nature neuroscience

Underlying principles of visual shape selectivity in posterior inferotemporal cortex

Scott L Brincat & Charles E Connor

Object perception depends on shape processing in the ventral visual pathway, which in monkeys culminates in inferotemporal cortex (IT). Here we provide a description of fundamental quantitative principles governing neural selectivity for complex shape in IT. By measuring responses to large, parametric sets of two-dimensional (2D) silhouette shapes, we found that neurons in posterior IT (Brodmann's areas TEO and posterior TE) integrate information about multiple contour elements (straight and curved edge fragments of the type represented in lower-level areas) using both linear and nonlinear mechanisms. This results in complex, distributed response patterns that cannot be characterized solely in terms of example stimuli. We explained these response patterns with tuning functions in multidimensional shape space and accurately predicted neural responses to the widely varying shapes in our stimulus set. Integration of contour element information in earlier stages of IT represents an important step in the transformation from low-level shape signals to complex object representation.

Past research has shown that IT neurons respond selectively to complex shape features or whole objects[1–3], carry shape information at the columnar and population levels[4–8], and are modulated by shape-related behavioral tasks and learning[9–12]. Despite these significant advances, the basic principles governing complex shape selectivity in IT remain to be determined. IT neurons respond differentially to complex objects, but what information are they conveying about the shapes of these objects? Can IT responses across a broad range of stimuli be explained by tuning in shape-related dimensions, in the same way that higher-level dorsal pathway processing has been characterized in spatial and motion-related dimensions? How are complex IT responses derived from simpler, lower-level inputs? These questions can only be answered by systematic, quantitative characterization of the principles underlying neural selectivity for shape in IT.

The main obstacle to quantifying IT neural selectivity is the extreme complexity of shape—the variety of possible shapes is virtually infinite, the neural representation of shape must necessarily be high-dimensional, and the nature of the dimensions is unknown. In the face of this complexity, previous investigators have been forced to rely on small numbers of example stimuli to support qualitative descriptions of IT shape selectivity. Here we attempt to explain one aspect of IT responses—selectivity for 2D boundary shape—at a quantitative level. We focused on more posterior, earlier processing stages in IT (Brodmann's areas TEO and posterior TE), where response properties are less abstract and experience-dependent, and therefore may be more analytically tractable. Our approach differed in two important respects from previous investigations. First, we used large stimulus sets (~1,000 stimuli) in which shape characteristics varied parametrically, to provide a rich and systematic dataset for quantitative analysis. Second, we fitted neural responses with tuning functions on a multidimensional shape domain in order to capture the full complexity of IT shape tuning. This approach yielded accurate predictions of neural responses and a straightforward explanation of boundary shape selectivity in posterior IT: neurons integrate information about multiple contour elements, presumably as a step toward more complete object representations.

RESULTS

To provide adequate data for mathematical analysis of shape selectivity, we needed a stimulus set covering a wide range of shapes but with quantifiable gradients of shape similarity between stimuli. To achieve this, we defined convex, straight and concave contour elements at specific orientations and object-relative positions, and crossed these basic geometric elements in a large permutation matrix of smooth, closed 2D silhouette shapes (**Fig. 1a** and **Supplementary Fig. 1** online). The factorial structure of this stimulus set allowed us to distinguish between tuning for simple and complex shape elements, and to measure nonlinearities in responses to conjunctions of elements. A simpler version of this factorial contour element approach to stimulus construction was effective in characterizing boundary shape coding in area V4 (ref. 13). We found the approach to be effective in IT as well—109 out of 137 cells (80%) showed significant selectivity within our main stimulus set ($P < 0.001$, ANOVA), and the average maximal response of these 109 cells was 30 spikes/s (63 spikes/s for the transient component of the response). The precise curvatures of the contour elements were varied in auxiliary tests on selected shapes (**Fig. 1b**); stimulus orientation, size and position were also sampled at finer resolution in other auxiliary tests (see **Supplementary Note** online).

The responses of the neuron in **Fig. 1a** (background gray levels; see scale bar) were typical in that activity was broadly distributed in

Zanvyl Krieger Mind/Brain Institute and Department of Neuroscience, Johns Hopkins University, Baltimore, Maryland 21218, USA. Correspondence should be addressed to C.E.C. (connor@jhu.edu).

