fixed for a particular neuron. Each neuron showed significant response modulation to the shapes, which was tested by an ANOVA.

The degree of response modulation to each possible stimulus pair was assessed by calculating the normalized Euclidean distance between the response of the neurons, using the following formula:

$$\left(\frac{\sum_i (R_i^1 - R_i^2)^2}{n} \right)^{1/2},$$

where R_i^1 is the response of neuron i , averaged over trials, to stimulus 1, and n is the number of neurons tested with that pair of stimuli. The resulting distance matrix allowed inferences as to the representation of the shapes of the different stimulus groups in IT.

Parametric (ANOVA) and, when possible, non-parametric statistical tests were used to compare responses to different stimuli.

MDS Analyses

Multidimensional scaling (MDS) analyses were done using Statistica software (StatSoft, Tulsa, OK). We employed the standard non-metric MDS algorithm using a matrix of the distances between each pair of shapes as input. The distances were computed using either neural modulations (Euclidean distance [see above]; same procedure as Young and Yamane, 1992; Op de Beeck *et al.*, 2001; Vogels *et al.*, 2001), or the different image similarity measures (see above). The MDS algorithm arranges the stimuli in a low-dimensional space while trying to maintain the observed distances.

The MDS analyses were done on the 32 shapes of two stimulus groups. MDS produces a low-dimensional configuration that can be freely translated, scaled and rotated. Scree plot analysis as well as inspection of the Shepard plots indicated that three dimensions were satisfactory in that they accounted for >90% of the variance of the neural distance matrix in each of the analyses (see Results). The three-dimensional solutions were rotated to find a maximum separation between the groups along one of these three dimensions. This was done in Matlab by orthogonally rotating the configurations in steps of 1° and finding the rotation at which the overlap between the members of groups along one of the dimensions was the smallest.

We used randomization statistics to assess whether the separations we found were significantly different from those obtained from a random configuration of two sets of 16 points without *a priori* separation. This was done by randomizing the neural responses of each neuron between stimulus groups, i.e. in the high-dimensional neural space. Then, as for the original data, MDS was used to obtain a three-dimensional configuration. The optimization procedure described previously was then used to maximally separate the groups. Thus, the three-dimensional configurations of the randomized data were rotated to find the optimal separation between the two groups. The randomization procedure was repeated 1000 times and we computed the proportion of random configurations that yielded a separation at least as great as the separation obtained with the neural data. The test was judged to be significant when the latter proportion was <0.05.

We supplemented this analysis with one aimed not at minimizing the degree of overlap between the groups, but at maximizing the overall distance between the groups. We maximized the t -value of the difference in mean position between the shapes of the two stimulus groups along the direction of maximum group separation. We report the maximal t -value after rotating the three-dimensional configurations in steps of 1°. The significance of this value was again assessed by randomizing the original neural data between groups, performing MDS to obtain a three-dimensional configuration, rotating to optimize the t -value and measuring the resulting t -value a 1000 times, and then computing the proportion of t -values that were as high or higher than the t -values found with the original subdivision.

Comparison of Within- and Between-group Distances

The distances among the stimuli of two different groups were computed using the neural responses (Euclidean distances), pixel gray levels, Lades, HMAX model outputs and perceived distances (see

below). We compared the distribution of the distances of the stimulus pairs belonging to the same group to that of distances of the stimulus pairs belonging to different groups. To quantify the degree of overlap of these two distributions we performed receiver operator characteristic (ROC) analysis (Green and Swets, 1974). The proportion of distances of the within-group and between-group distributions that was smaller than a particular criterion distance defined the proportion of 'hits' and 'false alarms', respectively. The area under the ROC curve plotting the hits and false alarms for all possible criterion distances is a quantitative measure of the separation of the two distributions. An area of 100% (or 0%) indicates that all between-group distances were larger (or smaller), than the within-group distances, i.e. complete separation, while 50% corresponds to no separation. The area under the ROC curve can also be understood as the percent correct that would be obtained if one had to decide whether two shapes belong to the same group solely on the basis of the distance between the two shapes. We will refer to this area under the ROC curve as the 'separation probability.' The latter probability can be computed for the distributions of the neural distances, pixel-based distances, Lades model distances, HMAX distances or perceived distances (see below).

Quantitative Measures of Similarity of Configurations

The congruence coefficient was used to assess the correspondence between different spatial (Euclidean) configurations, e.g. between that defined by neural similarities and image-based distances among the shapes. Its advantage, compared to the Pearson correlation, for assessing similarities among configurations is discussed in Borg and Leutner (1985). It is computed as follows:

$$c = \frac{\sum_i d_{A_i} d_{B_i}}{\left(\sum_i d_{A_i}^2 \sum_i d_{B_i}^2 \right)^{1/2}},$$

where d_{A_i} and d_{B_i} are the distances for the shape pair i of the configurations A and B, respectively.

The relationship between the degree of similarity and the congruence coefficient is a monotonically increasing but highly non-linear function. Thus, congruence coefficients for dissimilar configurations can already be quite large, and small differences among large congruence coefficients can be meaningful.

We evaluated the significance of the congruence between our neural data and our image similarity values using the following randomization test. The position of 16 stimuli of a group in an N -dimensional space, using MDS on the original neural distance matrix, was computed with N varying from 5 to 9. Then the stimulus position assignment in this N -dimensional space was permuted and this permutation was performed 1000 times. For each permutation, we rebuilt the distance matrix and recalculated the congruence coefficient ($n = 1000$). The test on the actual data was judged to be significant when the proportion of permuted matrices that yielded congruence coefficients at least as high as the one under testing was smaller than 0.05 for each of the N -dimensional configurations.

The significance of the difference between the two congruence coefficients of two different groups (e.g. comparing the congruence coefficient for the neural data and the pixel-based distances of the regular group with that of the neural data and the pixel-based distances of an irregular group) was tested by randomly splitting the neural data into two equal parts. The first part of the data was used to recalculate the congruence coefficients for both stimulus groups and to check whether they still differed in the original direction. This randomization and computation was repeated 1000 times. If the proportion that the direction of the difference reversed that of the original was <0.05, the original direction was judged to be significant.

Sorting Task

This is a classical task to measure judged image similarities (Hahn and Ramscar, 2001). We printed each of the 64 shapes of Figure 2 on 6.4 × 6.4 cm high-quality photopaper. These 64 cards were given to 23 naive human subjects who were required to sort the stimuli in groups based on shape similarity. No further definition of similarity was given and they could freely choose the number of groups (at least 2 and at most 63). Dissimilarity values between pairs of stimuli were calculated by counting the number of subjects that put the two members in different groups.

Results

Neural Responses

We recorded the responses of 119 anterior IT neurons that showed significant response selectivity to the stimuli of the set of Figure 2 (ANOVA, $P < 0.05$). They were obtained from two monkeys (76 in monkey 1 and 43 in monkey 2). The depth readings and white/gray matter transitions observed during the recordings, and the analysis of the CT and MRI images indicated that the neurons were located in the lower bank of the superior temporal sulcus and the lateral convexity of anterior IT (TEad).

