nature neuroscience

A neural code for three-dimensional object shape in macaque inferotemporal cortex

Yukako Yamane[1], Eric T Carlson[1,2], Katherine C Bowman[1,3], Zhihong Wang[1] & Charles E Connor[1,3]

Previous investigations of the neural code for complex object shape have focused on two-dimensional pattern representation. This may be the primary mode for object vision given its simplicity and direct relation to the retinal image. In contrast, three-dimensional shape representation requires higher-dimensional coding derived from extensive computation. We found evidence for an explicit neural code for complex three-dimensional object shape. We used an evolutionary stimulus strategy and linear/nonlinear response models to characterize three-dimensional shape responses in macaque monkey inferotemporal cortex (IT). We found widespread tuning for three-dimensional spatial configurations of surface fragments characterized by their three-dimensional orientations and joint principal curvatures. Configural representation of three-dimensional shape could provide specific knowledge of object structure to support guidance of complex physical interactions and evaluation of object functionality and utility.

A primary goal in the study of object vision is to decipher the neural code for complex object shape. At the retinal level, object shape is represented isomorphically (that is, replicated point for point) across a two-dimensional map comprising approximately 10^6 pixels. This isomorphic representation is far too unwieldy and unstable (as a result of continual changes in object position and orientation) to be useful for object perception. The ventral pathway of visual cortex[1,2] must transform the isomorphic image into a compact, stable neural code that efficiently captures the shape information needed for identification and other aspects of object vision.

Previous studies of complex shape coding have focused on two-dimensional pattern representation. These studies have shown that neurons at intermediate (areas V2 and V4) and final (IT) stages in the monkey ventral pathway process information about two-dimensional shape fragments. V2 and V4 neurons encode curvature, orientation and object-relative position of two-dimensional object boundary fragments[3–7]. At the population level, these signals combine to represent complete boundary shapes as spatial configurations of constituent fragments[8]. In posterior IT, individual neurons integrate information about multiple two-dimensional boundary fragments, producing explicit signals for more complex shape configurations[9,10]. In central/anterior IT, the homolog of high-level object vision regions in human cortex[11–13], neurons are selective for a variety of patterns, and that selectivity is organized across the cortical surface in a columnar fashion[14–18]. At each stage, neurons appear to be tuned for component-level shape, although holistic shape tuning can evolve in IT through learning[19]. Holistic object representation may be more fully realized in medial temporal brain structures associated with long-term declarative memory[20] and in prefrontal areas processing categorical object information[21].

The question addressed here is whether and how complex three-dimensional shapes are encoded by IT neurons. Our specific hypothesis is that IT neurons encode three-dimensional spatial configurations of surface fragments. Under this hypothesis, the two-dimensional structural representations described above could be considered to occupy a subspace in the three-dimensional structure domain (that is, surface fragments forming the two-dimensional self-occlusion boundary of an object would be a special case of three-dimensional surface fragments). This hypothesis is consistent with classic shape-coding theories in which objects are represented as three-dimensional spatial configurations of simple three-dimensional parts[22,23]. The alternative hypothesis, advanced in more current theories, is that complex shape perception is based primarily on two-dimensional image processing. According to these theories, consistent recognition of three-dimensional objects from different vantage points is achieved by learning associations between multiple two-dimensional views[24]. The multiple-views hypothesis avoids the time-consuming computational complexity of inferring three-dimensional structure. This hypothesis is supported by psychophysical results showing that view-invariant recognition is learning dependent[25,26] and by computational studies showing that two-dimensional image processing can support rapid, accurate object identification[27,28]. However, these results and the multiple-views hypothesis itself are also compatible with three-dimensional representation[29] (see Discussion).

The classic hypothesis that complex shapes are represented as three-dimensional spatial configurations of three-dimensional parts has yet to be tested at the neural level. Previous studies have shown differential responses across a small number of three-dimensional shapes[30] or tuning along a single depth-related dimension[31–34], but such results cannot demonstrate or explain complex three-dimensional

[1]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, USA. [2]Department of Biomedical Engineering, Johns Hopkins University, School of Medicine, 720 Rutland Avenue, Baltimore, Maryland 21205, USA. [3]Department of Neuroscience, Johns Hopkins University, School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205, USA. Correspondence should be addressed to C.E.C. (connor@jhu.edu).

Received 13 February; accepted 3 September; published online 5 October 2008; corrected online 12 October 2008 (details online); doi:10.1038/nn.2202

shape representation (similar responses in dorsal pathway cortex have been interpreted as signals for orientation in depth[35]). Representation of three-dimensional object shape would require neurons with much more complex, multi-dimensional tuning properties. That kind of tuning can only be measured with large stimulus sets in which a wide range of three-dimensional shape elements are combined in many different ways, in which quantitative analyses can be used to disambiguate which three-dimensional