Published online 4 July 2004; doi:10.1038/nn1278

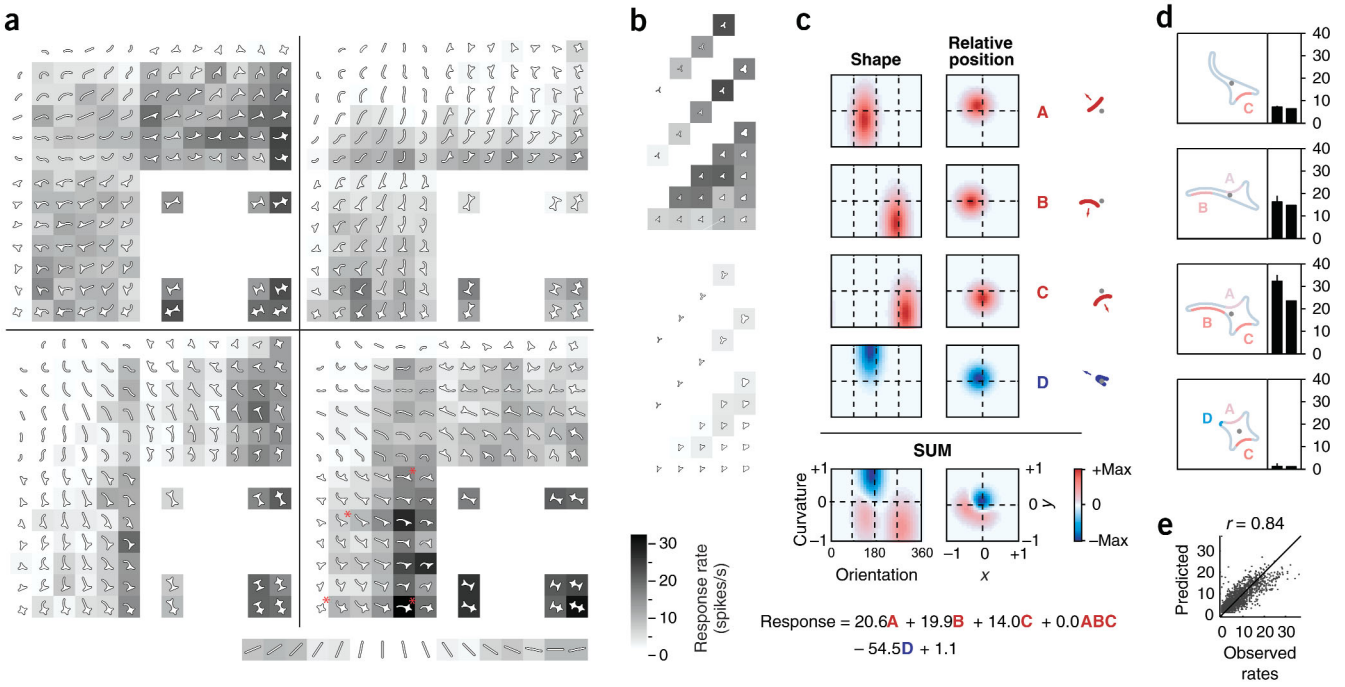

Figure 1 Quantitative analysis of shape selectivity for a typical IT neuron. (**a**) Main stimulus set. The shape stimuli (white icons) were generated by systematically combining convex, straight and concave contour elements at varying orientations and positions (see Methods and **Supplementary Note** online). Background gray levels indicate this cell's mean response to each shape (see gray level scale bar). In the experiments, stimuli were displayed as silhouettes against a blank background; outlines and background squares are shown here for illustrative purposes only. (**b**) Auxiliary stimulus set used to test fine-scale tuning for local curvature (conventions as in **a**). (**c**) Optimal model of this cell's shape selectivity, based on its response profile across 1,687 different stimuli (see **Supplementary Note** online). The four Gaussian subunits, three excitatory (red, A–C) and one inhibitory (blue, D), are each depicted by two plots, showing tuning for shape (curvature × orientation) and for position. For each subunit, the contour element corresponding to the Gaussian peak is shown at the right, positioned relative to stimulus center of mass (gray dot). The subunits are combined into one set of plots at the bottom (SUM). For each stimulus, the predicted response derives from the positions of its constituent contour elements on the Gaussian tuning functions. The resulting excitatory and inhibitory subunit values are combined according to the equation at the bottom, which, for this cell, indicates purely linear summation of contour elements (0 weight on nonlinear interaction term ABC). (**d**) Subset of main test stimuli (corresponding to asterisks in **a**) illustrating model predictions. In the stimulus icons, the color of each segment represents its proximity to the Gaussian peak indicated by the adjacent letter. Bar plots show observed (left; mean ± s.e.m.) and predicted (right) responses. (**e**) Scatterplot showing correlation between observed and predicted response rates for 1,687 stimuli.

a complex pattern not defined by any single optimal shape feature. The complex response pattern was not due to noise, as response rates exceeded 30 spikes/s and average standard error was only 1.4 spikes/s. Moreover, as with most cells we tested, shape selectivity was consistent across substantial variations in retinotopic position and size (see below). To quantify shape selectivity, we first defined each contour element in terms of its shape (curvature and orientation) and *x*,*y* position. Selectivity for single or multiple contour elements was characterized by fitting each cell's response pattern with a tuning function composed of 1–6 Gaussian subunits on the shape × position domain (**Fig. 1c**; see Methods). Each subunit represents a range of contour element shapes and positions that have an excitatory (red) or inhibitory (blue, see scale bar) effect on responses. For this cell, subunit A represents the excitatory effect of concave contour elements (negative curvatures), oriented toward the upper left (~135°) and positioned at the upper left relative to the object center, as illustrated by the icon to the right, which corresponds to the Gaussian peak. Subunit B represents concavities oriented toward the bottom and positioned to the left. Subunit C represents concavities oriented toward the lower right and positioned below object center. Subunit D, the only inhibitory factor in this model, represents sharp convexities oriented toward and positioned near the upper left. The full tuning function (summed across all four subunits) is shown at the bottom (**Fig. 1c**, SUM), together with the equation describing how subunit responses are integrated. Our model equations included a sum of the individual subunits (A, B, C and D) and products of collected same-sign subunits (ABC), in order to account for simple additive (linear) integration of multiple contour elements as well as nonlinear interactions between elements that generate supra-additive responses. (Essentially equivalent results were obtained with models that included the three pair-wise nonlinear interaction terms; see Methods.)

The stimuli in **Figure 1d** (corresponding to asterisks in **Fig. 1a**) illustrate how this model captures the neuron's responses. The first (top) stimulus contains only one contour element near any of the model's excitatory tuning peaks—a concavity that falls on the shoulder of subunit C (proximity to peak indicated by color; see scale in **Fig. 1c**). The predicted response (based on the 'C' term in the model equation) is relatively weak, similar to the observed response (bar plot at right). The second stimulus contains A- and B-type excitatory contour elements; the combination of these factors in the equation accurately predicts a moderate response. The third stimulus contains all three types of excitatory elements, and evokes a strong response roughly equal to the sum of the individual element responses. The additive effects of the excitatory elements are reflected in the model equation, where the linear terms (A, B and C) have strong weights but the nonlinear interaction term (ABC) has a weight of zero. The fourth stimulus exemplifies how responses to excitatory elements are sup-

