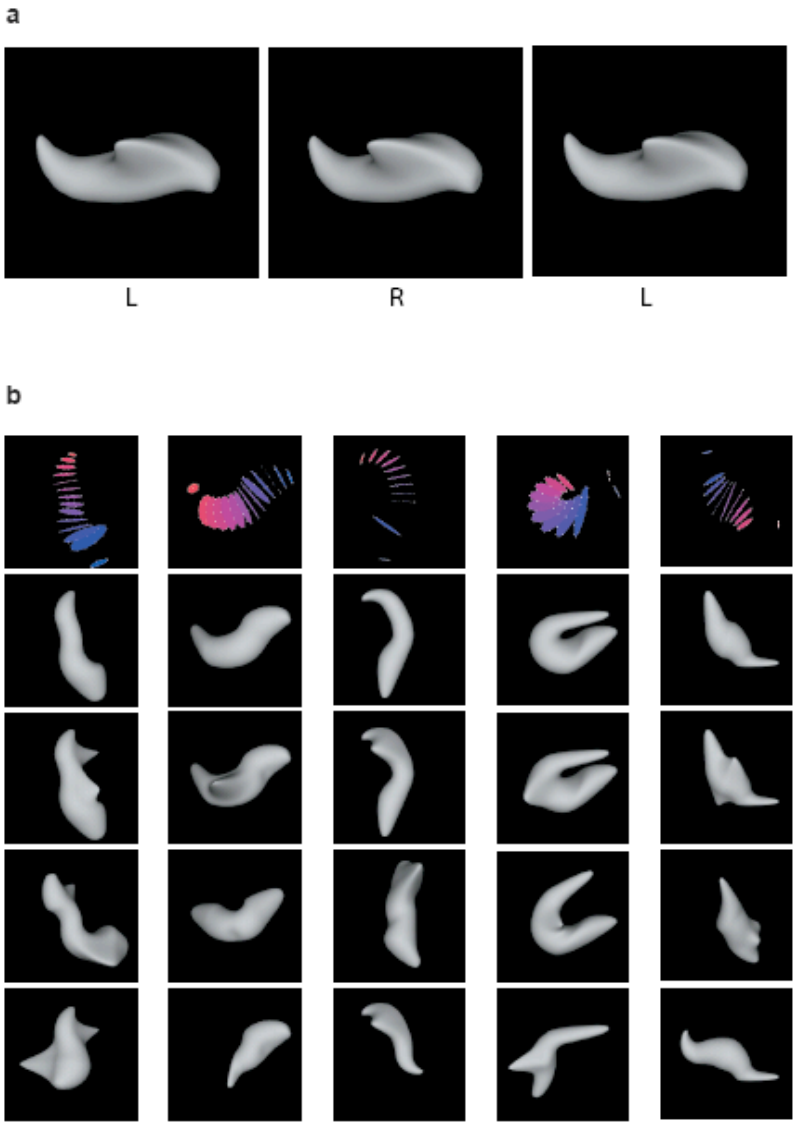
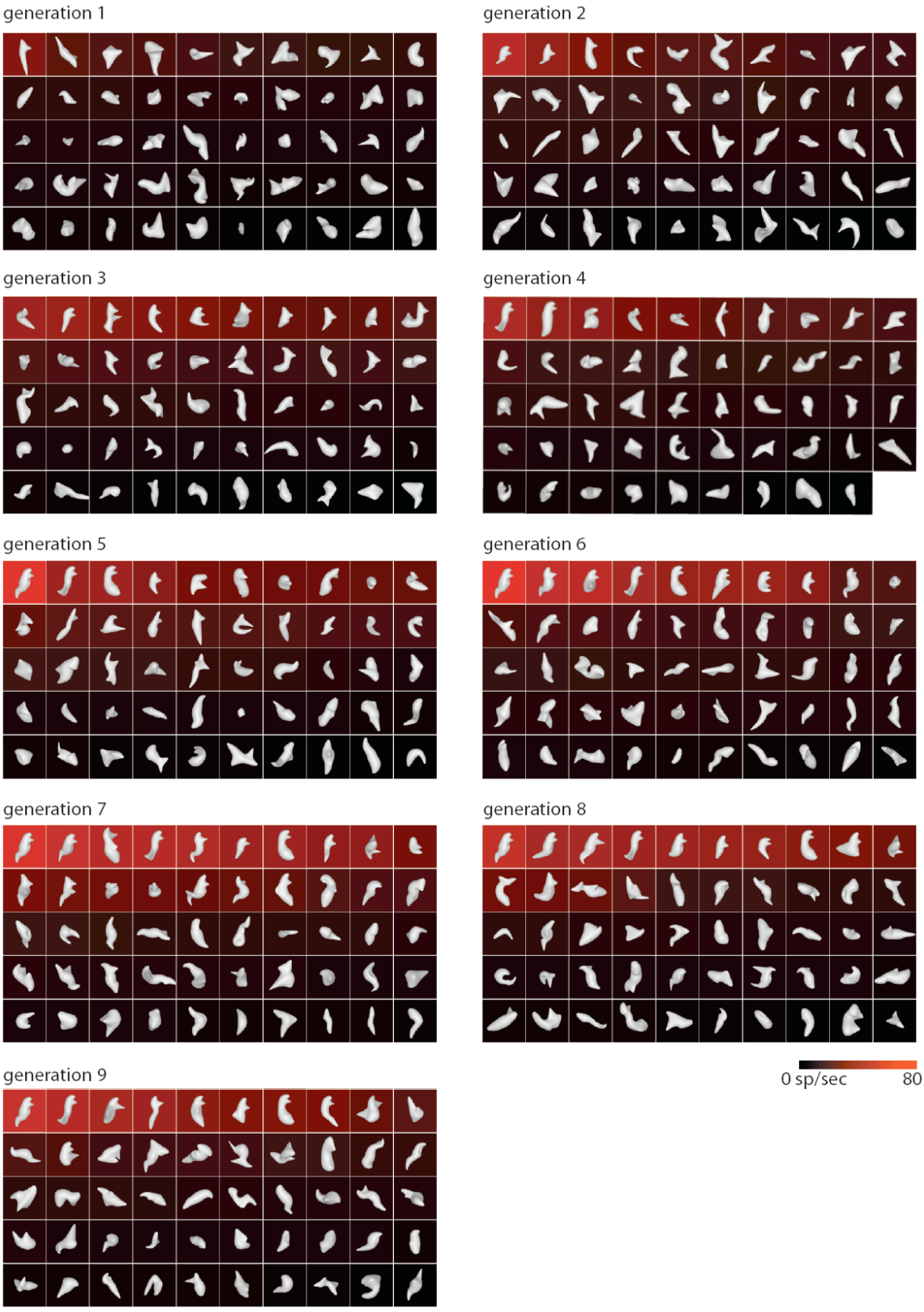
A neural code for three-dimensional object shape in macaque inferotemporal cortex