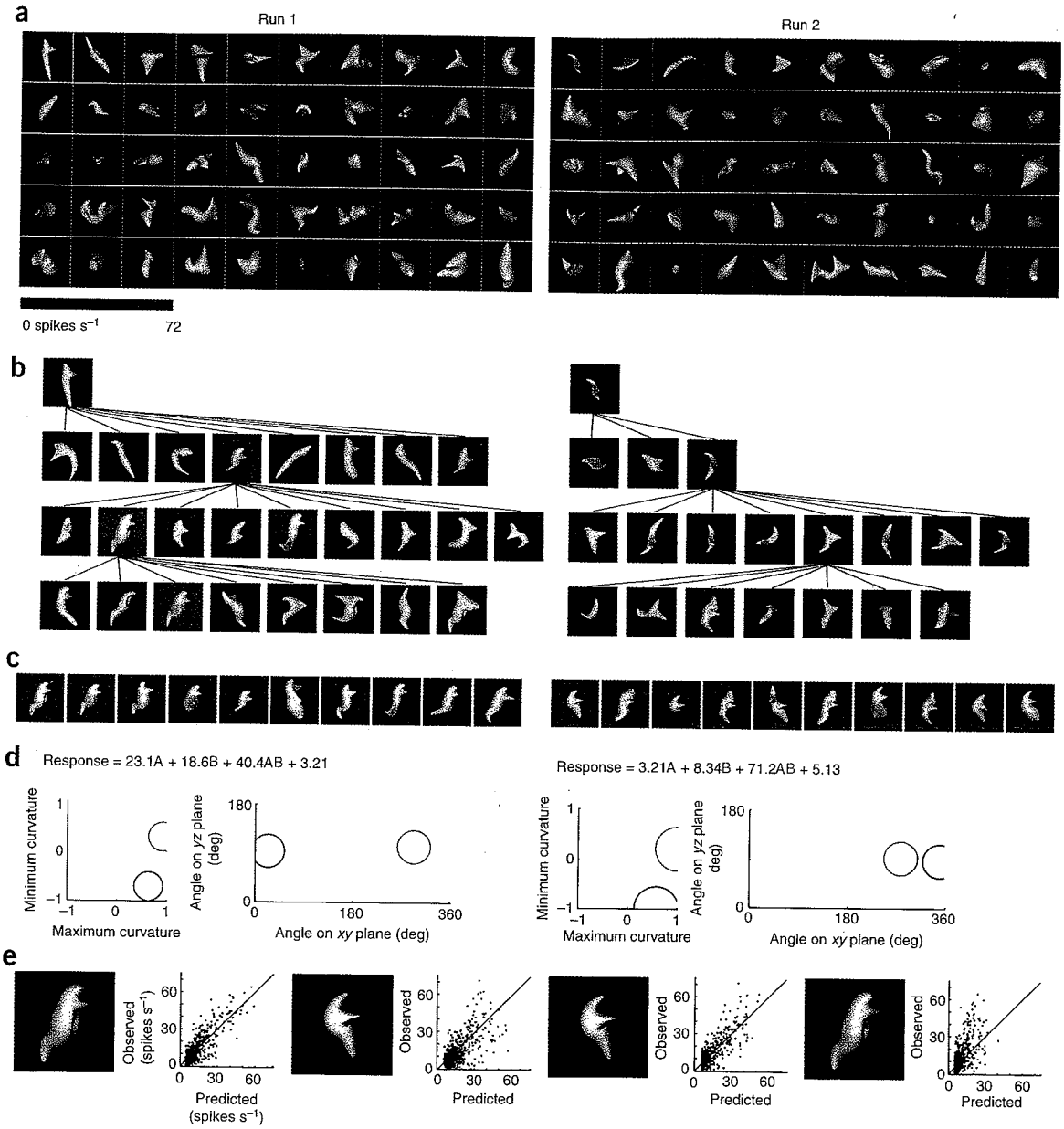

Figure 1 Evolutionary three-dimensional shape experiment. Two independent stimulus lineages (Run 1 and Run 2) are shown in the left and right columns, respectively. Background color (see scale bar) indicates the average response to each stimulus of a single IT neuron recorded from the ventral bank of the superior temporal sulcus (6.45 mm anterior to the interaural line). (**a**) Initial generations of 50 randomly constructed three-dimensional shape stimuli. Stimuli are ordered from top left to bottom right according to average response strength. (**b**) Partial family trees showing how stimulus shape and response strength evolved across successive generations. (**c**) Highest response stimuli across ten generations (500 stimuli) in each lineage. (**d**) Linear/nonlinear response models based on two Gaussian tuning functions. The Gaussian functions describe tuning for surface fragment geometry, defined in terms of curvature (principal, that is, maximum and minimum, cross-sectional curvatures), orientation (of a surface normal vector, projected onto the *xy* and *yz* planes) and position (relative to object center of mass in *xyz* coordinates). The curvature scale is squashed to a range between −1 and 1 (see Methods). The 1.0 s.d. boundaries of the two Gaussians (magenta and cyan) are shown projected onto different combinations of these dimensions. These boundaries appear circular because standard deviations in the curvature, orientation and position dimensions were constrained (respectively) to have the same values. The equations show the overall response models, with fitted weights for the two Gaussians, the product or interaction term, and the baseline response. (**e**) The two Gaussian functions are shown projected onto the surface of a high-response stimulus from each run. The stimulus surface is tinted according to the tuning amplitude in the corresponding region of the model domain. In this and subsequent displays, the projection areas are extended to include strongly correlated surface regions (see Methods). The scatter plots show the relationship between observed and predicted responses (left, self-prediction; right, cross-prediction).

shape factors (if any) are uniquely and consistently associated with neural responses. This has not been attempted before because of the intractable size of three-dimensional shape space. In this virtually infinite domain, a conventional random or systematic (grid-based) stimulus approach can never produce sufficiently dense combinatorial sampling.