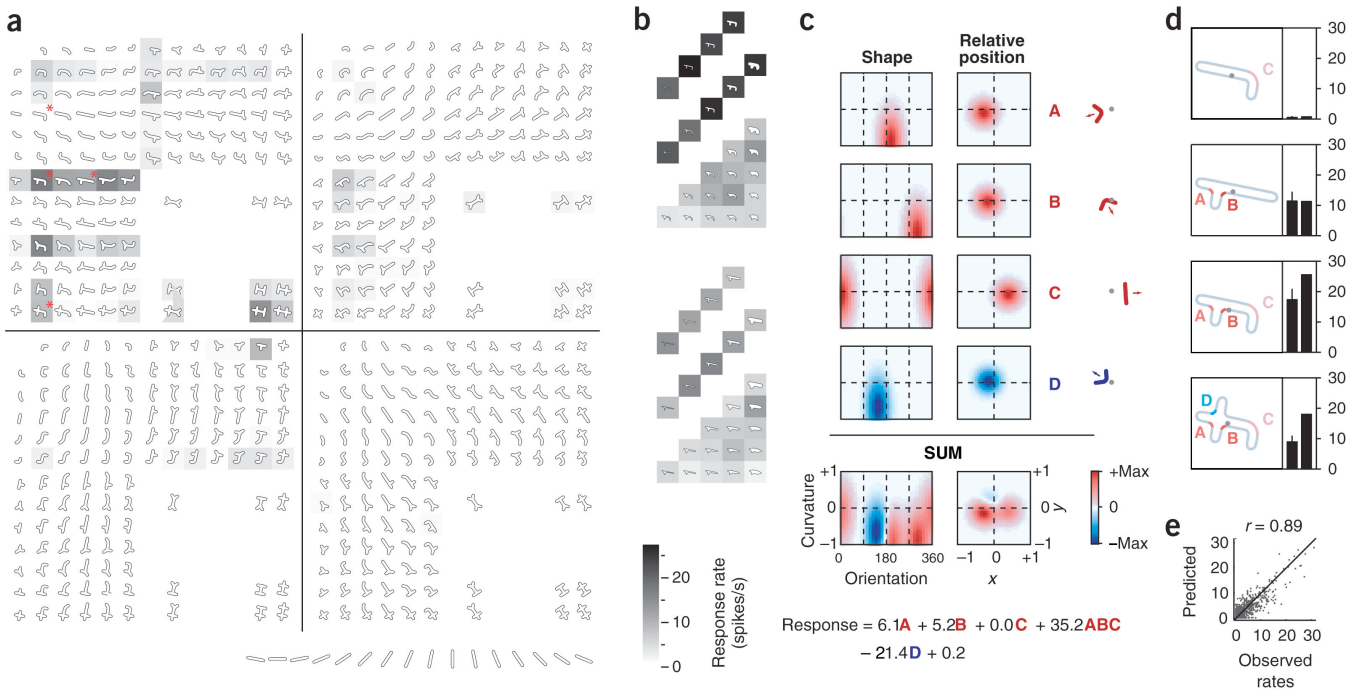

Figure 2 Quantitative analysis of shape selectivity for a highly selective IT neuron (conventions as in **Fig. 1**). (**a**) Main stimulus set. (**b**) Auxiliary test of local curvature tuning. (**c**) Optimal model of this cell's shape selectivity, based on its response profile across 1,882 different stimuli. The model equation is notable for its strong nonlinear term (ABC). (**d**) Subset of main test stimuli illustrating model predictions. (**e**) Scatterplot showing correlation between observed and predicted response rates for 1,882 stimuli.

pressed by the inhibitory convexity (D), as predicted by the negative term in the model equation. The linear summation of excitatory and inhibitory contour element effects in this model provides a relatively simple and accurate (cross-validation $r = 0.84$; **Fig. 1e**) explanation of the neuron's complex response pattern across the main stimulus set (**Fig. 1a**) and other stimuli (1,687 stimuli total). This model fit the data significantly better than simpler models ($P < 10^{-6}$, partial F test versus best-fitting three-subunit model).

Some IT cells had sparse response patterns limited to a few stimuli with clear shape similarity (as in **Fig. 2**). This neuron's responses were best fit by a highly nonlinear model (**Fig. 2c**) with a large excitatory interaction term (ABC) representing co-occurrence of adjacent sharp concavities oriented toward the lower left (A) and lower right (B), both positioned to the left, and with shallow-to-flat curvature oriented toward and positioned near the right (C). The conjunction of these contour elements defines the quadruped-like shapes that elicited robust responses from this cell. The model also features inhibitory tuning for concave elements oriented toward and positioned near the upper left (D). The first two stimuli (top) in **Figure 2d** contain only subsets (C and A/B, respectively) of the excitatory contour elements, and elicited relatively weak responses, as predicted. The third stimulus, containing all three excitatory elements, evoked a response greater than the sum of the individual element responses, as predicted by the large nonlinear interaction term (ABC) in the model. The fourth stimulus illustrates the inhibitory effect of the D-type concavity. Again, the model accurately predicted responses to a large number of stimuli (1,882; cross-validation $r = 0.89$; **Fig. 2e**) and provided a significantly better fit than simpler models ($P < 10^{-6}$, partial F test versus best-fitting three-subunit model).

Similar models, based on linear-nonlinear integration of 1–6 subunits, were used to quantify shape selectivity for 109 neurons in posterior IT. In a cross-validation analysis (used to prevent overfitting; see Methods) the mean correlation between predicted and observed responses was 0.70 (**Fig. 3a**), showing that IT responses to 2D silhouette shapes can be accurately predicted by models based on tuning for parts-level contour information. Most cells were best characterized by tuning functions with 2–4 subunits (**Fig. 3b**), indicating that IT cells integrate information about multiple contour elements, possibly by combining lower-level inputs from area V4. Most cells were inhibited by some range of contour elements[14], and for these cells excitatory and inhibitory influences were roughly equal in magnitude (**Fig. 3c**). Our index of nonlinearity for excitatory contour element integration (**Fig. 3d**, red bars) indicated two distinct subpopulations of neurons, one showing purely linear summation (*e.g.*, **Fig. 1**), and one showing partially or largely nonlinear integration (*e.g.*, **Fig. 2**). In contrast, integration of inhibitory elements was almost exclusively linear (**Fig. 3d**, blue bars). This integration scheme corresponds closely to what would be expected from a linear-nonlinear cascade model (for example, linear summation of contour signals, followed by nonlinear rectification at a spiking threshold value that varies between neurons). Thus, our more general models point toward a specific, biologically plausible mechanism analogous to successful models of neural summation in other domains[15,16]. Nonlinearity was also significantly correlated with response sparseness across IT neurons (**Fig. 3e**)—linear cells typically responded to a broad range of shapes (relatively low sparseness), whereas nonlinear cells tended to respond to only a few shapes (relatively high sparseness).