First, the responses of single neurons to the full stimulus set, comparing the modulations to shapes of the same versus different stimulus groups, will be described. This will be followed by the population analysis using MDS of the separation of the four shape groups. Then the assessment as to whether the sensitivity of IT neurons differs between the groups will be presented

Separation of the Four Different Shape Groups

The neurons tended to prefer stimuli from one group over stimuli of another, although rarely was the preference com-

pletely confined to a single group. The general group preference and feature tuning are illustrated by the neuronal responses shown in Figure 4. The neuron presented in Figure 4A was highly selective to some (but not all) regular shapes (R), often those with straight contours and compact aspect ratios. Nevertheless, it also responded weakly but significantly to a few of the irregular simple, compact straight shapes (ANOVA; $P < 0.0001$ for the main effect of response, using the ISS shapes only).

Other neurons preferred stimuli from one of the irregular groups. Figure 4B,C shows neurons that preferred, respectively, IC and ISS shapes. The neuron depicted in Figure 4D responded well to curves: its strongest response was to the circle. In general it preferred compact ISC stimuli although it also responded relatively strongly to the ellipse. Other neurons, such as the one shown in Figure 4E, preferred ISC shapes without a strong response to the circle or ellipse. In this particular neuron, the strongest responses to shapes not belonging to ISC were elicited by the stimuli of the ISS group that had the same overall shape as the two most preferred ISC shapes. The neuron in Figure 4F required a globally elongated shape to produce a strong response, and this global shape property was more critical than the presence of certain features or local shape variations.

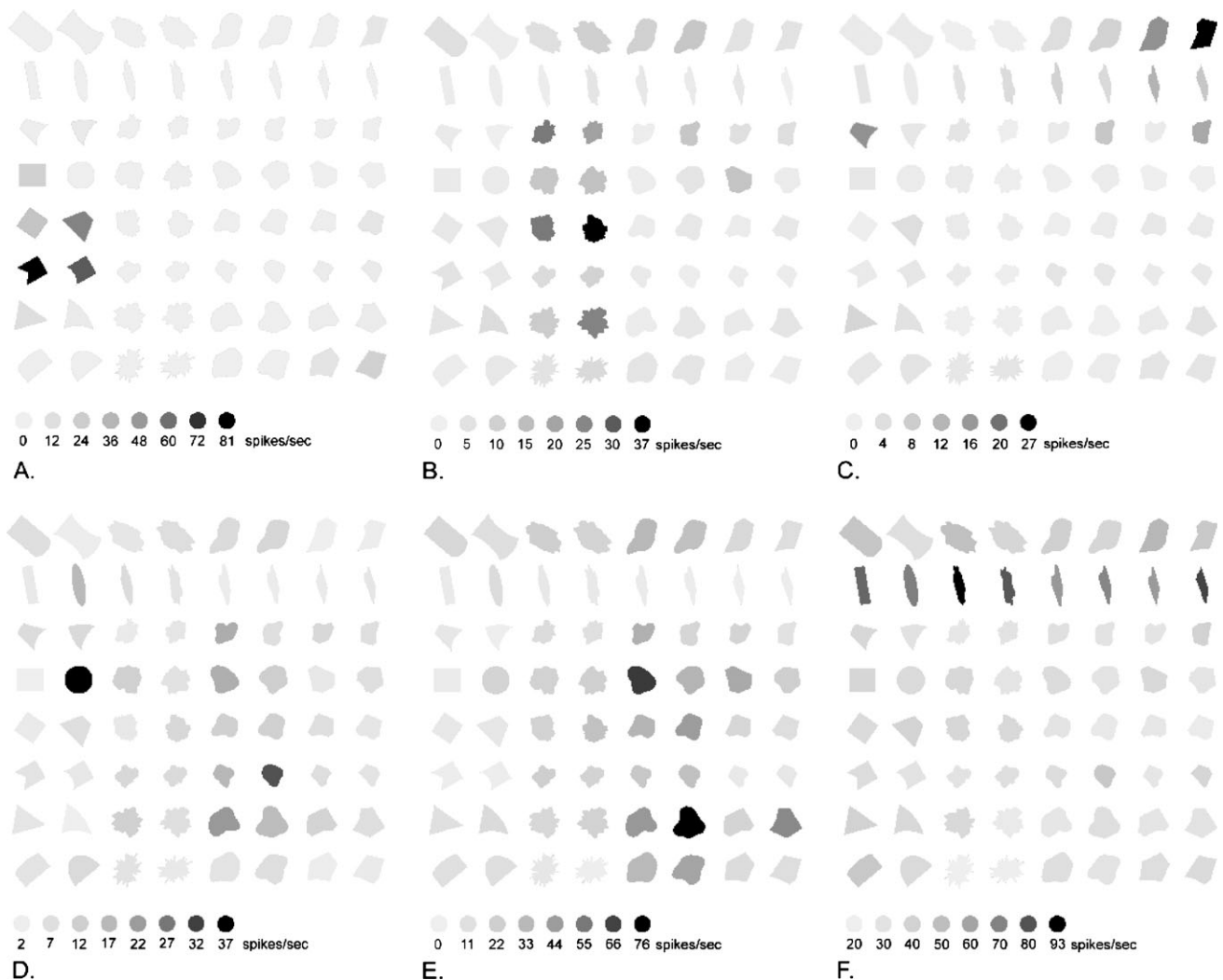


Figure 4. Responses of six neurons (A–F). The darker the depiction of the shapes, the stronger the response. The calibration bars underneath each of the sets indicate the response (in spikes/s) corresponding to the different shades.

The fact that our neurons showed no tendency to respond uniquely and completely to the members of a single group does not imply that the groups could not be separated based on the activity of the population of IT neurons. A complete separation might be found at a population level, which was analyzed with MDS. The purpose was to determine whether there exists a dimension that linearly separates the different groups. Due to the diverse shape differences in our stimulus set, the complete neural distance matrix was too high dimensional to fit in a sufficiently low-dimensional MDS solution. Therefore, in order to reduce the dimensionality of the neural shape space (and avoid using a model in which the explained variance is too low when forcing the distances into a manageable low-dimensional space), we conducted six different MDS analyses, with each a distance matrix consisting of the distances within and between two groups, thereby pairing all groups.

Figure 5 shows a two-dimensional projection of the MDS solution of a distance matrix containing the shapes from the regular (R) and the irregular complex (IC) groups. The linear separation between the groups within this low-dimensional subspace was perfect ($p < 0.001$, randomization statistics; see Materials and Methods). The solution was rotated such that the dimension separating the groups is shown on the horizontal axis.