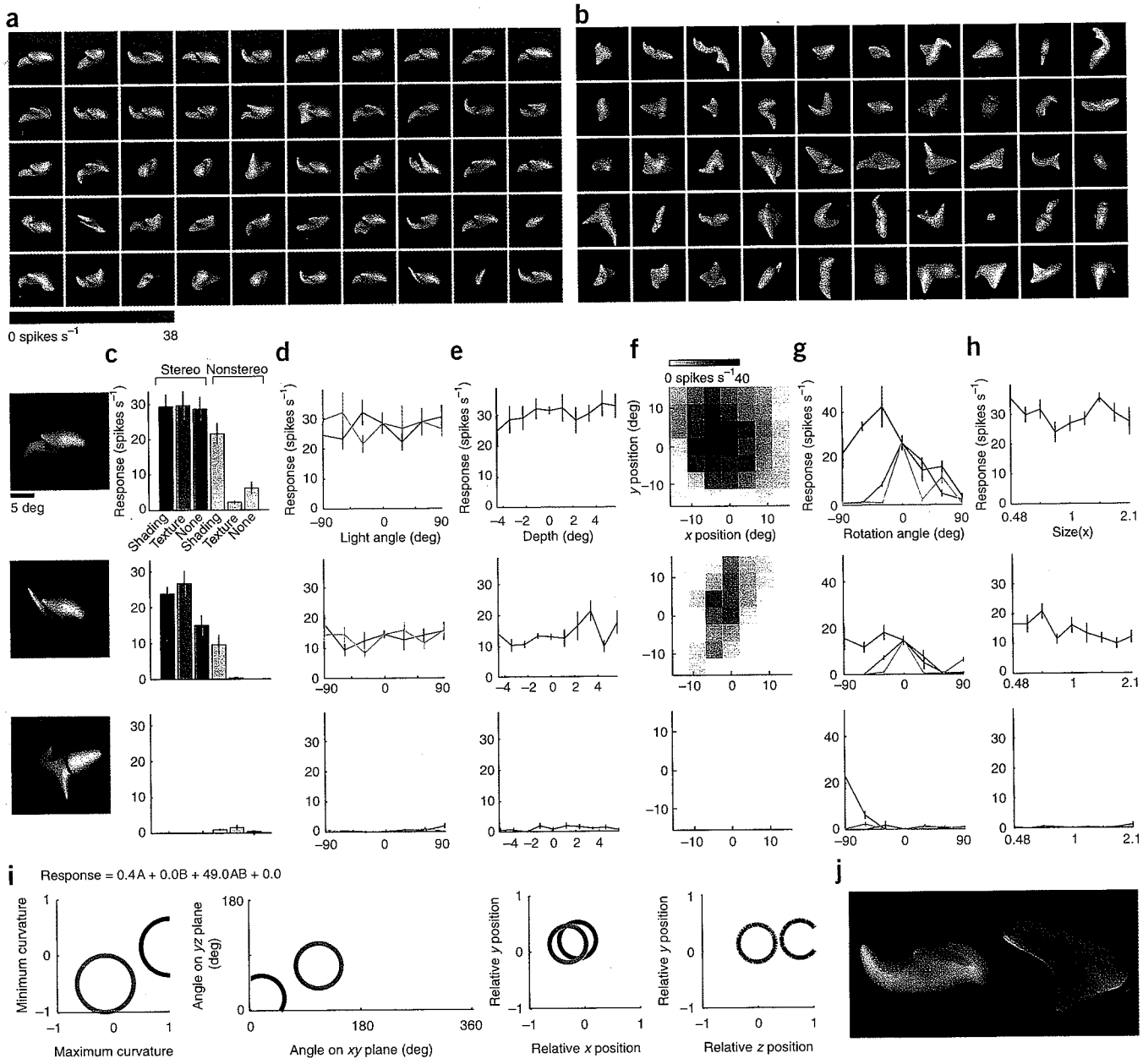

Figure 2 Neural tuning for three-dimensional configuration of surface fragments. (**a**) Top 50 stimuli across eight generations (400 stimuli) for a single IT neuron recorded from the ventral bank of the superior temporal sulcus (17.5 mm anterior to the interaural line). (**b**) Bottom 50 stimuli for the same cell. (**c**) Responses to highly effective (top), moderately effective (middle) and ineffective (bottom) example stimuli as a function of depth cues (shading, disparity and texture gradients, exemplified in **Supplementary Fig. 10**). Responses remained strong as long as disparity (black, green and blue) or shading (gray) cues were present. The cell did not respond to stimuli with only texture cues (pale green) or silhouettes with no depth cues (pale blue). (**d**) Response consistency across lighting direction. The implicit direction of a point source at infinity was varied across 180° in the horizontal (left to right, black curve) and vertical (below to above, green curve) directions, creating very different two-dimensional shading patterns (**Supplementary Fig. 11**). (**e**) Response consistency across stereoscopic depth. In the main experiment, the depth of each stimulus was adjusted so that the disparity of the surface point at fixation was 0 (that is, the animal was fixating in depth on the object surface). In this test, the disparity of this surface point was varied from −4.5° (near) to 5.6° (far). (**f**) Response consistency across *xy* position. Position was varied in increments of 4.5° of visual angle across a range of 13.5° in both directions. (**g**) Sensitivity to stimulus orientation. As with all neurons in our sample, this cell was highly sensitive to stimulus orientation, although it showed broad tolerance (about 90°) to rotation about the *z* axis (rotation in the image plane, blue curve); this rotation tolerance is also apparent among the top 50 stimuli in **a**. Rotation out of the image plane, about the *x* axis (black) or *y* axis (green), strongly suppressed responses. (**h**) Response consistency across object size over a range from half to twice that of the original stimulus. (**i**) Linear/nonlinear response model based on two Gaussian tuning functions (details as in **Fig. 2d,e**). (**j**) The tuning functions are projected onto the surface of a high-response stimulus, seen from the observer's viewpoint (left) and from above (right). Error bars indicate s.e.m. in all panels.

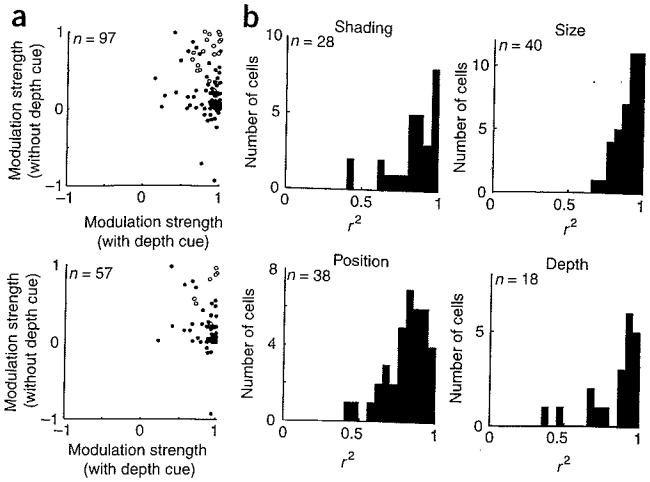

Figure 3 Prevalence of three-dimensional shape tuning in IT. (**a**) Response modulation depended strongly on three-dimensional cues. For each cell, three stimuli (with high, medium and low responses) were presented with depth cues (disparity and shading) and without depth cues (solid color silhouettes). A modulation index was calculated for each condition. The modulation index is the response difference between the high- and low-response stimuli, normalized by maximum response across all conditions. This normalization ensures that high values reflect robust responses. In some cases, removing three-dimensional cues reversed the rank order of responses and produced negative index values. The average modulation index of 0.85 with depth cues (horizontal axis) dropped to 0.26 without depth cues (vertical axis). The effect of depth cues on responses was significant ($P < 0.05$) for 76 of 97 cells (filled circles) on the basis of two-way ANOVA (main or interaction effects of stereo and shading). Of the 95 cells with significant three-dimensional shape tuning models, 57 were tested in this way. For these cells, the modulation index average dropped from 0.87 with depth cues to 0.23 without and 49/57 cells showed significant effects. (**b**) Shape tuning was independent of shading pattern, stereoscopic depth, stimulus position and stimulus size. Response consistency across these factors was tested as shown in **Figure 2**. Response consistency was measured by separability of tuning for shape (across the high-, medium- and low-response stimuli) and tuning for shading, depth, position or size. Separability is represented here by the fraction of response variance (r^2) explainable by a matrix product between separate tuning functions for shape and shading, depth, position or size. These tuning functions were the first pair of singular vectors in a singular value decomposition of the observed tuning matrix. For each factor, most neurons had r^2 values above 0.75, showing that three-dimensional shape tuning was largely independent of lighting direction, stimulus position, stimulus size and stimulus depth.

We addressed this obstacle with a new evolutionary stimulus strategy. Beginning with an initial generation of 50 random three-dimensional shapes, stimuli evolved through multiple generations under the guidance of neural feedback. Ancestor stimuli from previous generations were probabilistically morphed, either locally or globally, to produce descendant stimuli that varied the ancestor's shape characteristics and/or combined them with new shape features. The average neural response to each stimulus determined the probability with which it produced morphed descendants in subsequent stimulus generations. This strategy has two advantages. First, sampling becomes increasingly focused around the response range of the neuron, and thus far less experimental time is spent sampling null-response regions. Second, in the high-response stimulus lineages, the (initially unknown) shape characteristics encoded by the neuron are repeatedly varied and recombined with other shape features. The result is much denser, more combinatorial sampling in the most relevant region of the three-dimensional shape domain (in contrast, standard gradient descent search aims to identify a single, maximum response stimulus, which by itself cannot reveal what specific shape information is encoded by a neuron). This evolutionary stimulus strategy made it possible for the first time to test the three-dimensional configural coding hypothesis at the neural level.

RESULTS

Random three-dimensional shape stimuli were constructed by extensively deforming a closed ellipsoidal surface (**Supplementary Fig. 1** online). These stimuli were rendered in depth by a combination of binocular disparity and shading cues. Two rhesus monkeys were trained to maintain fixation on a small spot while stimuli were flashed (for 750 ms) sequentially at the center of gaze. We recorded the stimulus responses of individual neurons in central and anterior IT (between 5.8 and 21.0 mm anterior to the interaural line). Neurons were typically tested with 8–10 generations of 50 stimuli each. When recording time permitted, the entire procedure was run a second time to verify that stimulus evolution would converge in the same direction (**Fig. 1**). In each run, the initial generation was completely random and responses were generally low (**Fig. 1a**). Higher-response stimuli had a greater probability of producing morphed descendants in subsequent generations, and response levels progressed across generations in a way that was sometimes gradual and sometimes punctuated (**Fig. 1b**; also see **Supplementary Figs. 2–5** online). In any given generation, only a few descendants produced higher responses; most descendants evoked equivalent or lower responses. These lower response samples in neighboring shape space are essential for characterizing the shoulders and boundaries of tuning functions.