Shape tuning, in the sense of relative responses to different shapes, was consistent across changes in stimulus position and size, in agreement with previous reports[17,18]. (Absolute response rates can change markedly with position and size, as IT neurons typically have finite receptive fields and some degree of size tuning[1,17–19].)

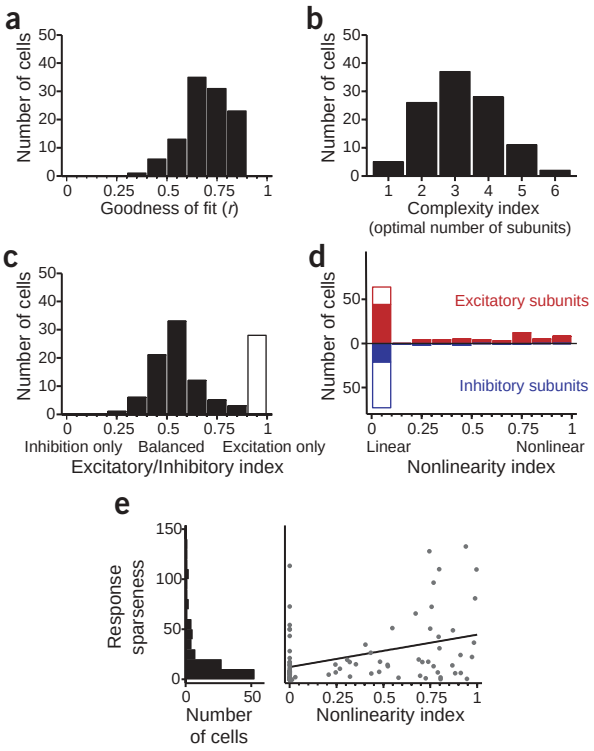

Figure 3 Population summary of IT shape selectivity. (**a**) Goodness of fit (correlation between observed and predicted responses based on a cross-validation procedure described in **Supplementary Note** online) for optimal models of shape selectivity for 109 IT neurons. (**b**) Tuning complexity (numbers of Gaussian subunits in optimal models). (**c**) Excitatory/inhibitory balance (see Methods). Optimal models for some cells indicated excitatory effects only (open bars). Optimal models for the majority of cells indicated combined excitatory and inhibitory effects (filled bars). (**d**) Linearity/nonlinearity of contour element integration (see Methods). Values for excitatory (red) and inhibitory (blue) factors are shown separately. Optimal models for some cells included only a single subunit of a given sign (open bars) and were necessarily linear. Optimal models including multiple excitatory subunits (filled red bars) split into distinct groups ($P = 0.002$, Hartigan's dip test for bimodality) showing linear and nonlinear integration, respectively. Models including multiple inhibitory subunits (filled blue bars) were mainly linear. (**e**) Sparseness of responses to main test stimuli as a function of nonlinearity (for excitatory integration). Linear regression (line) indicates a significant correlation between sparseness and nonlinearity ($r = 0.39$; $P = 0.0001$, permutation test).

This is exemplified for the cell in **Figure 2** by three stimuli (drawn from the top, middle and bottom of the cell's response range) tested across a 5 × 5 grid of positions (**Fig. 4a**) and a six-octave range of sizes (**Fig. 4b**; see **Supplementary Fig. 2** online for similar results for the cell in **Fig. 1**). To quantify such consistency, we tested all cells in a similar manner and calculated an index of position/size consistency based on separability of tuning in the shape and position/size dimensions (see **Supplementary Note** online). The distributions of index values indicate robust consistency of shape tuning across position (**Fig. 4c**) and size (**Fig. 4d**), within the position and size ranges to which neurons were responsive.

Previous studies have demonstrated significant changes over time in IT shape selectivity for small sets of highly familiar, behaviorally relevant stimuli[9,11,12]. In our experiments the number of stimuli was exceptionally large, and the monkeys were never required to perform a behavioral task involving the stimuli, but they were passively exposed to the stimuli over a period of months. If passive exposure produced the tuning properties we observed, model fits should have improved over time during the recording period, especially in the case of one monkey for whom the stimuli were novel at the start of experiments. However, we found no significant correlation between recording session number and model goodness of fit ($r = -0.01$, $P = 0.53$ for the experienced monkey; $r = 0.11$, $P = 0.29$ for the naive monkey; permutation test), nor were there any apparent nonlinear relationships. In any case, even the most complex tuning functions we described were still of such a general nature that they would encompass a wide range of natural stimuli. Thus, we postulate that the response properties we observed reflect stable, long-term mechanisms for general shape representation.

DISCUSSION

Our analyses provide the first quantitative explanation of neural shape tuning properties in IT cortex. We accurately predicted responses of neurons in posterior IT to 2D silhouette shapes with tuning function models that are computationally analogous (though with different input and output dimensions) to successful models of processing in other cortical areas. Computational analyses of neural responses in primary visual cortex (V1) to large, complex stimulus sets have shown how V1 neurons integrate retinotopic luminance information to generate explicit signals for local orientation and spatial frequency[15,20–22]. Our analyses show that IT cells integrate information about the shapes and relative positions of multiple contour elements (of the type represented at lower levels in the ventral pathway like V4; ref. 13) to generate explicit signals for object part relationships. This result lends general support to parts-based shape representation theories[23–29]. It is particularly consistent with those that posit explicit (as opposed to implicit) coding of structural relations between parts[23–25], and those based on graded, metric tuning (as opposed to categorical selectivity) for part shape and position[25–27,29]. Our findings speak only to the issue of 2D boundary shape coding for abstract objects, and not to more complex issues (realistic objects, natural images, internal detail, color and contrast variation, 3D shape, cue invariance, scene segmentation, attention and learning) addressed by previous studies[1–6,8–12,30]. However, 2D boundary shape may dominate IT responses to realistic objects[31]. Further, it should be possible to extend our analytical methods to other aspects of shape coding.