Yukako Yamane, Eric T. Carlson, Katherine C. Bowman, Zhihong Wang & Charles E. Connor



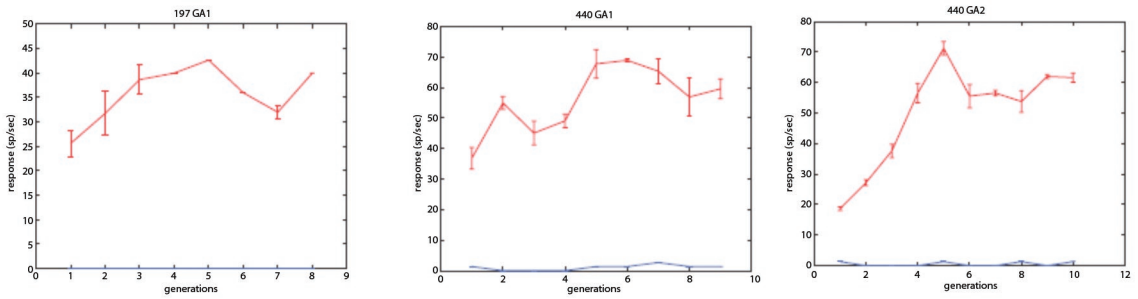
Supplementary Figure 1 | Random 3D stimulus construction and morphing. (**a**) Stereo pairs illustrating how 3D stimuli were rendered with binocular disparity and shading cues based on an implicit source from the viewer's direction at infinite distance. (**b**) Five examples (columns) of stimulus construction and morphing. The starting point