For this example cell, both lineages converged toward shapes that varied on the global level but contained consistent local structure comprising sharp protrusions and indentations, oriented toward the right and positioned to the upper right of object center (**Fig. 1c**). We quantified this response pattern with a multi-component model analogous to those applied recently to dorsal pathway area MT and in posterior IT cortex[9,10,36]. In the MT analyses, the model components represent tuning for different movement directions in the orientation domain. Here the model components represent tuning for different surface fragments in a three-dimensional curvature/orientation/position domain. Each of the two tuning models that we created on the basis of the two runs for this neuron (**Fig. 1d**) comprised two component Gaussian functions. One set of functions (magenta) was centered on strongly positive (convex) maximum curvature and negative (concave) minimum curvature, indicating a hyperbolic or saddle-shaped surface. This surface was oriented to the right (0°) and positioned to the upper right in the *xy* plane. The other set of functions (cyan) were centered on convex/convex curvature, oriented downwards (270°) and positioned to the upper right.

The predicted response for a given stimulus depended on how closely any of its constituent surface fragments matched these tuning functions. The overall predicted response was a linear combination of predictions based on the two separate functions plus a nonlinear interaction term (**Fig. 1d**). In this case, the nonlinear interaction terms had the highest fitted weights, showing that the neuron was relatively selective for the combination of both types of surface fragments. Higher-response stimuli in both runs included surface fragments near these two Gaussian functions (**Fig. 1e**). In this and all subsequent stimulus plots, the tinted regions include any additional surface structure that was strongly correlated with the Gaussian tuning function, to better capture the entire surface configuration associated

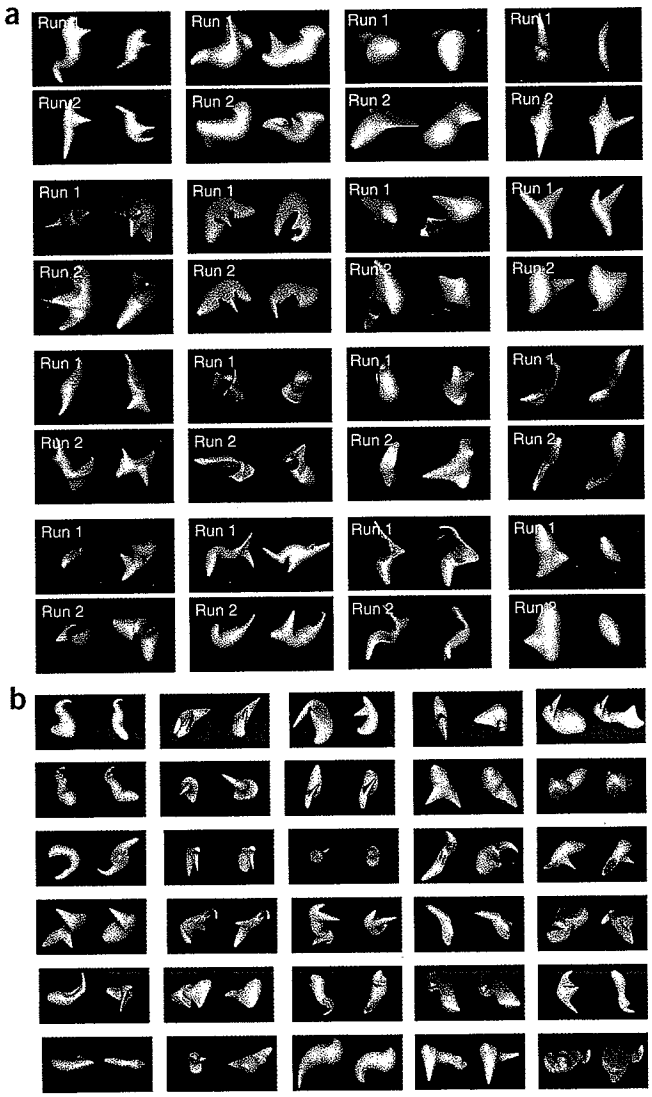

Figure 4 Three-dimensional surface configuration tuning patterns. (**a**) All neurons for which two independent evolutionary stimulus lineages were obtained. In each case, two high-response stimuli are shown from the first run (top) and the second run (bottom). Best-fit two-component models were projected onto these stimuli as in **Figure 1e**. (**b**) Example neurons for which only one lineage was obtained. In each case, two high-response stimuli are shown with the best-fit model projected onto the surface.

with neural responses. This is necessary because the geometry of closed, continuous surfaces imposes strong local structure correlations. For example, sharp points (tangent discontinuities) are necessarily correlated with the conjoined surfaces that define them (as in **Fig. 1e**). From a mechanistic point of view, the neuron's response could be driven by a sub-portion of that structure, but geometric constraints make that difficult or impossible to test in an experiment of this kind.

These tuning models showed strong cross-prediction of responses in the other run ($r = 0.67$ for run 1 model cross-prediction of run 2; $r = 0.63$ for run 2 model cross-prediction of run 1). The two runs provide a rigorous cross-validation test, as they were generated completely independently (geometric similarities between runs are imposed by the neuron itself and thus serve to confirm the generality of the tuning model). Cross-validation analyses of all cells with double lineages (see **Supplementary Fig. 6** online) established that models based on two Gaussian tuning regions had greater statistical validity than simpler models based on one Gaussian and more complex models based on three, four or five Gaussians. Restricting models to subunits explaining at least 5% additional variance produced a corresponding predominance of two-Gaussian models (see **Supplementary Fig. 7** online). The results presented below are based on two-Gaussian models for 95 neurons that showed statistically significant ($P < 0.05$) cross-validation between two runs ($n = 16$) or fivefold cross-validation in one run ($n = 79$). We corrected the fivefold cross-validation procedure for s.e.m. response measurements to estimate that these models accounted for 32% of the explainable response variance on average (mean $r = 0.57$; see **Supplementary Fig. 6**).

Figure 1 exemplifies the expected pattern for cells that encode two-dimensional boundary shape. Even though the response model domain encompasses three-dimensional surface fragments, the fitting procedure found surfaces at the two-dimensional occlusion boundary, with surface normals lying in the image plane, essentially corresponding to two-dimensional boundary fragments. Correspondingly, control tests showed that this cell responded nearly as strongly to the same shapes when all depth cues were removed (producing a silhouette shape with no internal detail, 54 spikes s^{-1} with stereo and shading cues and 44 spikes s^{-1} without; **Supplementary Fig. 8** online). However, this result did not typify the majority of neurons in our sample.

The more common finding was tuning for surface fragments outside the image plane (as in **Fig. 2**). The highest response shapes for this neuron were characterized by a ridge facing out of the image plane (**Fig. 2a**; also see **Supplementary Fig. 9** online). Stimuli lacking this surface characteristic evoked little or no response (**Fig. 2b**). Control tests performed on stimuli drawn from the top, median and bottom of the response range (**Fig. 2c–h**) confirmed that this neuron's shape tuning was dependent on depth structure and robust to other image changes. The neuron remained responsive when either disparity or shading cues for depth were present, but failed to respond when both were eliminated (**Fig. 2c**; see also **Supplementary Fig. 10** online). Three-dimensional shape selectivity remained consistent across lighting directions over 180° vertical and horizontal ranges (**Fig. 2d**), in spite of the resulting marked changes in shading patterns across the object surface (**Supplementary Fig. 11** online). Selectivity was likewise consistent across changes in stereoscopic depth (**Fig. 2e**) and *xy* position (**Fig. 2f**), although not across changes in stimulus orientation (**Fig. 2g**). Selectivity was consistent across stimulus size (**Fig. 2h**). The response model (**Fig. 2i**) comprised tuning for forward-facing ridges in front and upward-facing concavities near object center. The model for this neuron was highly nonlinear, as shown by the large AB interaction term representing combined energy in the two tuning regions. Thus, this neuron provided a relatively explicit signal for the ridge/concavity configuration. By explicit, we mean having a simple, easily decoded relationship to the shape configuration in question. Linear integration of information about multiple fragments would produce more ambiguous, less explicit signals, in which the same response level could correspond to either part A or part B. The AB configuration in this model characterized the high-response stimuli in the evolutionary test (**Fig. 2j**; one high-response stimulus is shown from both the front and above for greater visibility of the two surface components; see also **Supplementary Fig. 12** online).