The same projection of the R/IC configuration is presented in Figure 6*IA*. Figure 6 provides an overview of all the rotated MDS solutions of the neural modulations, model differences (discussed later) and human dissimilarities (also discussed later). The first column in Figure 6 (i.e. 6*IA*–6*6A*) presents the two-dimensional projections of the neural MDS solutions of the six possible group pairs. In each of the six cases a three-dimensional solution was acceptable, accounting for 92–96% of the variance of the original distance matrix. In each case, we rotated the three-dimensional space so that one of the dimensions separated the two groups (shown on the horizontal axis). For each pair of stimulus groups, the linear separation between the groups was perfect within these low-dimensional subspaces ($P < 0.001$, randomization statistics). Also the average difference in position between the groups was large (see Table 1 for maximal t -values; see Materials and Methods) and significantly higher than chance in all cases ($P < 0.001$, randomization statistics; see Materials and Methods). After this rotation,

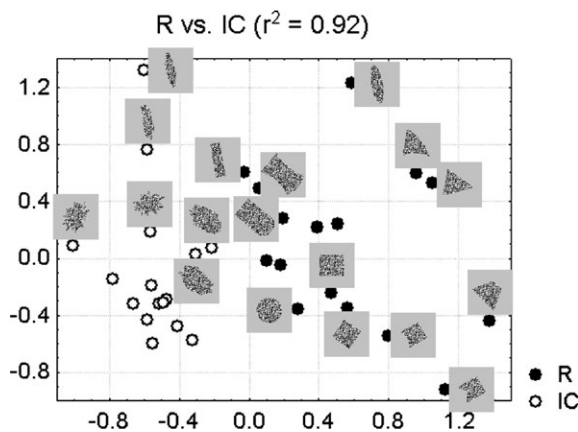


Figure 5. Two-dimensional representation of a three-dimensional shape configuration containing the stimuli from group R and IC. The distances are Euclidean distances between the neural firing rates. Some data points are accompanied by a picture of the shape they represent.

another dimension invariantly separated the stimuli of row 2 in Figure 2 (indicated by squares in Figure 6) from the others (shown on the vertical axis). This dimension appears to be one of aspect ratio or orientation. In general, stimulus variations along the third dimension could not be interpreted and thus are not shown.

The variation along the horizontal dimension within each of the groups does not have a simple interpretation. It should be noted that these dimensions might not be ‘homogenous’. They could be determined by various neurons that have nothing in common but the fact that they segregate both groups. For example, the neurons in Figure 4*A* and Figure 4*D* will both contribute to the segregation between group R and IC (Fig. 5), preferring stimuli from R, and so will the neuron in Figure 4*B*, preferring stimuli from IC, and the subset of neurons (not presented) that responded selectively and strongly to (only) the two IC stimuli in row 8 of Figure 2. The tuning of these neurons differs, but they contribute to the segregation of the two stimulus groups.

To conclude, IT neurons show tendencies to prefer one of the four stimulus groups and when pooled together the combined tendencies result in a linear separation of the stimulus groups in a three-dimensional space. Neurons also separate the stimuli from row 2 in Figure 2 from the others, individually (e.g. Fig. 4*F*), as well as on a population level.

Do these MDS configurations derived from the neural data merely reflect the pixel-based image differences? To facilitate comparison of the neural and physical configurations, the pixel-based distance matrices (six in total, each containing the distances within and between two stimulus groups) were reduced to three dimensions as well. These three-dimensional solutions explained 94–96% of the variance in the original pixel-based distance matrix and are shown in column B of Figure 6. The horizontal dimension of these rotated configurations had the least number of stimuli lying on the ‘wrong’ side of the optimal border that separated the groups (see Materials and Methods). As can be seen in the horizontal overlap of open and filled symbols in Figure 6, none of the six pixel-based configurations provided a full linear separation between the groups. As shown in Table 1, the t -values of the pixel-based group differences were much lower than those for the neural data. Thus, the between-group segregation observed in the IT response data must have emerged in the visual system. The pixel-based configuration does, however, separate the stimuli of row 2 in Figure 2 (squares in Fig. 6) from the others in all six configurations (second column in Fig. 6, vertical axis within each configuration).

Within-group Neural Modulations Compared for the Different Shape Groups

The above MDS analyses indicate that a population of IT neurons linearly separates four shape groups. The next analysis compares the degree of the response modulation between the shape groups.

Each group can be divided into eight pairs, depicted by the eight rows of Figure 2, one member belonging to column ‘a’ of each group and the other to column ‘b’. The pairs on the same row are calibrated in pixel difference and its members have approximately equal aspect ratios. Figure 7*A* shows the mean neural modulation between the members of a pair for each of the four shape groups and the ISC/ISS comparison (straight versus curved irregular shapes), averaged over the eight pairs of