for each random stimulus was a polar grid of NURBS (non-uniform rational B-spline) control points defining an ellipsoid. The relative positions, sizes, aspect ratios and orientations of the latitudinal sections were randomly altered (first row; colored disks). The control point grid was used to define a continuous surface (second row). This surface was elaborated with randomly generated 2D elliptical Gaussian distortions applied to the NURBS grid, creating hills, ridges, bowls and channels (third row). In subsequent stimulus generations, higher response ancestor stimuli were represented by descendants morphed at the global (fourth row) or local (fifth row) levels. Morphed stimuli were created by altering the number, positions, orientations, and radii of the latitudinal sections, by altering the number, amplitudes, and standard deviations of the Gaussian distortions, and by changing 3D orientation. The global morphs (fourth row), in which all parameters were altered by a restricted amount, allowed us to test variations in many shape characteristics simultaneously. This ultimately led to dense sampling of structural variations that were relevant to neural responses. The local morphs (fifth row), in which one part of the shape was held constant while the rest was completely redefined, allowed us to decorrelate structural elements in the stimulus set, to the extent possible given the geometric constraints of closed, continuous surfaces. For example, in the rightmost column, the rightward protrusion that appears in the original stimulus (third row) is combined with an entirely different set of structures (fifth row). This ultimately produces a highly factorial stimulus set that provides a basis for disambiguating which specific structures were consistently associated with neural responses.