Tuning for three-dimensional surface configurations (as in **Fig. 2**) was more common in our sample than tuning for two-dimensional boundary configurations (as in **Fig. 1**). We established this by follow-up control tests measuring how responses depended on the presence of three-dimensional cues (as in **Fig. 2c**). The critical comparison from these tests is between the condition in which the

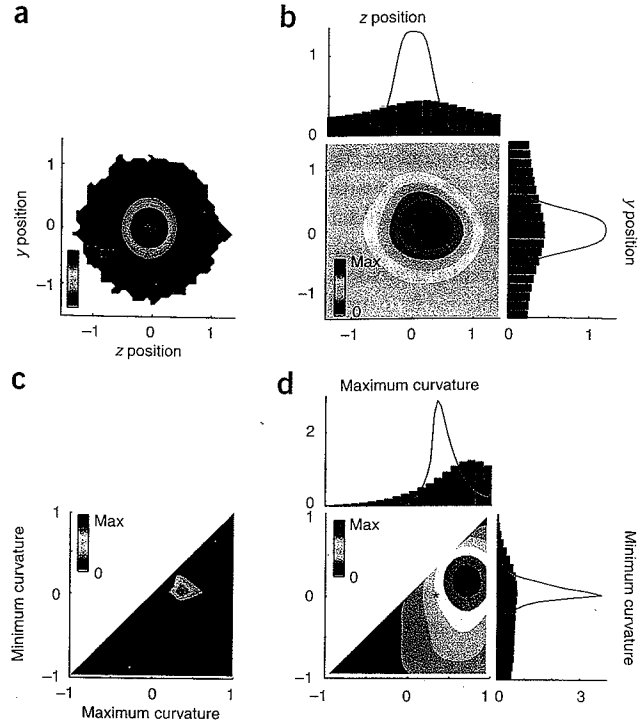

Figure 5 Distribution of three-dimensional shape tuning. (**a**) Comparison distribution of surface point positions in the *yz* plane (relative to object center of mass) in random stimuli (first generation stimuli for all 95 neurons described here). The scale is in arbitrary units approximately corresponding to stimulus size (maximum span in any direction averaged across stimuli = 1.08). (**b**) Distribution of Gaussian tuning peaks in best-fit models for 95 neurons. The stimulus distribution peak is shown in the surface plot (asterisk) and the stimulus distributions are shown in the marginal histograms (red curves). The distribution was biased toward positive values in the *z* dimension: that is, positions in front of object center. (**c**) Comparison distribution of surface curvatures across random (first generation) stimuli. The bias toward positive (convex) curvatures is characteristic of closed, topologically spherical surfaces. (**d**) Distribution of Gaussian tuning peaks in the curvature domain. The stimulus distribution peak is shown in the surface plot (asterisk) and the stimulus distributions are shown in the marginal histograms (red curves). Relative to the stimulus distribution, the tuning peaks were biased toward higher magnitude convexity in the maximum curvature dimension and higher magnitude concavity in the minimum curvature dimension.

standard three-dimensional cues (disparity and shading) were present and the condition in which no three-dimensional cues were present, so that the stimulus was a plain two-dimensional silhouette. In each condition, the strength of modulation was measured by the response difference between a stimulus that was drawn from the top of the main test response range and a stimulus that was drawn from the bottom of the range, normalized by the maximum response across all conditions. Response modulation was generally strong (near 1.0) when three-dimensional depth cues were present (**Fig. 3a**). When depth cues were removed, modulation sometimes remained strong, indicating that responses were determined by two-dimensional boundary shape. More commonly, however, modulation dropped to near 0, reflecting sensitivity to three-dimensional shape. The drop in response modulation with removal of three-dimensional cues does not simply reflect selectivity for two-dimensional shading patterns, as three-dimensional shape tuning was robust to changes in shading pattern produced by different lighting directions. We quantified this by measuring the separability of tuning for shape and tuning for lighting direction (**Fig. 3b**). Three-dimensional shape responses were likewise robust to changes in *xy* position, stereoscopic depth and stimulus size (**Fig. 3b**), analogous to previous results showing similar tolerance in two-dimensional shape responses[37].

Tuning models spanned a wide range of surface-fragment configurations (**Fig. 4a,b**). Component fragments were frequently discontiguous, consistent with a multi-fragment configural coding scheme. Even when tuning regions were extended to cover structures strongly correlated with the fitted Gaussians (as in **Fig. 4**), only a fraction (on average 23%) of the total object surface area was included (**Supplementary Fig. 13** online). Correspondingly, the global shape of high response stimuli varied at locations outside of these surface regions. Thus, three-dimensional shape representation in IT is not generally holistic. IT neurons represent spatially discrete three-dimensional shape components and must cooperate in a distributed coding scheme.

Tuning models showed a predictable bias toward surface positions near the front of the object, which is more visible and behaviorally relevant under normal circumstances (**Fig. 5a,b**). Tuning was markedly biased in the curvature domain toward high values, especially on the convex end of the scale. Thus, although object surface area was dominated by flat or broad curvature (**Fig. 5c**), the IT representation of three-dimensional shape emphasized sharper projecting points and ridges (**Fig. 5d**). Tuning for nonzero curvature reflects the coding advantage of higher-order derivatives[38]. The bias toward convexity may reflect the functional importance of protruding object parts and/or the well-established perceptual bias toward interpreting convexities as object parts and concavities as junctions between parts[39].

DISCUSSION

We tested the classic hypothesis that complex shapes are represented as three-dimensional spatial configurations of three-dimensional parts. This hypothesis requires that neurons encode the three-dimensional shape, three-dimensional orientation and relative three-dimensional position of object parts. Our analyses found that a substantial fraction of IT neurons did exactly that: they were simultaneously tuned for three-dimensional shape (maximum and minimum principle surface curvatures), three-dimensional orientation and relative three-dimensional position of constituent surface fragments. Moreover, they were tuned for multiple regions in this domain; that is, they responded to shapes that include a specific configuration of particular surface features. This result supports classic theories of three-dimensional configural shape representation[22,23] and extends those theories by suggesting that neurons encode not only individual parts but also configural relationships between multiple parts.