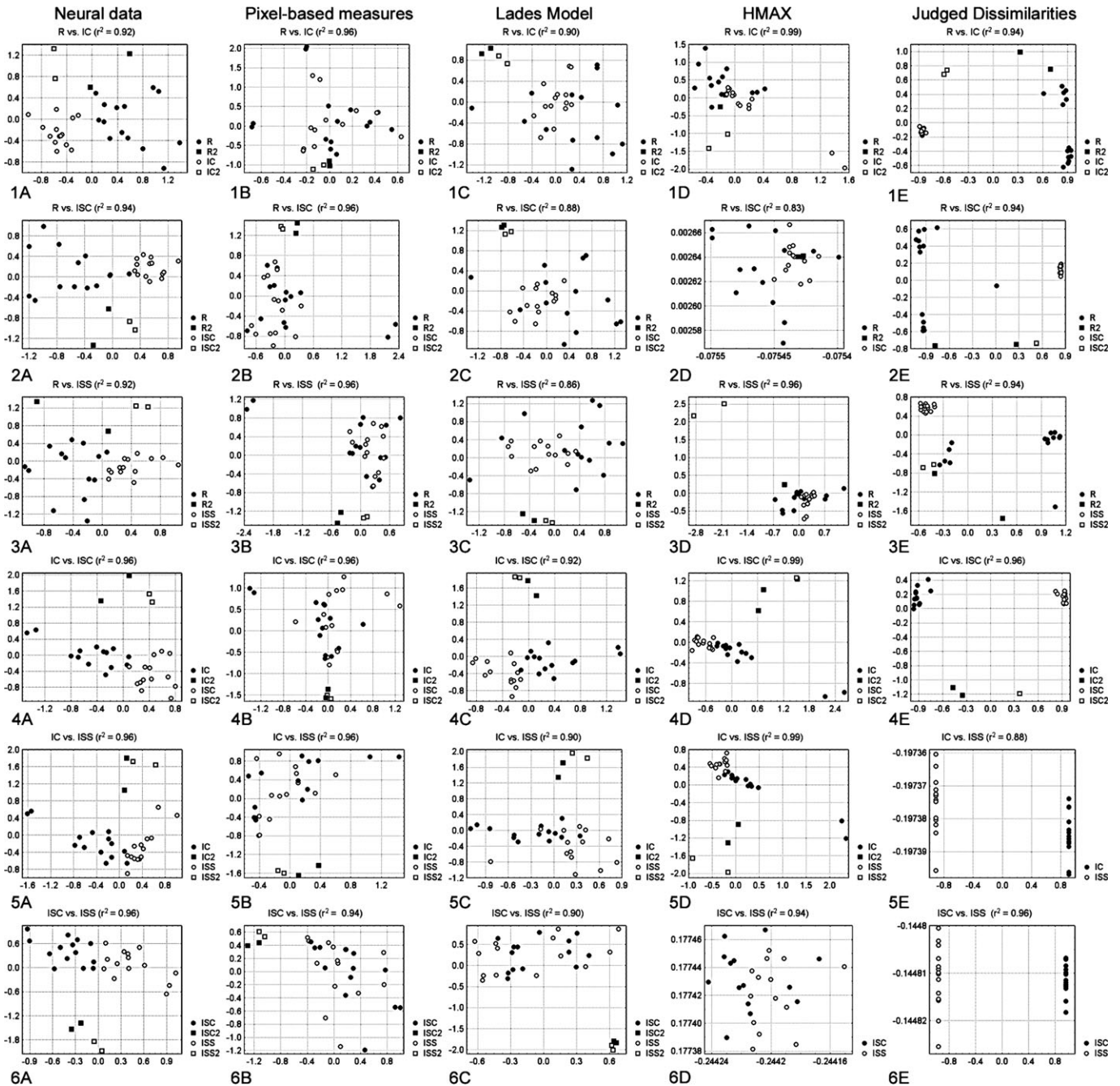


Figure 6. Two-dimensional projections of three-dimensional stimulus configurations. Each configuration contains the shapes of two groups, labeled above the graphs. Each of the six rows corresponds to one of the combinations of pairings of the four conditions enabling assessment of the comparisons shown in the upper portion of Figure 3. The distances between the shapes are measured, starting from the leftmost columns, neural Euclidean distances, pixel-based Euclidean distances, Lades model dissimilarities, Euclidean distances between Hmax C2 layer outputs and human judged dissimilarities. In panels 2d, 6d, 5e and 6e all or some of the stimuli of row 2 were omitted because the distances between those stimuli and the other rows were too high to be included in the panel without a compression of the axis so extreme that the other differences would be indistinguishable.

a group. The neural modulation is expressed as the normalized Euclidean distance (see Materials and Methods) between the two shapes of a particular pair. The neural modulation for the regular shape pairs is significantly larger compared to that for each of the three groups of irregular shapes (Wilcoxon signed rank, $P < 0.05$, $n = 8$, tested for each comparison separately; significant when averaging across monkeys and in each monkey separately). This effect must originate within the visual system, as the pixel differences between members of a pair were, on

average, lower for the regular group than for the three irregular groups (Fig. 7B). The neural modulation within the irregular complex group is approximately equal to the modulations within the irregular simple groups, making it unlikely that the difference between R and IC can be explained by the difference in complexity between the two groups. The modulation to ISC versus ISS is significantly greater than the modulations within IC, ISC and ISS [Wilcoxon signed rank, $P < 0.05$, $n = 8$, tested for each comparison separately (although the greater response

modulation for ISC versus ISS was only significant in monkey 1)], which suggests that it is a change in a NAP which underlies the greater modulation for the regular shapes. Note that the greater neural modulation for the irregular curved versus straight shapes comparison (i.e. ISC versus ISS) relative to configurational differences (within ISC or ISS) is probably underestimated

Table 1

Maximal *t*-values of difference in mean position on a separating dimension for the shapes of each combination of the four stimulus groups

Pair	Neural modulation	Pixel-based differences	Lades model	Hmax	Sorting task
R-IC	10.33	0.73	1.5	3.48	37.93
R-ISC	8.79	2.28	2.41	3.01	17.64
R-ISS	7.64	1.86	2.44	1.13	10.92
IC-ISC	8.49	2.31	6.66	4.86	25.89
IC-ISS	5.87	1.96	4.04	4.27	14.5
ISC-ISS	9.71	1.25	0.74	0.26	500.83

as the average pixel-difference within the ISC versus ISS-pairs was much lower than the pixel-differences within the other pairs (Fig. 7B).

Figure 7A shows that the modulation for the ISC versus ISS difference was approximately as large as the modulation within the R groups. Given that the stimulus differences were smaller in the ISC versus ISS comparisons than in the within-R comparisons (Fig. 7B), one might be tempted to conclude that NAP differences are at least as detectable, if not more so, within irregular shapes than regular, symmetric shapes. However, such an inference may be unwarranted. On average, there were more NAP differences between members of an ISC versus ISS pair — because every contour and every vertex (or extrema of curvature) differed in all eight pairs — whereas for all but two of the R pairs (rows 2 and 4 in Fig. 2) only some of the contours and few, if any, of the vertices changed between the ‘a’ and ‘b’ versions. The only inference that can be drawn from the ISC versus ISS comparison is that a robust effect of these NAP differences can be obtained with asymmetrical shapes.

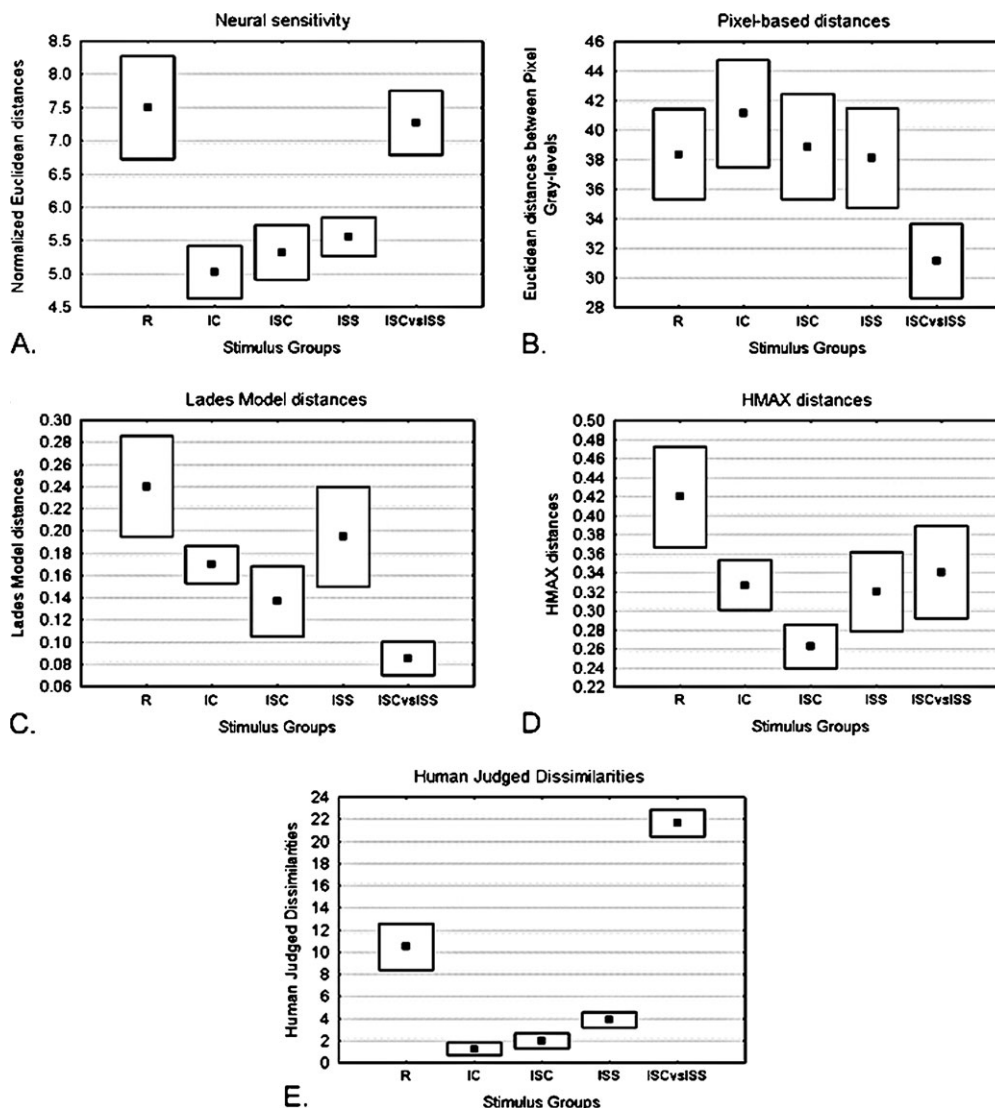


Figure 7. Dissimilarities for the five calibrated kinds of comparison (described in Figs 2 and 3), each averaged over the eight pairs (rows in Fig. 2). Panels (A-E) show the dissimilarities based on, respectively, neural Euclidean distances, pixel-based Euclidean distances, Lades model dissimilarities, Euclidean distances between Hmax C2-layer outputs and human judgment distances. The error bars indicate standard errors of the mean.