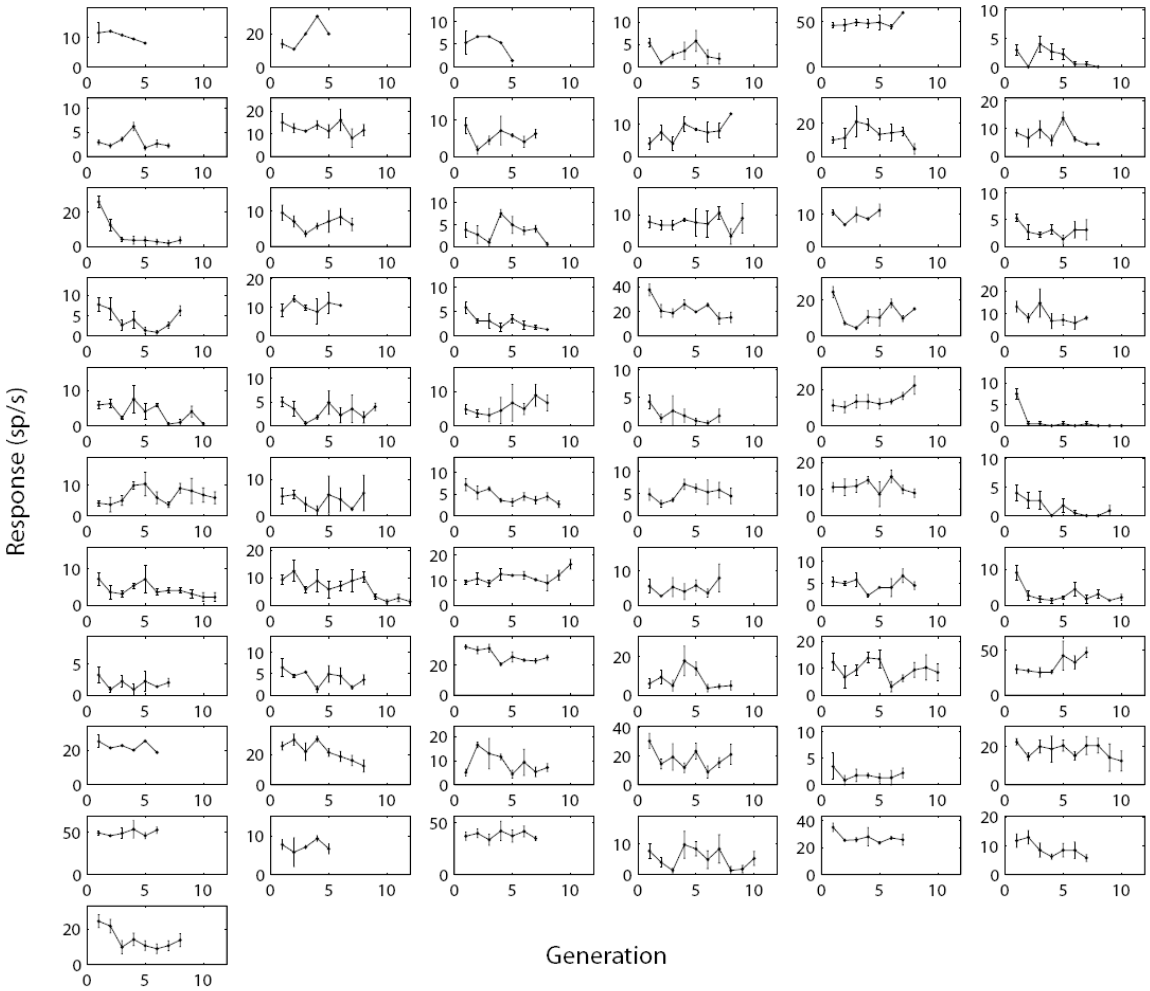


Supplementary Figure 2 | Successive stimulus generations for Fig. 1 neuron, Run 1. Average response to each stimulus is indicated by background color.

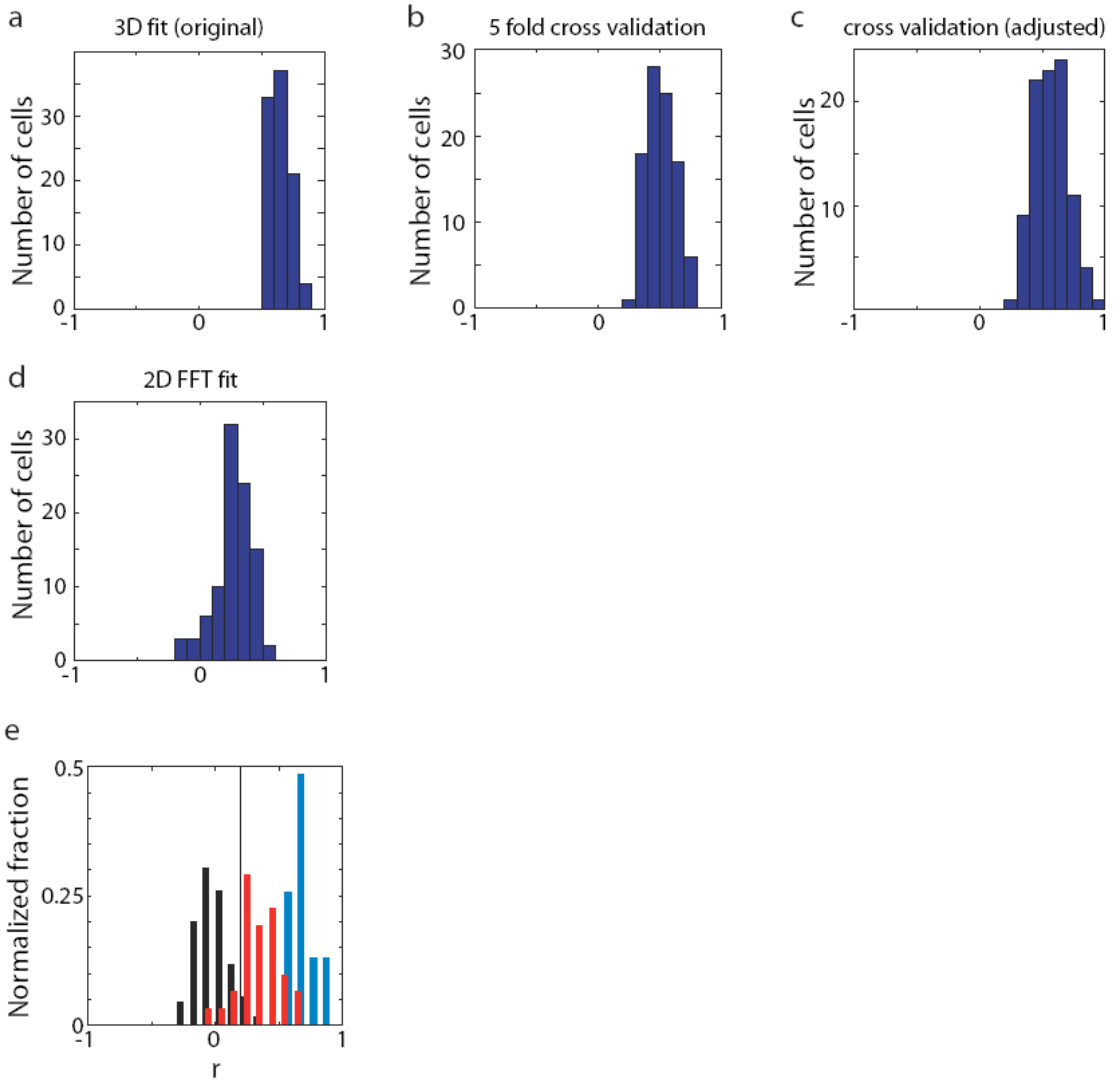
Supplementary Figure 3 | Successive stimulus generations for Fig. 1 neuron, Run 2. Average response to each stimulus is indicated by background color.