Some IT cells were highly selective for specific part combinations, consistent with previous descriptions of IT selectivity based on critical features[2,8]. Other cells exhibited complex response patterns, distributed across the stimulus set, that could not be reduced to a single optimal shape feature. We found that these different types of response profiles could be accurately explained by models differing in linearity of contour signal integration—linear integration predicted the distributed response profiles, whereas nonlinear integration predicted the sparse, combination-selective profiles. It has been proposed that achieving a sparse representation, in which a relatively small number of neurons are active at any time, is one goal of sensory processing—sparse coding is metabolically efficient, facilitates pattern matching, and increases memory storage capacity[32–34]. Theoretical[33] and empirical[34] work has suggested that the nonlinear response properties of V1 cells are instrumental in generating their relatively sparse representation. Our results suggest that nonlinear integration in IT cells might act to further increase the sparseness of the ventral pathway shape representation en route to areas critical for object memory.

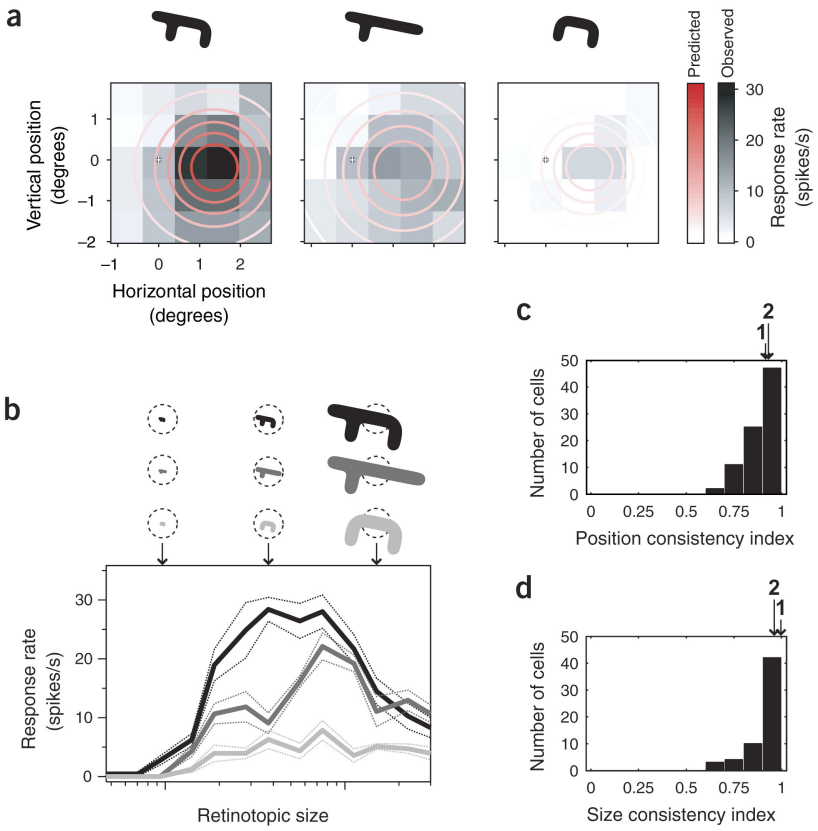

Figure 4 Consistency of shape selectivity across position and size changes. (**a**) Responses of the example neuron in **Figure 2** to three shapes spanning its response range, presented at a 5 × 5 grid of retinotopic positions. Gray level indicates actual responses; overlaid red contours show response predictions based on this neuron's optimal model. The icon above each gray level plot shows the shape used to generate it. Crosses represent center of gaze. (**b**) Responses of the **Figure 2** neuron to three shapes across a six-octave range of sizes. The icons above the plot show stimulus sizes, relative to estimated receptive field (dashed circle), at three selected points within the tested range. The gray level of each size-tuning function corresponds to the gray level of the corresponding shape icon above. Dotted lines show standard error of the mean. (**c**) Population distribution of position consistency index values (see **Supplementary Note** online). Arrows labelled '1' and '2' indicate values for the **Figure 1** and **2** example neurons, respectively. (**d**) Population distribution of size consistency index values.

We observed that IT neurons integrate highly specific information about the curvatures, orientations and positions (in both retinotopic and object-relative coordinates) of multiple (typically 2–4) contour fragments, resulting in complex shape selectivity and thus providing signals useful for high-level object representation. This phenomenon is different from previously described V1 facilitation and suppression effects that depend in a general way on collinearity with contours outside the classical receptive field and do not produce specific selectivity for complex shapes. These V1 collinearity effects, which peak at an inter-element separation of less than 1 degree and are predominantly suppressive at supra-threshold contrasts[35,36], serve to modulate V1 orientation signals depending on local contour continuity and orientation contrast. Contour representation at early (V1) and intermediate (V4)[13] stages is considerably simpler than what we have observed in IT.

The basic finding of our study is that IT neurons integrate information about the shapes and relative positions of contour elements. We quantified local contour shape with simple geometric dimensions known to be fundamental to visual coding: orientation and curvature (the first and second derivatives of contour position). Orientation is strongly represented throughout visual cortex; curvature is strongly represented at intermediate levels of the ventral pathway[13,37] and has been identified psychophysically as a basic dimension of visual perception[26,38]. We measured contour position in both retinotopic and object-centered coordinates. All of these dimensions were critical for describing the response patterns of most neurons, although tuning for retinotopic position was broadest and weakest (see **Supplementary Fig. 3** online). Selectivity for retinotopic position gradually decreases through successive stages of the ventral pathway, though selectivity for object shape is consistent across the larger receptive fields found at higher levels[1,17–19] (**Fig. 4**). Recent evidence[13,25,39–41] suggests that this is accomplished through a transformation to object-centered coordinates, producing an explicit representation of part relationships independent of retinotopic position.

Thus, the shape and position dimensions in our analysis were chosen for their simplicity and previously established relevance in visual coding. Essentially equivalent results would have been obtained using any other simple geometric description of contour element shape and position. Tuning in more complex spaces would manifest as distinctive patterns in these simpler dimensions. For example, it has been proposed that IT neurons could be tuned for global periodic variations in boundary orientation ('Fourier descriptors')[42]. In our analysis, such tuning would show up as patterns of alternating convex and concave peaks in the tuning function. We did not observe any such patterns, however (consistent with other negative results for the Fourier hypothesis[43]), nor did we detect any other common patterns that would signify tuning in more complex dimensions.