The Correspondence of the Neural Modulations and the Pixel-based Image Changes Compared for the Different Stimulus Groups

Human psychophysical studies (Biederman *et al.*, 1999) suggest that perceived shape similarity relates well to physical distances, as long as nonaccidental shape differences are not present. Thus one might expect a greater correspondence between neural and physical distances for the irregular groups compared to the regular groups.

Table 2 shows the congruence coefficients computed on the neural modulations and the image changes for each group separately. The congruence coefficients for the neural modulations and the image changes are significant in all cases ($P < 0.001$, randomization statistics, see Materials and Methods), indicating significant correspondences between neural modulation and image changes for each group separately. However, as predicted, the neural modulations within the irregular groups were significantly better related to the pixel-based image changes than those within the regular group ($P < 0.05$ for each comparison, randomization statistics, see Materials and Methods).

The correlation of the neural and pixel-based differences is shown in Figure 8A. The distances between the regular shapes are shown as filled circles; those within the irregular groups as open squares (the latter represent distances within but not between the three different irregular groups). The relation between the neural modulations to the shape differences of the irregular groups and the corresponding pixel-based image changes (Fig. 8A, thin fitted function) loosely follows a rising linear function ($r^2 = 0.36$), while the fitted linear function (thick line) between the neural modulation and image changes for the regular shapes is almost flat ($r^2 = 0.03$).

Comparison Between the Neural Modulations and the Outputs of the Lades Model

Separation of the Stimulus Groups: MDS Analysis

The Lades model preserves the segregation of the elongated stimuli of row 2 from the other compact stimuli, a separation

Table 2

Congruences between image differences and neural modulation for each of the four stimulus groups

Group	Pixel-based differences	Lades model	Hmax
R	0.9559	0.9484	0.9786
IC	0.9723	0.9805	0.9276
ISC	0.9834	0.978	0.8964
ISS	0.9799	0.9721	0.9328

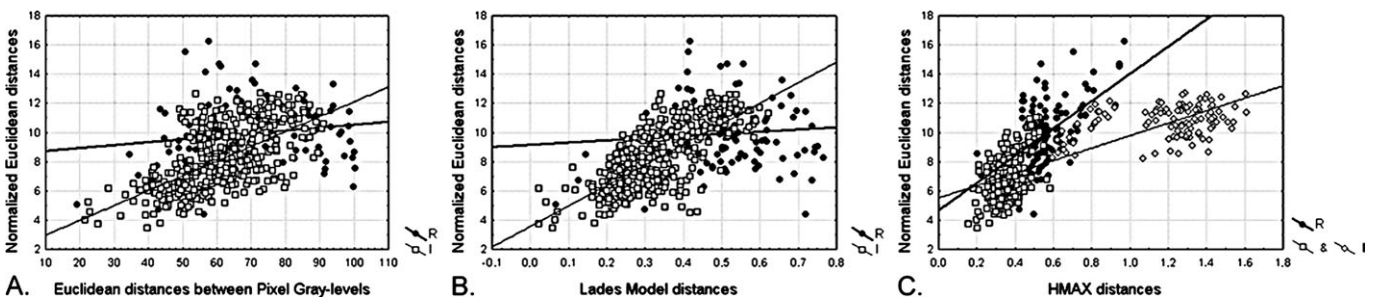


Figure 8. Neural Euclidean distances between the shapes plotted against the corresponding image-based differences. Panels (A–C) show, respectively, pixel-based Euclidean distances, Lades model dissimilarities and Euclidean distances between Hmax outputs. The distances between stimuli within R are shown as filled circles with the thick line showing the linear fit. The distances within each of the three irregular groups are shown as empty squares with the thin lines showing the linear fit. Panel (C) includes a marked subgroup containing the differences between the irregular stimuli from row 2 and the other rows (see Fig. 2) and those between the IC stimuli from row 8 and the IC stimuli from other rows (empty diamonds).

already present in the pixel-based distances (third column in Fig. 6, vertical axis within each configuration). However, this model does not produce linear separations between the regular and the irregular stimuli (Fig. 6, configurations 1C, 2C and 3C, horizontal axes), and between ISS and ISC (Fig. 6, configuration 6C, horizontal axis). The Lades model, does, however, separate the irregular complex and simple stimuli (Fig. 6, configurations 4C, horizontal axes). The t -values of the Lades based group differences were always lower than the neural t -values (Table 1).

Within-group Sensitivities Compared for the Regular and the Irregular Shapes

The Lades model-based distances within the pixel-calibrated pairs are somewhat greater for the R than the irregular groups, consistent with the neural modulation although to a lesser extent (Fig. 7C). Consistent with the pixel-based differences, but contrary to the neural data, there is a lower sensitivity to the straight-curved contrast of the ISC versus ISS comparison relative to the within-group ISC and ISS differences.

Correspondence Between the Neural Modulations and the Lades Model Sensitivities

The correlation between the Lades model differences and the neural modulations are similar to those obtained using the pixel-based image changes (compare Fig. 8A and 8B). Table 2 shows that the neural modulations within the irregular groups were also more strongly related to the Lades-based image changes than those within the regular group, an effect that was significant in each case ($P < 0.05$, randomization statistics, see Materials and Methods). The congruence coefficients of the neural modulations and the image changes were significant in all cases ($P < 0.001$, randomization statistics, see Materials and Methods).

Modulation for Lades Model Calibrated Stimulus Pairs of Regular and Irregular Shapes

The Lades model showed higher sensitivity to the differences between the members of the regular shape pairs. To what extent can the relatively greater neuronal modulation for these differences generally be derived from the output of V1-like spatial filters? To address this question, we tested IT neurons using a stimulus set (Fig. 1, from Cooper *et al.*, 1995) that was constructed of wavelet-calibrated rather than pixel-calibrated pairs. Each row depicts a group of four line drawings of objects, a pair with only regular parts and a pair containing one irregular part. The latter irregular part and its regular counterpart differ between the members of the pairs. The differences between the irregular parts were configurational differences, the differences between

their regular counterparts were NAP differences. For each row, the Lades model image differences, presented in Figure 9A, are lower for the pairs with only regular parts. The neural Euclidean distances, based on the responses of 69 IT neurons collected in two monkeys (35 in monkey 1 and 34 in monkey 2), are shown in Figure 9B. There is a significantly higher sensitivity for the pairs differing in a NAP of a regular part (mean 7.36; SE 1.36) compared to the pairs differing in the configuration of an irregular part (mean 5.32; SE 1.07) (Wilcoxon signed rank $P < 0.05$, $n = 7$). This suggests that the higher neural sensitivity to differences among regular shapes differing in NAPs compared to irregular shapes differing in metric properties cannot be explained by V1-like activity as simulated by the Lades model.