Supplementary Figure 4 | Maximum and minimum responses across generations. Maximum (red) and minimum (blue) response (mean +/- standard error) across generations for the Fig. 2 neuron (left) and the two runs in Fig. 1.

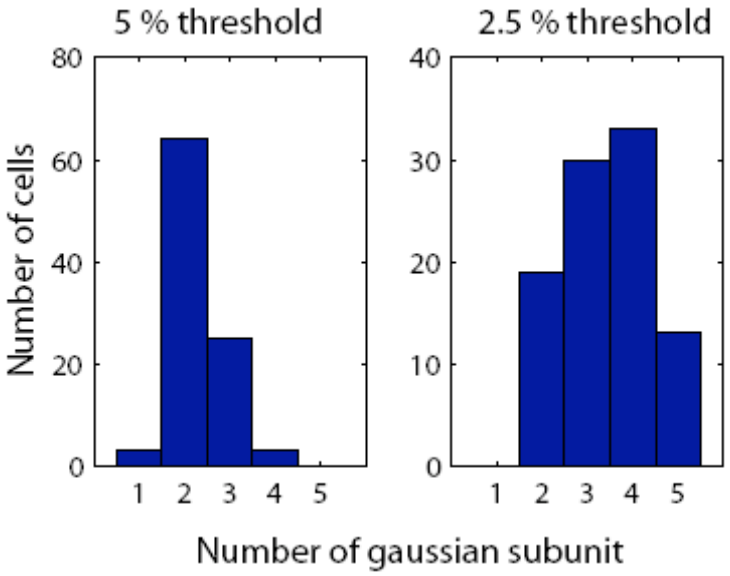


Supplementary Figure 5 | Response variability across generations. For 61 neurons, an identical stimulus was presented across at least 5 generations. Each plot shows the mean +/- standard error progression of responses to this stimulus across generations for one neuron. The general tendency for responses to decline across generations is probably due to the inclusion of more high response stimuli in late generations producing fatigue or habituation.

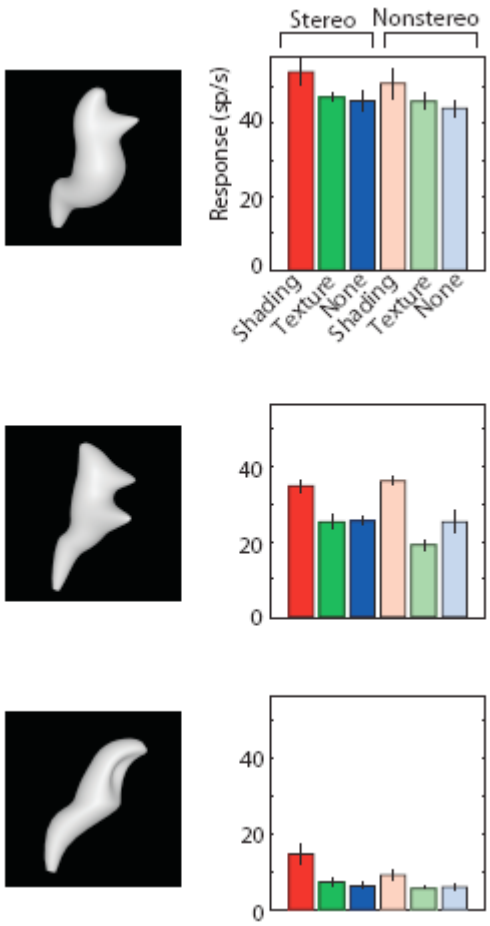


Supplementary Figure 6 | Correlation and cross-validation analyses. (**a**) Distribution of correlations between observed responses and responses predicted by the models presented for 95 neurons in the main text. The mean of this distribution is 0.64. (**b**) Distribution of correlations based on 5-fold cross-validation. For each neuron, stimuli were divided into 5 equal sets, and the responses to each set were predicted with a 2-Gaussian model fit only to the stimuli outside the set. Correlation coefficients were then averaged across the 5 fits for each neuron. The mean r value for this distribution is 0.51. (**c**) Distribution of 5-fold cross-validations corrected for measurement error using the method employed in Pasupathy and Connor, J. Neurophysiol 2001. For each neuron, we estimated response variance due to measurement error by summing squared standard error of the mean across stimuli. Squared standard error is the expected squared difference between the estimated mean response and the true mean response. The sum of squared standard errors across stimuli provides an estimate of response variance due