Other computational studies of IT have addressed different questions, such as the amount of information available in IT responses at the single-cell[44] and population[6,45] levels. Multidimensional scaling has been used to analyze the arrangement of shape stimuli in the space defined by IT population responses[5,7]. These studies have helped to characterize the general nature of IT shape information, whereas ours is the first quantitative study to analyze the specific informational content of IT responses.

The parts integration mechanisms we have described in posterior IT could culminate with holistic coding in anterior IT, that is, with each neuron representing information about global object shape[3,27,46]. Alternatively, parts-level population coding may itself be the ultimate form of neural shape representation, and may just become increasingly complex and sparse at higher levels[8,29,47]. We found no evidence for holistic coding in our data—most posterior IT cells conveyed information about only a few contour elements, and even the most selective cells in our sample responded to a variety of global shapes (*e.g.*, **Fig. 2** and **Supplementary Fig. 1**). Many of our neurons were tuned for object-relative position of contour elements, which does depend on at least crude global information (*e.g.*, lumi-

nance integration to derive center-of-mass). By 'holistic' shape coding, however, we and other authors[12] mean specific shape information that could support object recognition and discrimination. Center-of-mass is not specific (because many different shapes could have the same center-of-mass) and could not by itself support recognition. Holistic shape coding may evolve only for familiar or behaviorally relevant shapes[12]. A complete understanding of shape representation will depend on quantitative analysis at multiple cortical levels under multiple sensory and behavioral conditions.

METHODS

Behavioral and neurophysiological methods. We recorded spike activity of well-isolated single neurons in the left hemispheres of two awake rhesus monkeys (*Macaca mulatta*) trained to maintain fixation within a 0.75°-radius window. Eye position was monitored using an infrared eye tracker (IScan). Epoxy-coated tungsten microelectrodes (A-M Systems) were inserted through a transdural guide tube. Extracellular action potentials were isolated on-line using a dual time-amplitude window discriminator, and spike times were saved with a temporal resolution of 1 ms. Neurons were sampled in an unbiased manner from IT cortex, in the posterior to central lower bank and lip of the superior temporal sulcus (from –5 mm posterior to +8.5 mm anterior to the interaural line), corresponding to functionally defined areas PIT and CIT (cytoarchitectural areas TEO and posterior TE, respectively). IT cortex was identified based on neurons' receptive field (RF) sizes and response properties, the pattern of gray-white matter transitions during the penetration, and by reference to structural MRIs obtained from each animal prior to the recording experiments. No clear differences in tuning properties along the anterior-posterior axis were observed; all neurons were therefore grouped together in our analyses. All procedures were approved by the Johns Hopkins animal care committee and conformed to National Institutes of Health and US Department of Agriculture guidelines.

Shape stimuli. We generated parametric stimulus sets in which basic geometric elements were varied systematically and permuted into a large number of combinations (see **Supplementary Note** online). Stimuli were displayed as high luminance contrast (white, black or colored) silhouettes against a gray background (2.5 cd/m^2). Each cell was studied with the stimulus set shown (**Figs. 1a** and **2a**; 66 cells), with a modified version in which one end of each stimulus was centered in the RF (for 33 cells that were more sensitive to absolute position of object parts; **Supplementary Fig. 1** online) or, whenever possible, with both stimulus sets (10 cells). For each cell, these stimulus sets were optimized for color, orientation, local curvature, retinotopic position and size, based on preliminary tests. *Post hoc* auxiliary tests (**Figs. 1b** and **4a**,**b**), involving a selection of high- and low-response shapes from the main stimulus sets, were used to examine fine-scale tuning along those same dimensions and to confirm that shape tuning properties generalized across those dimensions. Across our sample of IT neurons, optimal retinotopic positions (RF centers) covered most of the contralateral visual hemifield, with a strong bias toward foveal and parafoveal locations (median eccentricity: 3.0°; 10th percentile: 0°; 90th percentile: 12.5°). For cells with RFs including the fovea, the 0.1° diameter fixation marker was dimmed to near-threshold luminance and occluded by any overlapping stimuli. In all tests, stimuli were flashed for 500 ms each, with a 250-ms interstimulus interval, in random order, during 4.5 s fixation trials (6 stimuli/trial).

Data analysis. For each stimulus, response rate was calculated by summing spikes over the 500-ms presentation period and averaging across 2–8 repetitions (mean: 4.1). We isolated and studied 137 cells, of which 109 (80%) showed significant selectivity ($P < 0.001$; ANOVA) across the main stimulus set. These selective cells constitute the neural sample described in this paper.

To quantify shape tuning, we first characterized the stimuli in terms of their component contour elements. This approach, used in a number of mathematical shape formulations[48–50], provides a versatile, compact, low-level geometric description with the capacity to represent more complex object parts as well as overall object shape. Contour elements of approximately constant curvature were defined by segmenting the stimulus boundary at points of rapid change in curvature. Each contour element was described by six values: mean curvature (rate of change in tangent angle per unit contour length, scaled to a range of [–1,1] using a sigmoidal mapping function[13]), orientation (of a perpendicular bisector pointing outward from the object interior), object-relative position (x,y relative to the object's center of mass) and absolute position (x,y in retinotopic coordinates; see ref. 19 for comparison; see **Supplementary Note** online). Thus, each element could be described as a point in a six-dimensional (6D) shape space, and each stimulus was represented by a set of such points corresponding to its constituent elements.