Comparison Between the Neural Modulations and the Output of HMAX

Separation of the Stimulus Groups: MDS Analysis.

The segregation of the row 2 stimuli (Fig. 2) is not consistently present in the configurations based on HMAX (third column in Fig. 6, vertical axis within each configuration). Indeed, HMAX does not fully separate the stimuli in row 2 of the regular group. The between-group separations are similar to those obtained using the Lades model: no complete linear separation of the regular and the irregular stimuli (Fig. 6, configurations 1D, 2D and 3D, horizontal axes), a segregation of the irregular complex and simple stimuli (Fig. 6, configurations 4D and 5D, horizontal axes) and no segregation of the straight and the curved simple irregular stimuli (Fig. 6, configuration 6D, horizontal axis). The t -values of the group differences were in all instances lower than the neural t -values (Table 1).

Within-group Sensitivities Compared for Different Shape Groups

HMAX is more sensitive to the differences for the pixel-calibrated regular shape pairs compared to the irregular pairs, in correspondence with the Lades model (Fig. 7D). Again the increase in sensitivity is slightly less pronounced than for the neural data. HMAX also encompasses a very small elevation in sensitivity to ISC versus ISS that is not apparent in either the Lades or the pixel-based measures.

Relationship Between the Neural Modulations and the HMAX Sensitivities

The relationship between the neural modulations and the HMAX sensitivities is stimulus group-dependent, but in a qualitatively

different manner than the correspondences between the neural modulations and the other image measures (Table 2). There is better correspondence of the neural and HMAX similarities for the regular shapes compared to the other groups ($P < 0.05$ in each case, randomization statistics, see Materials and Methods), which is opposite to that observed for the other image similarities. All the congruence coefficients for HMAX were significant ($P < 0.001$, randomization statistics, see Materials and Methods).

Figure 8C shows the correlation between the HMAX outputs and the neural modulation for both the regular group and the irregular groups. The correlation between the neural modulations and the HMAX sensitivity for the regular shapes is pronounced, while the configuration of the irregular HMAX distances is qualitatively different from the neural configuration. Some differences, namely those between the irregular shapes in row 2 and the other irregular shapes, and those between the IC shapes in row 8 and the other IC shapes (indicated by diamonds), are given excessive weight in HMAX compared to the neural modulation.

Comparison Between the Neural Modulations and the Perceived Similarities Obtained in a Sorting Task

We obtained an estimate of the judged perceptual differences between the images by having human subjects sort the stimuli of Figure 2 (see Materials and Methods).

Separation of the Stimulus Groups

The stimuli of row 2 are separated from the others in every stimulus group (Fig. 6, fifth column, vertical dimensions), in agreement with the neural behavior. In configuration 5E, however, the latter separation is not orthogonal to the dimension that segregates both groups. All stimulus groups are separated from each other. The segregations are generally more extreme than the neural segregations with the exception of configuration 3E where a subgroup of R is still situated relatively close to the ISS shapes. This subgroup consisted of all the regular shapes that contain only straight edges. The greater segregation of the shape groups by the human subjects than by the neural modulation is reflected in the t -values of the group differences, which are always much higher for the configurations based on the sorting task (Table 1).

Within-group Perceived Similarities Compared for the Different Shape Groups

The similarities inferred from the human sorting behavior (shown in Fig. 7E) differ from the image similarity measures

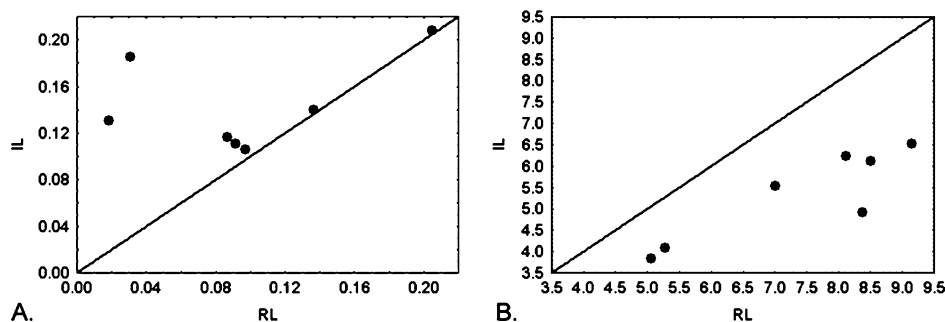


Figure 9. Average distances within the pairs in the stimulus set depicted in Figure 1, each point representing a row. The pairs within RL (distance presented on the abscissa) consist only of regular parts, with one part differing between the members of a pair. In the corresponding pairs within IL (distances presented on the ordinate), this regular part is replaced by an irregular part, that is again differing between the members of the pairs. Distances are Lades model distances (A) and neural Euclidean distances obtained from 69 neurons in two monkeys (B).

and the neural modulations in the sense that there was a relatively higher sensitivity to the straight versus curved contrast of the ISC versus ISS comparison. The sensitivity to ISC versus ISS was significantly higher than to the pairs within each of the irregular stimulus groups (Wilcoxon signed rank, $P < 0.02$, $n = 8$, tested for each of the four comparisons separately). Every single subject put the majority of the ISC and ISS shapes in different groups. The only exceptions involved the stimuli in row 2: some subjects made a separate group containing all the stimuli in row 2. The high sensitivity to ISC versus ISS could be related to the neural segregation between both groups, as described previously. [This segregation was apparent in the neurons of both monkeys separately, with respectively 0 and 3 exceptions ($P < 0.01$, randomization statistics); the exceptions in monkey 2 included, probably not by coincidence, the stimuli in row 2.]

In accordance with the neural data, there is a significantly higher sensitivity to comparisons within the regular than within the irregular groups (Wilcoxon signed rank, $P < 0.02$, $n = 8$). There is also a marginal but significant (Wilcoxon signed rank, $P < 0.05$, $n = 8$) higher sensitivity to comparisons within ISS compared to comparisons within IC. This might be related to the slightly but not significantly higher neural (and Lades model) sensitivities to ISS compared to IC. In general, there is a consistent and striking similarity between the perceived similarities (Fig. 7E) and the neural modulations (Fig. 7A) which both differ from the image similarities in qualitatively the same way.

Within-group versus Between-group Similarities for Neural Responses: The Pixel-based Measure, Lades Output, HMAX Output and Human Sorting Behavior Compared

As a quantitative measure of the overlap of the distributions of the within-group and between-group distances, we employed the ROC-based separation probability (see Materials and Methods). This allows a quantitative comparison of the separation between stimulus groups that does not involve a MDS analysis. We computed the separation probabilities for, respectively, the distances based on IT neuronal responses, pixel-based gray levels, the wavelet-based Lades model, the HMAX C2 units and the human sorting data. Table 3 shows the separation probabilities for each of the six pairwise comparisons of the different stimulus groups.

Using responses of IT neurons to compute the shape distances produced separation probabilities that are larger than the near-chance probabilities obtained with the pixel-based, local gray-level stimulus measure. This shows that the separation observed in IT is really due to processing of image features that are more complex than local intensities. In general, the Lades model and HMAX model outputs produced similar separation probabilities, usually smaller than those of the IT neurons and the judged distances by human subjects. For the IC-ISC and IC-

ISS comparisons, HMAX outperformed the Lades model. In fact, for these comparisons the separation probabilities were very similar to those obtained with the IT responses. These separation probabilities fit well the separations found in the MDS analysis of the HMAX and wavelet-based distances for these group comparisons (see above).