to measurement error. This value is then subtracted from total variance in order to obtain explainable variance. The average fraction of this explainable variance captured by the models was 0.32. The corresponding mean r value in the distribution shown here is 0.57. (**d**) Distribution of correlations between observed responses and responses predicted by 2-Gaussian models fit to Fourier transforms of each stimulus (log power in the frequency/orientation domain). The mean r value for this distribution is 0.27 (i.e. approximately 0.07 fractional variance explained). (**e**) Cross-validation analysis of model validity based on double-run experiments. Cross-validation between runs provides the strictest test of model validity, because it requires predicting responses to a wholly unrelated stimulus set that was free to evolve from a different starting point toward a completely different region of complex 3D shape space. Convergence of multiple lineages toward the same result is the best guarantee against finding a local error minimum. We used a permutation/randomization method to test for significant cross-prediction between runs. This test is an adaptation of standard permutation regression methods (B.F.J. Manly, Randomization and Monte Carlo Methods in Biology, Chapman & Hall, 1991, Chap. 6) to our dataset. The question to be addressed is whether correlations between observed responses and model-predicted responses predicted are significant. The null hypothesis is that these correlations are no higher than those expected by chance. If this hypothesis is true, then randomizing (permuting) the relationship between models and cells should not significantly affect correlation values. (This is the basic rationale for every permutation test.) Randomizing this relationship many times (in this case, we used all 528 possible pairings) produces a distribution of correlation values expected under the null hypothesis, against which the original correlation values can be compared. For each pairing, we calculated the correlation between observed responses from one cell with responses to the same stimuli predicted by another cell's best-fit 2-Gaussian model. The average cross-prediction in this chance distribution is –0.0086 (black bars). The distribution of actual (within-cell) cross-predictions (red bars) for 31 models with self-prediction r values (blue bars) exceeding 0.5 had a mean of 0.35. Most (27/31) of these models exceeded the upper 5% point on the chance distribution (0.20; vertical line). The separation between distributions was less complete for simpler (1-Gaussian) and more complex (3-5 Gaussian) shape tuning models. Restricting models to include only Gaussian subunits that explained at least an additional 5% of response variance produced a corresponding preponderance of 2-Gaussian models (see Supplementary Fig. 7).

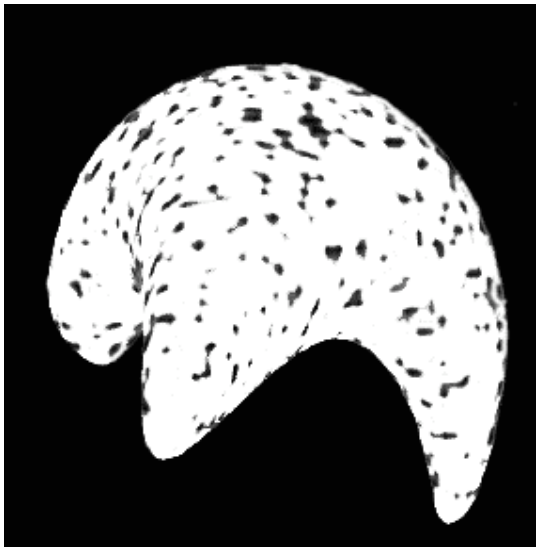


Supplementary Figure 7 | Model order based on explained variance threshold. As an alternate method for determining the appropriate number of Gaussian tuning regions to describe each neuron, we restricted model complexity on the basis of additional variance explained by each successively added subunit. A 5% additional explained variance threshold (left) produced a preponderance of 2-Gaussian models, consistent with the cross-validation analysis described in Supplementary Fig. 6. A more lenient 2.5% threshold produced a range of 2-5 subunits, consistent with our previous analysis of 2D shape configuration tuning using this threshold (Brincat & Connor, Nature Neuroscience 7: 880-886, 2004).