In many cases, our analyses and follow-up control tests revealed exclusive tuning for two-dimensional boundary shape. The majority of neurons in our sample, however, were clearly tuned for three-dimensional spatial configurations of three-dimensional surface fragments. Our control tests eliminated explanations that are based on stimulus properties other than three-dimensional shape. For example, neural responses might have been associated with two-dimensional boundary shape features, as these can be strongly correlated with three-dimensional surface shape[40]. But our depth cue test showed that, for most cells, removing three-dimensional information and presenting only the two-dimensional boundary resulted in a marked loss of tuning and responsiveness (**Fig. 3a**). Alternatively, neural responses might have been associated with the two-dimensional shading patterns that we used to help convey three-dimensional shape. But our lighting direction test showed that marked changes in the two-dimensional shading pattern (**Supplementary Fig. 11**) had little effect on shape

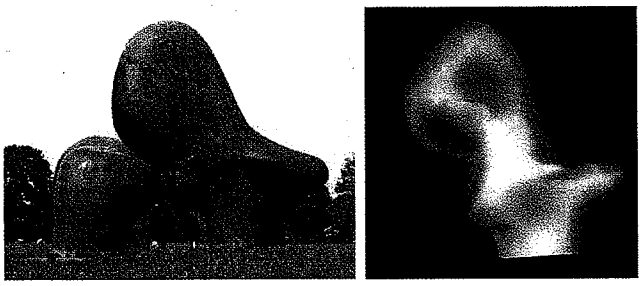

Figure 6 Configural coding of three-dimensional object structure. To illustrate how complex three-dimensional shape could be encoded at the population level, five two-Gaussian tuning models (red, green, blue, cyan and magenta) from our neural sample are projected onto a three-dimensional rendering (right) of the larger figure in Henry Moore's "Sheep Piece" (1971–1972, left; reproduced by permission of the Henry Moore Foundation, http://www.henry-moore-fdn.co.uk). Tuning models were scaled and rotated to optimize correspondence. A small number of neurons representing surface fragment configurations would uniquely specify an arbitrary three-dimensional shape of this kind and would carry the structural information required for judging its physical properties, functionality (or lack thereof) and aesthetic value.

tuning (**Fig. 3b**). Neural response differences might have reflected tuning for binocular disparity, but our depth position test showed that large changes in disparity had little effect on tuning (**Fig. 3b**). These three control tests eliminated alternate explanations based on any stimulus characteristic associated with the two-dimensional boundary, two-dimensional shading pattern or disparity values, and these three elements constitute all of the image information present in these stimuli. Likewise, large changes in stimulus size and position typically had little effect on tuning (**Fig. 3b**). No other potentially explanatory stimulus properties apart from three-dimensional surface shape itself would have survived all of these major image changes. As a further control, we carried out an analysis to show that response variations could not be explained in terms of image spatial-frequency content (**Supplementary Fig. 6**).

Our results are largely consistent with classic theories of configural shape representation[22,23]. According to these theories, objects are represented as spatial configurations of canonical three-dimensional parts. The parts are generalized cones or more complex volumetric components called 'geons'. Their configuration is represented in an object-defined reference frame that translates, scales and rotates with the object (producing invariance to viewpoint). Correspondingly, we observed explicit signals for configurations of three-dimensional surface fragments that appear to be encoded in an object-relative three-dimensional reference frame. By explicit, we mean easily decoded signals with clear, simple relationships to large-scale three-dimensional object structure. In contrast, although the same information is necessarily present in V1 (from which IT responses ultimately derive), the V1 representation is highly implicit and difficult to decode because it is distributed across a much larger population of neurons with complex, highly variable relationships to large-scale three-dimensional object structure.

Our results differed in two ways from classic theories. First, although the neural reference frame appeared to translate and scale with the object (given the consistency of responses across stimulus position and size), we did not find any evidence that it rotated with the object. Generalization across object rotations may depend on learned associations between views[24,41], although there is some capacity for recognizing rotated views of novel objects[42]. Representation of three-dimensional structure is not incompatible with this hypothesis[29]. Second, classic theories envision that individual neurons would represent single, spatially discrete parts. Instead, consistent with our previous two-dimensional studies, we found that neurons represent configurations of multiple parts, frequently at distant, discontiguous locations on the object surface (**Supplementary Fig. 14** online). Integration across multiple boundary and surface regions to derive larger, more complex configurations may be a ubiquitous aspect of visual processing[43]. Complete three-dimensional shapes might be represented in terms of a small number of component surface configurations (**Fig. 6**). A coding scheme such as this still has the combinatorial productivity of parts-based representation; a finite number of signals can be combined in many different ways to represent a virtually infinite number of objects. At the same time, it would constitute a step toward holistic shape coding, which has greater potential for sparseness. A sparse object representation based on just a few signals can be more efficiently stored in memory and decoded by other parts of the brain[44,45]. The configural coding that we observed may reflect a compromise between productivity and sparseness in higher-level visual cortex.

The neural code that we observed here is a three-dimensional generalization of the two-dimensional coding scheme we have previously described[8,9]. Boundary fragments in two dimensions generalize to surface fragments in three dimensions. The circular orientation domain for two-dimensional boundary fragments in the image plane generalizes to the spherical orientation domain for three-dimensional surface fragments. Tuning for two-dimensional boundary fragment curvature generalizes to joint tuning for the two principal curvatures characterizing any three-dimensional surface fragment. Tuning for the relative position of boundary fragments in the two-dimensional image plane generalizes to tuning for the relative position of surface fragments in three-dimensional space. Conceivably, the two-dimensional shape tuning properties that we and others have studied previously are not in any way distinct and simply occupy a subspace in the three-dimensional domain. In other words, neurons tuned for two-dimensional shape might simply be a subpopulation in a three-dimensional shape representation, encoding surface fragments near the occlusion boundary. The predominance of three-dimensional shape sensitivity in our neural sample supports this suggestion. However, this bias might be the result of denser sampling in the superior temporal sulcus, which has been reported to contain more three-dimensional sensitive neurons[46] (see **Supplementary Fig. 15** online).

To summarize, we used a new evolutionary stimulus strategy to reveal for the first time, to the best of our knowledge, an explicit structural code for three-dimensional object shape in high-level visual cortex. This code is embodied by neurons tuned for configurations of multiple object surface fragments. This result rules out the alternative explanation that complex shape representation is primarily two-dimensional, as in most current computational models of object vision. It supports classic theories of three-dimensional structural representation[22,23] but qualifies those theories in two ways. First, individual neurons represent not single object parts but instead represent configurations of multiple object parts, providing more explicit signals for spatial relationships between shape elements. Second, the spatial reference frame is only partially defined by the object. The reference frame is centered with respect to the object but, contrary to classic theories, does not rotate with the object. Thus, neural representations are very different depending on viewpoint, and relationships between different views of the same object would have to be learned, as proposed in view-dependent theories of object representation[24–26]. Finally, these results apply only to the population of neurons that we were able to study effectively with our stimulus strategy, and they do not rule out other object-processing mechanisms

in other neural populations. In particular, rapid, coarse object categorization might depend on faster feedforward processing of two-dimensional image information[28].

Why would the brain explicitly represent complex three-dimensional object shape, considering the computational expense of inferring three-dimensional structure from the two-dimensional retinal image and the higher neural tuning dimensionality required? In contrast, two-dimensional shape representations are simpler and can be derived quickly and directly from the retinal image. Moreover, computational studies of biologically inspired hierarchical network models show that two-dimensional image processing alone can produce rapid, accurate object identification[27,28]. We speculate that representation of three-dimensional object structure instead supports other aspects of object vision beyond identification. Direct cognitive access to three-dimensional object structure makes even unfamiliar objects comprehensible in terms of their geometric similarity to familiar objects, inferred physical properties, potential functionality and utility, and aesthetic qualities. Knowledge of three-dimensional structure is also necessary for accurate prediction of physical events and control over complex physical interactions with objects. These cognitive and behavioral requirements may have driven the emergence of an explicit neural code for three-dimensional object structure in visual cortex. Similar tuning might emerge from hierarchical network models given three-dimensional input information and more diverse task demands.