The separation probabilities based on the sorting task in the human subjects were close to maximum, except for the R-ISS comparison. This might not be surprising given that 13 of the 16 R shapes have at least one straight contour, as do all the ISS images. Note that each of the separation probabilities computed using the sorting data is larger than those based on the neural responses.

Discussion

We found that, first, the activity of a population of IT neurons can distinguish groups of shapes based on perceptually salient properties, such as curved versus straight contours, irrespective of large variations in the configurations of the shapes. It should be noted that our monkeys were not trained to categorize the groups of shapes. Thus, the separations we found cannot be induced by category learning, at least during the experiment, but reflect the normal, unbiased state of the system. Except for the distinction between irregular complex and simple, the separations of these shape categories are not present when computing shape similarity using pixel-based or wavelet-based measures or HMAX C2 layer output, but they are consistent with the even greater separations obtained using the data of an unsupervised sorting task in human subjects.

Second, in agreement with human psychophysical data, the cells were more selective to NAP differences than to differences in the configuration of convexities and concavities. This was true both for the various NAP differences present in the regular groups and the distinction between curved and straight edges in the irregular shapes. Third, the response selectivity of IT neurons correlates with a pixel- or wavelet-based measure of image similarity for the 'irregular' shapes but less so for 'regular' shapes differing in NAPs. In general, our results demonstrate clearly that IT response selectivity correlates better with *perceived* similarities than with low-level image similarities.

The three main points of the study will now be further discussed. The MDS, as well as the separation probabilities analysis, indicated that IT neurons separate the different groups of shapes. The Lades and the HMAX model distances provide a better account of these IT neural sensitivities than the pixel-based ones. This is to be expected since the former models incorporate further processing of the raw pixel intensity input which is, to some degree, similar to that performed by some types of neurons in early visual areas. Nevertheless, they only come near the neural separations in 2 of the 6 cases, e.g. when separating the irregular complex from the irregular simple shapes. The HMAX C2 layer is somewhat better at separating these two groups of shapes than the Lades model, probably because HMAX has additional processing stages. None of the image-based models could separate the regular from the irregular groups, as the neurons and humans did. Also, none of the image-based models were able to reflect the marked separation between shapes with curved versus straight contours evident in shape sorting by humans, although this separation was well characterized by the neurons.

Still, the separation probabilities for the IT neuron population were rather low, i.e. 75% at most. This is expected, however,

Table 3
Separation probabilities for each combination of the four stimulus groups

Pair	Neural modulation	Pixel-based differences	Lades model	Hmax	Sorting task
R-IC	0.72	0.48	0.56	0.55	0.99
R-ISC	0.75	0.49	0.55	0.57	0.97
R-ISS	0.65	0.50	0.54	0.54	0.79
IC-ISC	0.67	0.49	0.61	0.67	0.99
IC-ISS	0.62	0.49	0.55	0.63	0.99
ISC-ISS	0.63	0.48	0.50	0.52	0.99

since IT neurons can be tuned to other shape features than those that differentiate systematically the different groups. An analogy can be made to V1 neurons, which are tuned to both spatial frequency and orientation. Presenting gratings of different spatial frequency and orientation will also not produce complete separations of the within- and between-group distances for different orientation groups since the neuronal distances between two gratings of the same orientation but differing in spatial frequency (within-group) can be as substantial as distances between gratings of different orientations (between-group). Because orientation and spatial frequency tuning are separable (i.e. the same orientation and spatial frequency preferences at different spatial frequencies and orientations, respectively; Mazer *et al.*, 2002), an MDS analysis would easily separate ‘groups’ of gratings according to their orientation in a manner that is virtually identical to how the MDS analysis was able to separate our groups.

The possibility of completely separating the groups by extracting orthogonal dimensions from the data suggests that IT neurons — to continue the analogy with the orientation and spatial frequency tuning of V1 neurons (Mazer *et al.*, 2002) — may code for different aspects of a shape in a separable manner. For example, the MDS analyses of the IT population responses to the different groups of stimuli indicate that the neurons are sensitive to differences in a shape property that at least covaries with aspect ratio (vertical dimension in Fig. 6) and are also sensitive to the feature(s) that separate the two groups of shapes (horizontal dimension in Fig. 6). Thus these two (sets of) shape properties seem to be encoded independently. Neurons that respond strongly to the shapes of row 2 (Fig. 2) and less so for the shapes in row 3 in the regular group will show a similar ‘aspect ratio’ preference for the irregular groups. Analogously, a separation between, for example, the curved and the straight shapes will arise if at least most of the neurons show a general preference for either curved or straight shapes, with the selectivity reduced or even absent (but not reversed), depending on other shape aspects.

The linear separability provides a basis for the categorization of the groups of shapes which, as shown by our sorting experiment, humans spontaneously do. The partitioning of different items that are linearly separated in a low-dimensional space is computationally straightforward and could easily be accomplished by appropriately weighting the connections to the next layer. Of course the actual mapping of the partitioning onto response categories is likely to occur not in IT cortex, but in regions such as prefrontal cortex or the striatum (e.g. Vogels, 1999; Ashby and Ell, 2001; Vogels *et al.*, 2002; Freedman *et al.*, 2003; Tanaka, 2003).

Vogels *et al.* (2001) and Kayaert *et al.* (2003) found that IT neurons are more sensitive to the difference between simple volumetric primitives or object parts that differ in a nonaccidental property than in a metric property. The present results extend these findings to two different kinds of two-dimensional shapes, simple geometric shapes and Fourier descriptor based shapes.

It is tempting to attribute the entire enhanced sensitivity to the shape differences within the regular group and between the straight and curved simple irregular groups to NAPs. However, other factors need to be taken into account. It is possible that already at earlier levels of the visual system, differences among, for example, the regular shapes are more salient than those among the irregular shapes. Indeed, simulations with the

wavelet-based Lades model suggested that V1-like units might already respond more selectively to the regular than to the irregular shapes that we used. The HMAX simulations show a similar trend, which is expected given that the first layer in this model also does orientation- and spatial-frequency-dependent spatial filtering. The regular shapes contain longer smooth, oriented contours than the irregular shapes and thus will stimulate end-free V1 neurons optimally. The irregular shapes have shorter line segments that vary in curvature or corners and thus will stimulate end-stopped V1 neurons more optimally than end-free neurons (Dobbins *et al.*, 1987; Versavel *et al.*, 1990). The members of a pair of irregular shapes differ mostly in the configurations of their short curved or straight segments, and thus end-stopped neurons are likely to contribute more to signaling the irregular shape differences than do end-free neurons. Neither the Lades model nor the HMAX model incorporate end-stopped receptive fields and thus both will underestimate the neural distances between the irregular shapes compared to the regular shapes at the level of V1. It is thus possible that the greater selectivity for the regular shapes originates at levels in between V1 and IT, e.g. in V2 (Hegd  and Van Essen, 2000) or V4 (Pasupathy and Connor, 1999, 2001, 2002), or in IT itself. In fact, the presence of a regular shape advantage for the line drawing stimuli (Fig. 1), which were equated using the wavelet system, supports such a suggestion.