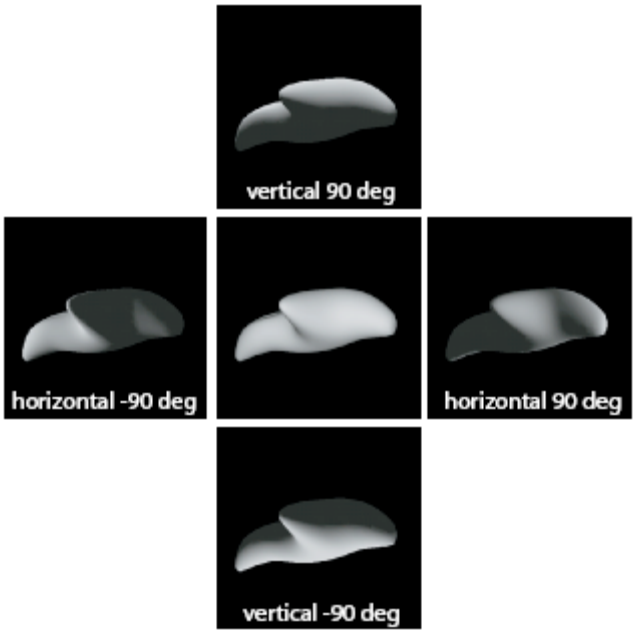


Supplementary Figure 8 | Depth cue test for Fig. 1 neuron. The test was performed for 3 stimuli drawn from the top, middle, and bottom of the response range for this neuron (top, middle, bottom, respectively). Responses were marginally affected by changing depth cues or removing depth cues entirely in the non-stereo, “none” condition.

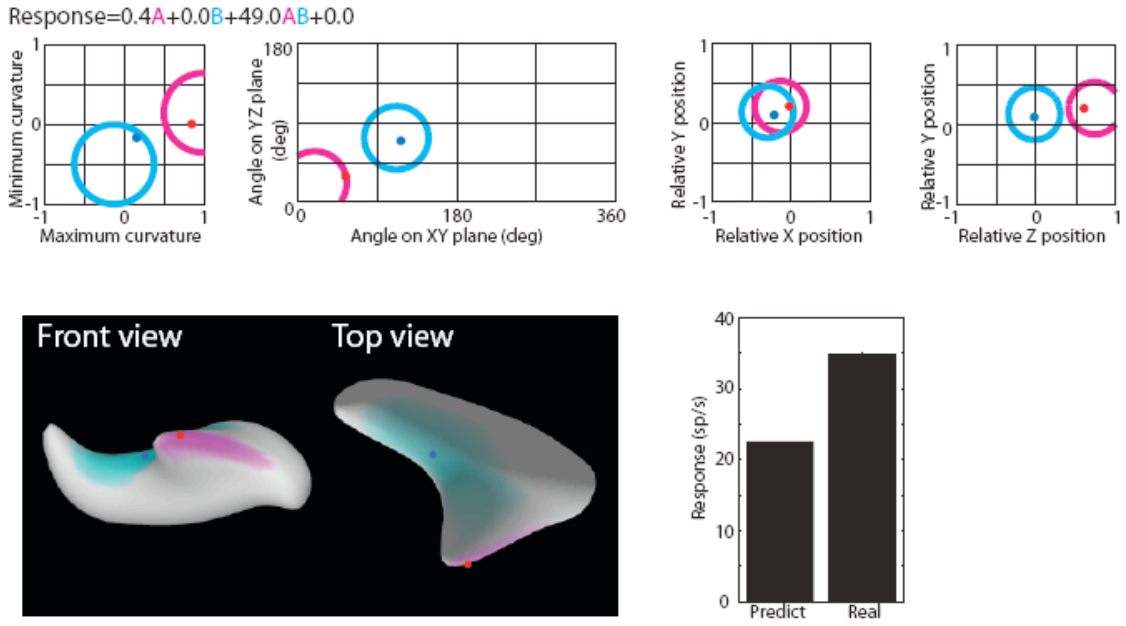
Supplementary Figure 9 | Successive stimulus generations for the Fig. 2 neuron. Average response to each stimulus is indicated by background color.