METHODS

Behavioral task. Two head-restrained rhesus monkeys (*Macaca mulatta*), a 9.5-kg male and a 5.3-kg female, were required to maintain fixation within 0.75° (radius) of a 0.1° spot for 4 s to obtain a juice reward. Eye position was monitored with an infrared eye tracker (ISCAN). Separate left- and right-eye images were presented via mirrors. Binocular fusion was verified with a random dot stereogram search task. All animal procedures were approved by the Johns Hopkins Animal Care and Use Committee and conformed to US National Institutes of Health and US Department of Agriculture guidelines.

Electrophysiological recording. The electrical activity of well-isolated single neurons was recorded with epoxy-coated tungsten electrodes (A-M Systems) and amplified and filtered in a Tucker-Davis Technologies acquisition system. We studied 250 neurons with at least 300 stimuli. Of these, 95 (59 from the male and 36 from the female) produced models that passed our statistical thresholds (see Results). These neurons were sampled from the central/anterior lower bank of the superior temporal sulcus and lateral convexity of the inferior temporal gyrus (5.8–21.0 mm anterior to the interaural line; see **Supplementary Fig. 15**). IT cortex was identified on the basis of structural magnetic resonance images.

Visual stimuli. Three-dimensional shape stimuli were rendered with shading and binocular disparity cues using the NURBS facility in OpenGL (gluNurbsSurface). NURBS control point positions were varied by distorting a polar grid (see **Supplementary Fig. 1**). Stimulus depth was adjusted so that the depth of the object surface at the fovea matched the fixation depth (screen distance).

Neurophysiological testing protocol. A roughly optimal stimulus color (white, red, green, blue, cyan, yellow or magenta) was selected on the basis of responses to stimuli under experimenter control. During all subsequent tests, following initiation of fixation, four stimuli were flashed one at a time for 750 ms each, with inter-stimulus intervals of 250 ms. Control tests of sensitivity to depth cues, *xy* position, stereoscopic depth, size, orientation and lighting direction (see **Fig. 2**) were performed on three stimuli, drawn from the top, median and bottom of the response range in the main experiment.

Evolutionary morphing algorithm. The first stimulus generation contained only randomly generated three-dimensional stimuli. Stimulus responses (averaged across five repetitions) were ranked into ten bins with equal numbers of stimuli. In the second generation, 10–20% of stimuli were randomly generated. The rest were morphed descendants of ancestor stimuli from the first generation, selected randomly in equal numbers from the ten bins. Thus, a typical second generation would contain ten stimuli generated *de novo*, four descendants of stimuli in the highest response bin, four descendants from the second highest bin, etc. In subsequent generations, ancestor stimuli were pooled across all preceding generations and re-binned. Descendant stimuli had an equal probability of being either locally or globally morphed (see **Supplementary Fig. 1**). The amplitude of control point changes was inversely proportional to the response rate to produce denser sampling in higher response ranges.

Three-dimensional shape response models. We characterized each three-dimensional stimulus in terms of its component surface fragments. The stimulus surface was densely sampled across the NURBS control point grid. At each point, seven values were determined: the *x*, *y* and *z* positions relative to the stimulus center of mass, the surface normal orientation on the *xy* and *yz* planes, and the maximum and minimum (principal) cross-sectional curvatures, squashed to a range from −1 to 1 using a sigmoidal function. Component surface fragments were defined by mathematically fitting elliptical regions on this grid with approximately constant curvature in either the minimum or maximum curvature dimensions. Each ellipse was shifted and scaled to be as large (that is, cover as many grid points) as possible without violating the constraints on maximum deviation of curvature values. Successive ellipses were fitted to remaining regions on the object surface. Curvature maxima and minima were also used to define surface fragments. On average, a given stimulus comprised 240 component surface fragments. For each fragment, maximum and minimum curvature were averaged across the included surface points, and the position and orientation values were measured at the point of maximum or minimum curvature, yielding seven measurements. Thus, each fragment corresponded to a point in a seven-dimensional domain and each stimulus was represented by a constellation of such points. All of the models described here were based on seven-dimensional Gaussian tuning functions (model subunits) in this domain. The predicted response component resulting from a given subunit was determined by the Gaussian function amplitude at the stimulus point closest to the Gaussian peak:

$$\text{Response}_{\text{subunit}} = A \cdot \max(\text{Gaussian}(ka_c kb_c \theta x_c \theta y_c rx_c ry_c rz_c))$$

$$\text{Gaussian}(ka_c kb_c \theta x_c \theta y_c rx_c ry_c rz_c) = -\exp\left(\frac{(ka_c - \mu_{ka})^2}{2\sigma_k^2} + \frac{(kb_c - \mu_{kb})^2}{2\sigma_k^2} + \frac{(\theta x_c - \mu_{\theta x})^2}{2\sigma_\theta^2} + \frac{(\theta y_c - \mu_{\theta y})^2}{2\sigma_\theta^2} + \frac{(rx_c - \mu_{rx})^2}{2\sigma_r^2} + \frac{(ry_c - \mu_{ry})^2}{2\sigma_r^2} + \frac{(rz_c - \mu_{rz})^2}{2\sigma_r^2}\right)$$

where ka_c, kb_c, θx_c, θy_c, rx_c, ry_c and rz_c are the curvature, normal and relative position values for a surface fragment in the stimulus; μ and σ are the fitted Gaussian peaks and the s.d. on each of these seven dimensions; and A is the fitted Gaussian amplitude. The s.d. in the curvature, orientation and position dimensions were constrained to be the same (respectively) for all Gaussian functions to limit model complexity (that is, only 3 s.d. values were required for any given model, one each in the curvature, orientation and position dimensions). Including separate s.d. parameters for each dimension slightly increased the variance explained, by a mean of 3.7% for the one-Gaussian models and 4.5% for the two-Gaussian models (fivefold cross-validation, $n = 95$). Because the difference in explained variance between one- and two-Gaussian models increases, this would not affect our conclusion that neurons are tuned for configurations of multiple surface fragments rather than single fragments. The analyses presented here were based on the models with just 3 s.d. parameters.

For each cell, we fitted models comprising 1–5 excitatory Gaussian subunits. The predicted response to each stimulus was a weighted combination of the individual subunit responses (the linear component) and a product of subunit responses (the nonlinear interaction component). All possible combinations of subunits for nonlinear components were tested, but each subunit could participate in only one nonlinear interaction term in the final models. For the two-Gaussian models emphasized in this report, there was

only one possible interaction term.

$$\text{Response} = G\left(\sum_{s=1}^{\#\text{subunits}} (W_s R_s) + \sum_{k=1}^{\#\text{NLterms}} (W_k R_k^{NL})\right) + b_0$$

R_s is the unweighted response predicted by each subunit (derived from the subunit response equation above) and $R^{NL}{}_s$ is the unweighted response predicted by each interaction term. W_s is the fitted weight (amplitude) for each subunit, W_k is the fitted weight for each interaction term, G is the overall gain and b_0 is the baseline firing rate. The total numbers of fitted parameters for the models with 1–5 subunits were 13, 21–22, 29–30, 37–39 and 45–47, respectively. The variability in parameter number for higher-order models is the result of variability in the number of possible interaction terms. Overfitting was controlled by cross-validation analyses (described in Results and presented in detail in **Supplementary Fig. 6**).