As noted in the Introduction, in addition to their NAP differences, the shapes within the regular group differ from those in the irregular groups in a number of respects. The regular shapes are simpler than those from any of the irregular groups, which could make their features, and perhaps differences among these features, more salient. The regular stimuli are not only symmetric, they also contain larger sections of smooth contour (which may be more visible and salient) and fewer distinct contour sections (perhaps contributing to salience through less competition for perceptual resources). However, if salience due to simplicity (directly or indirectly) is an underlying factor, it is not necessarily a ‘more saliency more modulation’ effect. Such an effect would predict that modulation within the simple irregular would be greater than that within the complex irregular groups, which was not the case. Also the congruence coefficients as well as the graphs in Figure 8 show that the modulations to the regular shapes are not just larger but also less related to the pixel differences than those of the irregular shapes, suggesting that the regular and irregular shapes are coded in a different manner.

The higher discriminability between the curved and straight irregular shapes (ISC/ISS pairs) confirms the importance of NAPs in determining sensitivity. However, other factors such as number and consistency of contour differences might contribute to this higher sensitivity in humans and perhaps in IT neurons. Indeed, although both the configurational and curvature changes were spread over the entire contour of the shapes, and the configurational changes were much larger physically, the perceptually salient change of the different vertices in the curved versus straight condition produces a greater number of discrete differences. Also the fact that the changes are ‘all curved’ versus ‘all straight’ could very well have made the curvature changes even more salient.

Consistent with the psychophysical observations on similar Fourier boundary based shapes and faces (Biederman and Kalocsai, 1997; Biederman and Subramaniam, 1997), there was a positive linear correlation between the neural distances and

the Lades-based distances for the irregular shapes. This correlation was much lower for the regular shapes. Note that the feature-based HMAX model showed the opposite behavior as the other similarity measures, i.e. a better correspondence to the regular compared to the irregular shapes. This discrepancy is not so surprising as HMAX gives high weight to feature differences, but not to the relative placing of features within the stimuli (Riesenhuber and Poggio, 1999). Whatever the underlying mechanisms of the correspondence between the different image similarity measures and the neural modulations, our analysis demonstrates that the relative success of the three models in predicting IT neural modulations depends on the sorts of shapes and/or the properties by which they differ.

The present investigation is consistent with prior research in demonstrating that *in the absence* of differences in NAPs, image-based measures of similarity, i.e. pixel or Gabor jet measures, are correlated with both neural and psychophysical similarity (e.g. Op de Beeck *et al.*, 2001). However, when shapes are distinguished by NAPs, the latter will override the image-based similarity measures in accounting for both the neural tuning and perceptual behavior. The representation of shape in IT thus exceeds a mere faithful representation of physical reality, emphasizing instead perceptually salient features relevant for essential categorizations.***

Notes

The technical help of M. De Paep, P. Kayenbergh, G. Meulemans, G. Vanparijs and W. Depuydt is gratefully acknowledged. K. Okada, M. Kouh and M. Nederhouser assisted with the image scaling, and thanks to H. Op de Beeck for discussions. Supported by Human Frontier Science Program Organization RG0035/2000-B (R.V. and I.B.), and Geneeskundige Stichting Koningin Elizabeth (R.V.), and James S. McDonnell Foundation 99-53 (I.B.).

Address correspondence to Rufin Vogels, Laboratorium voor Neuro- en Psychofysiologie, KU Leuven Medical School, Onderwijs en Navorsing, Campus Gasthuisberg, B3000 Leuven, Belgium. Email: rufin.vogels@med.kuleuven.ac.be.

References

Ashby FG, Ell SW (2001) The neurobiology of human category learning. *Trends Cogn Sci* 5:204-210.

Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115-147.

Biederman I (1995) Visual object recognition. In: *An invitation to cognitive science*. Vol. 2: Visual cognition, 2nd edn (Kosslyn SM, Osherson DN, eds), pp. 121-165. Cambridge, MA: MIT Press.

Biederman I, Bar M (1999) One shot-viewpoint invariance in matching novel objects. *Vision Res* 39:2885-2899.

Biederman I, Gerhardstein PC (1993) Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J Exp Psychol Hum Percept Perform* 19:1162-1182.

Biederman I, Kalocsai P (1997) Neurocomputational bases of object and face recognition. *Philos Trans R Soc Lond B Biol Sci* 352:1203-1219.

Biederman I, Subramaniam S (1997) Predicting the shape similarity of objects without distinguishing viewpoint invariant properties (VIPs) or parts. *Invest Ophthalmol Vis Sci* 38:998.

Biederman I, Subramaniam S, Bar M, Kalocsai P, Fiser J (1999) Subordinate-level classification reexamined. *Psychol Res* 62:131-153.

Borg I, Leutner D (1985) Measuring the similarity of MDS configurations. *Multivar Behav Res* 20:325-334.

Cooper EE, Subramaniam S, Biederman I (1995) Recognizing objects with an irregular part. *Invest Ophthalmol Vis Sci* 36:473.

Dean P (1976) Effects of inferotemporal lesions on the behavior of monkeys. *Psychol Bull* 83:41-72.

Dobbins A, Zucker SW, Cynader MS (1987) Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature* 329:438-441.

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

Green DM, Swets JA (1974) Signal detection theory and psychophysics. New York: Robert E. Krieger.

Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophysiol* 35:96-111.

Hahn U, Ramsar M (2001) Similarity and categorization. Oxford: Oxford University Press.

Hegd  J, Van Essen D (2000) Strategies of shape representation in macaque visual area V2. *J Neurosci* 20(RC61):1-6.

Janssen P, Vogels R, Orban GA (2000) Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288:2054-2056.

Kayaert G, Biederman I, Vogels R (2003) Shape tuning in macaque inferior temporal cortex. *J Neurosci* 23:3016-3027.

Kimia BB (2003) On the role of medial geometry in human vision. *J Physiol Paris* 97:155-190.

Koffka K (1935) Principles of gestalt psychology. London: Lund Humphries.

Lades M, Vorbruggen JC, Buhmann J, Lange J, von der Malsburg C, Wurtz RP, Konen W (1993) Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans Comput* 42:300-311.

Logothetis NK, Pauls J, B lthoff HH, Poggio T (1994) View-dependent object recognition by monkeys. *Curr Biol* 4:401-414.

Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant, JL (2002) Spatial frequency and orientation tuning dynamics in area V1. *Proc Natl Acad Sci USA* 99:1645-1650.

Op de Beeck H, Wagemans J, Vogels R (2001) Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat Neurosci* 4:1244-1252.

Pasupathy A, Connor CE (1999) Responses to contour features in macaque area V4. *J Neurophysiol* 82:2490-2502.

Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86:2505-2519.

Pasupathy A, Connor CE (2002) Population coding of shape in area V4. *Nat Neurosci* 5:1332-1338.

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

Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415:318-320.

Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109-139.

Tanaka K (2003) Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb Cortex* 13:90-99.

Versavel M, Orban GA, Lagae L (1990) Responses of visual cortical neurons to curved stimuli and chevrons. *Vision Res* 30:235-248.

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

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.

Vogels R, S ry G, Dupont P, Orban G (2002) Human brain regions involved in visual categorization. *Neuroimage* 16:401-414.

Young MP, Yamane S (1992) Sparse population coding of faces in the inferotemporal cortex. *Science* 256:1327-1331.