We modeled each cell's shape selectivity with a combination of excitatory and inhibitory Gaussian subunits on this 6D domain (analogous to the ON and OFF lobes in a V1 simple cell RF). For a given stimulus, the response component predicted by each subunit was the sum of Gaussian amplitude values at the points corresponding to the stimulus contour elements (inner product between Gaussian and contour element points):

$$\begin{aligned}\text{Response}_{\text{subunit}} &= A\cdot\sum_{c=1}^{\substack{\#\text{ contour}\\ \text{elements}}}\text{Gaussian}\,(k_c,\theta_c,rx_c,ry_c,ax_c,ay_c)\\ &= A\cdot\sum_{c=1}^{\substack{\#\text{ contour}\\ \text{elements}}}e^{-\left(\frac{(k_c-\mu_k)^2}{2\sigma_k^2}+\frac{(\theta_c-\mu_\theta)^2}{2\sigma_\theta^2}+\frac{(rx_c-\mu_{rx})^2}{2\sigma_r^2}+\right.}\\ &\qquad\left.\frac{(ry_c-\mu_{ay})^2}{2\sigma_r^2}+\frac{(ax_c-\mu_{ax})^2}{2\sigma_a^2}+\frac{(ay_c-\mu_{ay})^2}{2\sigma_a^2}\right)\end{aligned}\qquad(1)$$

where k_c, θ_c, rx_c, ry_c, ax_c and ay_c are the (fixed) curvature, orientation, relative position and absolute position values for each contour element in the stimulus, μ and σ are the fitted Gaussian centers and standard deviations on each of these six dimensions, and A is the fitted Gaussian amplitude (positive (excitatory) or negative (inhibitory)). Similar results were obtained when a maximum operation (using only the contour element closest to the Gaussian center) was used instead of the sum.

For each cell, we fitted models based on combinations of 1–6 Gaussian subunits, which sufficed to capture the range of response complexity we observed in IT. A model's predicted response to each stimulus was a weighted combination of the individual subunit responses (the linear component) and products of subunit responses of the same sign (the nonlinear component). By varying these weights, our models could range continuously from linear summation across contour elements to nonlinear selectivity for element combinations. In equation form, the predicted response was:

$$\text{Response} = \text{gain}\cdot\left[\sum_{s=1}^{\#\text{ subunits}}(w_sR_s)+w_{NL+}\prod_{s=1}^{\substack{\#\text{ excitatory}\\ \text{subunits}}}(R_s)-w_{NL-}\prod_{s=1}^{\substack{\#\text{ inhibitory}\\ \text{subunits}}}(R_s)\right]^{+}+b_0\qquad(2)$$

where R_s is the unweighted response predicted by each subunit alone (derived from equation 1), w_s is the fitted linear weight (amplitude) for each subunit, w_{NL+} and w_{NL-} are the fitted weights for the excitatory and inhibitory nonlinear terms, respectively, b_0 is the baseline neuronal firing rate, and the brackets represent rectification of the predicted response at 0 spikes/s. The total numbers of fitted parameters for models with 1–6 subunits were 12, 18, 23, 29, 34 and 39 respectively.

These models were fitted to each cell's response data (main and auxiliary stimulus sets) using an iterative nonlinear least-squares algorithm (Matlab, lsqnonlin function; Mathworks Inc.) to minimize the sum of squared differences between observed and predicted responses. For each neuron, we used a stepwise regression procedure to determine the minimum number of tuning subunits needed to explain responses. We calculated the additional amount of data variance explained by each additional subunit; if this was ≥2.5% of total variance, then the subunit was added and the process was repeated, up to a maximum of six subunits (beyond which the threshold increase in explained variance was never observed). This procedure was much more conservative than other criteria commonly used to constrain model complexity. All multiple subunit models described here produced a significant improvement in fit relative to models containing one less subunit ($P < 0.01$, partial F test). To evaluate model goodness of fit, we calculated the correlation between observed and predicted responses, using a standard cross-validation procedure to control for overfitting (see **Supplementary Note** online).

To characterize the relative contributions of excitation and inhibition, we calculated an excitatory/inhibitory index for each cell:

$$\text{E/I index} = \sum(\text{all excitatory weights}) \Big/ \sum(\text{all weights}) \quad (3)$$

An E/I index value of 1 indicates a lack of inhibitory influences, and 0.5 indicates that total excitation and inhibition were balanced in magnitude. To characterize the degree of linearity/nonlinearity in responses to multiple contour elements, we calculated a nonlinearity index for each cell:

$$\text{NL index} = \text{nonlinear term weight} \Big/ \sum(\text{all weights}) \quad (4)$$

This index was calculated separately for the excitatory and inhibitory terms in each model. It has a value of 0 for purely linear summation and 1 for purely multiplicative, nonlinear responsiveness to contour element combinations; values in between correspond to supra-linear (greater than additive) summation. Response sparseness was defined for each neuron as the kurtosis of the distribution of its responses across the main stimulus set; to convert observed response distributions into symmetric, unimodal distributions appropriate for calculation of kurtosis, we reflected the data about the origin (see ref. 34 for details).

We used the particular model form described above (equation 2), with one catch-all (n^{th} order) nonlinear interaction term for each sign, because it effectively and parsimoniously characterized nonlinear integration in our IT cells. Essentially equivalent results were obtained with models that included terms characterizing all 2nd order nonlinear interactions between pairs of same-sign subunits: mean goodness of fit (r) increased by only 0.005, mean optimal number of subunits was virtually identical (3.17 versus 3.18), inhibitory integration remained almost exclusively linear (mean NL_i index: 0.13 versus 0.11), linearity of excitatory integration exhibited a similar bimodal distribution (mean NL_e index: 0.09 versus 0.04 and 0.72 versus 0.74 for linear and nonlinear subpopulations, respectively), and nonlinearity was similarly correlated with sparseness (r: 0.42 versus 0.39).