Response rates were calculated by counting the number of spikes during the 750-ms stimulus presentation period and averaged across repetitions. The Matlab function lsqnonlin was used to adjust model parameters to minimize the sum of squared differences between the observed and predicted responses. The fitting procedure was repeated using 20–163 (average 93) starting points based on the constituent surface fragments for the three highest response stimuli. For each neuron, the best-fitting model across all starting points was used.

In some cases, these Gaussian tuning models may fail to capture the complete shape configurations associated with neural responses, instead focusing on smaller surface fragments within the complete configurations. This problem is inevitable because of the local surface structure correlations in any closed, continuous surface (Results). To better capture the complete shape configurations signaled by the neurons, we identified additional surface structure components that were highly correlated with those identified by the fitted Gaussian tuning regions. We then extended the tuning regions to include those correlated structures, up to the point at which explained variance dropped by 5%. The extended tuning regions are bound to have a non-Gaussian shape, which we approximated with a cluster of neighboring Gaussians. Predicted responses were based on average stimulus matches across the Gaussians in each cluster. These extended models were the basis for projections onto stimulus surfaces (**Figs. 1e,2j** and **4**), tuning distributions (**Fig. 5** and **Supplementary Figs. 16** and **17** online) and analysis of fractional surface coverage (**Supplementary Fig. 12**). All goodness of fit results were based on the original fitted two-Gaussian models (computational procedure for extending tuning regions is described in **Supplementary Fig. 13**).

Statistical analysis. Randomization analysis of cross-validation between double lineages was used to compare the statistical validity of models based on 1–5 Gaussian tuning regions (see **Supplementary Fig. 6**). A fivefold cross-validation procedure was used to estimate explained variance without overfitting (Results).

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

ACKNOWLEDGMENTS

We thank B. Nash, B. Quinlan, C. Moses and L. Guruvadoo for technical support, and J. Bastian, A. Bastian, T. Poggio and M. Riesenhuber for comments on the manuscript. This work was supported by a grant from the US National Institutes of Health to C.E.C.

Published online at http://www.nature.com/natureneuroscience/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Ungerleider, L.G. & Mishkin, M. *Analysis of Visual Behavior* (eds. Ingle, D.G., Goodale, M.A. & Mansfield, R.J.Q.) 549–586 (MIT Press, Cambridge, Massachusetts, 1982).
2. Felleman, D.J. & Van Essen, D.C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
3. Anzai, A., Peng, X. & Van Essen, D.C. Neurons in monkey visual area V2 encode combinations of orientations. *Nat. Neurosci.* **10**, 1313–1321 (2007).
4. Ito, M. & Komatsu, H. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J. Neurosci.* **24**, 3313–3324 (2004).
5. 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).
6. Pasupathy, A. & Connor, C.E. Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* **86**, 2505–2519 (2001).
7. Pasupathy, A. & Connor, C.E. Responses to contour features in macaque area V4. *J. Neurophysiol.* **82**, 2490–2502 (1999).
8. Pasupathy, A. & Connor, C.E. Population coding of shape in area V4. *Nat. Neurosci.* **5**, 1332–1338 (2002).
9. Brincat, S.L. & Connor, C.E. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* **7**, 880–886 (2004).
10. Brincat, S.L. & Connor, C.E. Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* **49**, 17–24 (2006).
11. Malach, R. *et al.* Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. USA* **92**, 8135–8139 (1995).
12. Kourtzi, Z. & Kanwisher, N. Representation of perceived object shape by the human lateral occipital complex. *Science* **293**, 1506–1509 (2001).
13. Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B. & Tootell, R.B.H. Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* **6**, 989–995 (2003).
14. 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).
15. 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).
16. 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).
17. Fujita, I., Tanaka, K., Ito, M. & Cheng, K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**, 343–346 (1992).
18. 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).
19. 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).
20. Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005).
21. Freedman, D.J., Riesenhuber, M., Poggio, T. & Miller, E.K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).
22. Marr, D. & Nishihara, H.K. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B* **200**, 269–294 (1978).
23. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
24. Vetter, T., Hurlbert, A. & Poggio, T. View-based models of 3D object recognition: invariance to imaging transformations. *Cereb. Cortex* **5**, 261–269 (1995).
25. Bulthoff, H.H., Edelman, S.Y. & Tarr, M.J. How are three-dimensional objects represented in the brain? *Cereb. Cortex* **5**, 247–260 (1995).
26. Tarr, M.J. & Pinker, S. Mental rotation and orientation-dependence in shape recognition. *Cognit. Psychol.* **21**, 233–282 (1989).
27. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999).
28. Serre, T., Oliva, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. USA* **104**, 6424–6429 (2007).
29. Tarr, M.J. & Barenholtz, E. Reconsidering the role of structure in vision. in *The Psychology of Learning and Motivation* (eds. Ross, B. & Markman, A.) 157–180 (Academic Press, London, 2007).
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. Uka, T., Tanaka, H., Yoshiyama, K., Kato, M. & Fujita, I. Disparity selectivity of neurons in monkey inferior temporal cortex. *J. Neurophysiol.* **84**, 120–132 (2000).
32. Watanabe, M., Tanaka, H., Uka, T. & Fujita, I. Disparity-selective neurons in area V4 of macaque monkeys. *J. Neurophysiol.* **87**, 1960–1973 (2002).
33. Hinkle, D.A. & Connor, C.E. Three-dimensional orientation tuning in macaque area V4. *Nat. Neurosci.* **5**, 665–670 (2002).
34. Janssen, P., Vogels, R. & Orban, G.A. Three-dimensional shape coding in inferior temporal cortex. *Neuron* **27**, 385–397 (2000).
35. Sakata, H. *et al.* Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. *Phil. Trans. R. Soc. Lond. B* **353**, 1363–1373 (1998).
36. Rust, N.C., Mante, V., Simoncelli, E.P. & Movshon, A. How MT cells analyze the motion of visual patterns. *Nat. Neurosci.* **9**, 1421–1431 (2006).
37. 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).
38. Connor, C.E., Brincat, S.L. & Pasupathy, A. Transformation of shape information in the ventral pathway. *Curr. Opin. Neurobiol.* **17**, 140–147 (2007).
39. Hoffman, D.D. & Richards, W.A. Parts of recognition. *Cognition* **18**, 65–96 (1984).
40. Koenderink, J.J. What does the occluding contour tell us about solid shape? *Perception* **13**, 321–330 (1984).
41. Edelman, S. & Poggio, T. Models of object recognition. *Curr. Opin. Neurobiol.* **1**, 270–273 (1991).
42. Wang, G., Obama, S., Yamashita, W., Sugihara, T. & Tanaka, K. Prior experience of rotation is not required for recognizing objects seen from different angles. *Nat. Neurosci.* **8**, 1768–1775 (2005).
43. Roelfsema, P.R. Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* **29**, 203–227 (2006).
44. Rolls, E.T. & Treves, A. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* **1**, 407–421 (1990).
45. Vinje, W.E. & Gallant, J.L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
46. Janssen, P., Vogels, R. & Orban, G.A. Selectivity for 3D shape that reveals distinct areas within macaque inferotemporal cortex. *Science* **288**, 2054–2056 (2000).