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank W. Nash, W. Quinlan, B. Sorensen and L. Ziehm for technical assistance, and A. Bastian, D. Hinkle, K. Johnson and R. von der Heydt for helpful comments. Our implementation of Hartigan's dip test is based on public domain code provided by F. Mechler and D. Ringach. This work was supported by the National Eye Institute and by the Pew Scholars Program in the Biomedical Sciences.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 10 March; accepted 25 May 2004
Published online at http://www.nature.com/natureneuroscience/

1. Gross, C.G., Rocha-Miranda, C.E. & Bender, D.B. Visual properties of neurons in inferotemporal cortex of the Macaque. *J. Neurophysiol.* **35**, 96–111 (1972).
2. Kobatake, E. & Tanaka, K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* **71**, 856–867 (1994).
3. Booth, M.C. & Rolls, E.T. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex* **8**, 510–523 (1998).
4. Fujita, I., Tanaka, K., Ito, M. & Cheng, K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**, 343–346 (1992).
5. Young, M.P. & Yamane, S. Sparse population coding of faces in the inferotemporal cortex. *Science* **256**, 1327–1331 (1992).
6. Rolls, E.T., Treves, A. & Tovee, M.J. The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Exp. Brain Res.* **114**, 149–162 (1997).
7. Op de Beeck, H., Wagemans, J. & Vogels, R. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat. Neurosci.* **4**, 1244–1252 (2001).
8. 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).
9. Sakai, K. & Miyashita, Y. Neural organization for the long-term memory of paired associates. *Nature* **354**, 152–155 (1991).
10. Miller, E.K. & Desimone, R. Parallel neuronal mechanisms for short-term memory. *Science* **263**, 520–522 (1994).
11. Sigala, N. & Logothetis, N.K. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415**, 318–320 (2002).
12. Baker, C.I., Behrmann, M. & Olson, C.R. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat. Neurosci.* **5**, 1210–1216 (2002).
13. Pasupathy, A. & Connor, C.E. Population coding of shape in area V4. *Nat. Neurosci.* **5**, 1332–1338 (2002).
14. Wang, Y., Fujita, I. & Murayama, Y. Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nat. Neurosci.* **3**, 807–813 (2000).
15. Carandini, M., Heeger, D.J. & Movshon, J.A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17**, 8621–8644 (1997).
16. Mechler, F. & Ringach, D.L. On the classification of simple and complex cells. *Vision Res.* **42**, 1017–1033 (2002).
17. Ito, M., Tamura, H., Fujita, I. & Tanaka, K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* **73**, 218–226 (1995).
18. DiCarlo, J.J. & Maunsell, J.H. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J. Neurophysiol.* **89**, 3264–3278 (2003).
19. Op de Beeck, H. & Vogels, R. Spatial sensitivity of macaque inferior temporal neurons. *J. Comp Neurol.* **426**, 505–518 (2000).
20. Jones, J.P. & Palmer, L.A. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58**, 1187–1211 (1987).
21. Ringach, D.L., Sapiro, G. & Shapley, R. A subspace reverse-correlation technique for the study of visual neurons. *Vision Res.* **37**, 2455–2464 (1997).
22. Theunissen, F.E. *et al.* Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* **12**, 289–316 (2001).
23. Marr, D. & Nishihara, H.K. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond B Biol. Sci.* **200**, 269–294 (1978).
24. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
25. Salinas, E. & Abbott, L.F. Invariant visual responses from attentional gain fields. *J. Neurophysiol.* **77**, 3267–3272 (1997).
26. Wilson, H.R., Wilkinson, F. & Asaad, W. Concentric orientation summation in human form vision. *Vision Res.* **37**, 2325–2330 (1997).
27. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999).
28. Mel, B.W. & Fiser, J. Minimizing binding errors using learned conjunctive features. *Neural Comput.* **12**, 731–762 (2000).
29. Edelman, S. & Intrator, N. Towards structural systematicity in distributed, statically bound visual representations. *Cognit. Sci.* **27**, 73–109 (2003).
30. Janssen, P., Vogels, R. & Orban, G.A. Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proc. Natl. Acad. Sci. USA* **96**, 8217–8222 (1999).
31. Kovacs, G. *et al.* Effects of surface cues on macaque inferior temporal cortical responses. *Cereb. Cortex* **13**, 178–188 (2003).
32. Barlow, H. Redundancy reduction revisited. *Network* **12**, 241–253 (2001).
33. Olshausen, B.A. & Field, D.J. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res.* **37**, 3311–3325 (1997).
34. Vinje, W.E. & Gallant, J.L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
35. Kapadia, M.K., Westheimer, G. & Gilbert, C.D. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *J. Neurophysiol.* **84**, 2048–2062 (2000).
36. Polat, U., Mizobe, K., Pettet, M.W., Kasamatsu, T. & Norcia, A.M. Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* **391**, 580–584 (1998).
37. Gallant, J.L., Braun, J. & Van Essen, D.C. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* **259**, 100–103 (1993).
38. Wolfe, J.M., Yee, A. & Friedman-Hill, S.R. Curvature is a basic feature for visual search tasks. *Perception* **21**, 465–480 (1992).
39. Olson, C.R. & Gettner, S.N. Object-centered direction selectivity in the macaque supplementary eye field. *Science* **269**, 985–988 (1995).
40. Connor, C.E., Preddie, D.C., Gallant, J.L. & Van Essen, D.C. Spatial attention effects in macaque area V4. *J. Neurosci.* **17**, 3201–3214 (1997).
41. Zhou, H., Friedman, H.S. & von der Heydt, R. Coding of border ownership in monkey visual cortex. *J. Neurosci.* **20**, 6594–6611 (2000).
42. Schwartz, E.L., Desimone, R., Albright, T.D. & Gross, C.G. Shape recognition and inferior temporal neurons. *Proc. Natl. Acad. Sci. USA* **80**, 5776–5778 (1983).
43. Albright, T.D. & Gross, C.G. Do inferior temporal neurons encode shape by acting as fourier descriptor filters? *Proc. Int. Conf. on Fuzzy Logic & Neural Networks* **2**, 375–378 (1990).
44. Optican, L.M. & Richmond, B.J. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol.* **57**, 162–178 (1987).
45. Gochin, P.M., Colombo, M., Dorfman, G.A., Gerstein, G.L. & Gross, C.G. Neural ensemble coding in inferior temporal cortex. *J. Neurophysiol.* **71**, 2325–2337 (1994).
46. Ullman, S. Three-dimensional object recognition based on the combination of views. *Cognition* **67**, 21–44 (1998).
47. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
48. Hoffman, D.D. & Richards, W.A. Parts of recognition. *Cognition* **18**, 65–96 (1984).
49. Leyton, M. *A Generative Theory of Shape* (Springer, Berlin, 1998).
50. Kimia, B.B. On the role of medial geometry in human vision. *J. Physiol. (Paris)* **97**, 155–190 (2003).