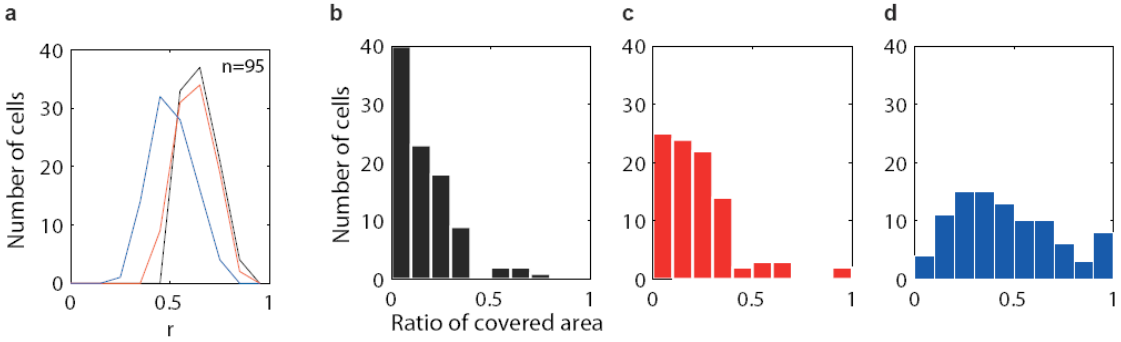
Supplementary Figure 10 | Texture gradient depth cue. This example illustrates the appearance of stimuli when only texture gradient depth cues were present.



Supplementary Figure 11 | Image changes produced by variation in lighting direction. This figure shows stimulus appearance at the extreme values for horizontal and vertical lighting direction, for the high response stimulus in the Fig. 3d test. The shading patterns change dramatically, but the neuron responds consistently to the 3D shape implied by all these shading patterns (Fig. 2d).



Supplementary Figure 12 | Mapping between model domain and 3D surface fragments. The model for the Fig. 2 neuron is shown with the addition of dots on both the model domain and the surface of an example high response stimulus indicating the surface fragments with maximum correspondence to the two Gaussian tuning regions for this neuron. This histogram shows the predicted and observed responses based on these two surface fragments. Other details as in Fig. 2.

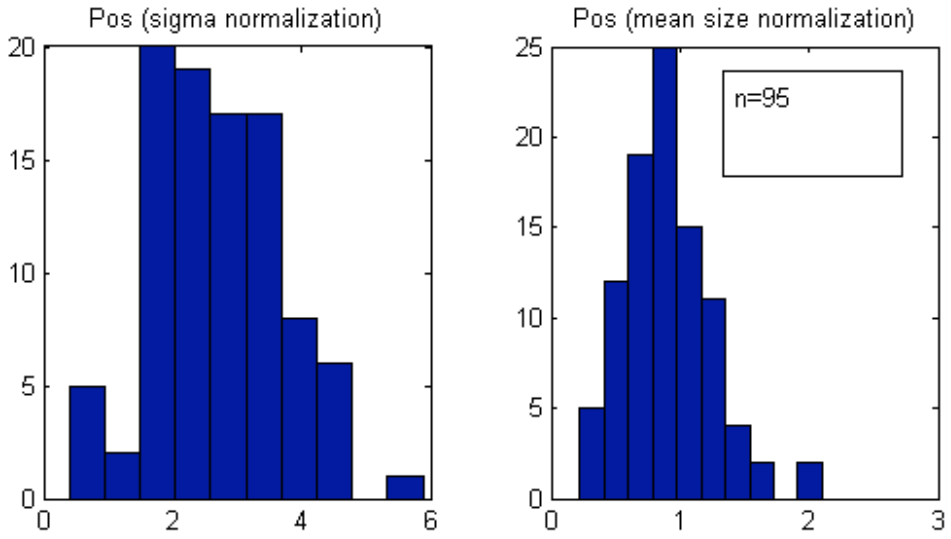


Supplementary Figure 13 | Fraction of stimulus surface area covered by best-fit models. This analysis demonstrates that neural responses were associated with parts-level shape and not holistic or global shape. To estimate the surface structures associated with neural responses as completely as possible, the original fitted tuning regions were extended to include highly correlated parts of the surface domain. To do this, the 7D surface structure domain was binned at a resolution of 0.2 in the curvature dimensions, 0.2 in the position dimensions, and 30° in the orientation dimensions. To calculate the correlation pattern for a given Gaussian subunit, correlation strength for each bin was accumulated across stimuli, by calculating the predicted subunit response for each stimulus (i.e., how closely the stimulus approximated the fitted Gaussian) and summing that value into all the bins occupied by that stimulus. The final value in each bin was divided by the number of stimuli in that bin in order to normalize for sampling.

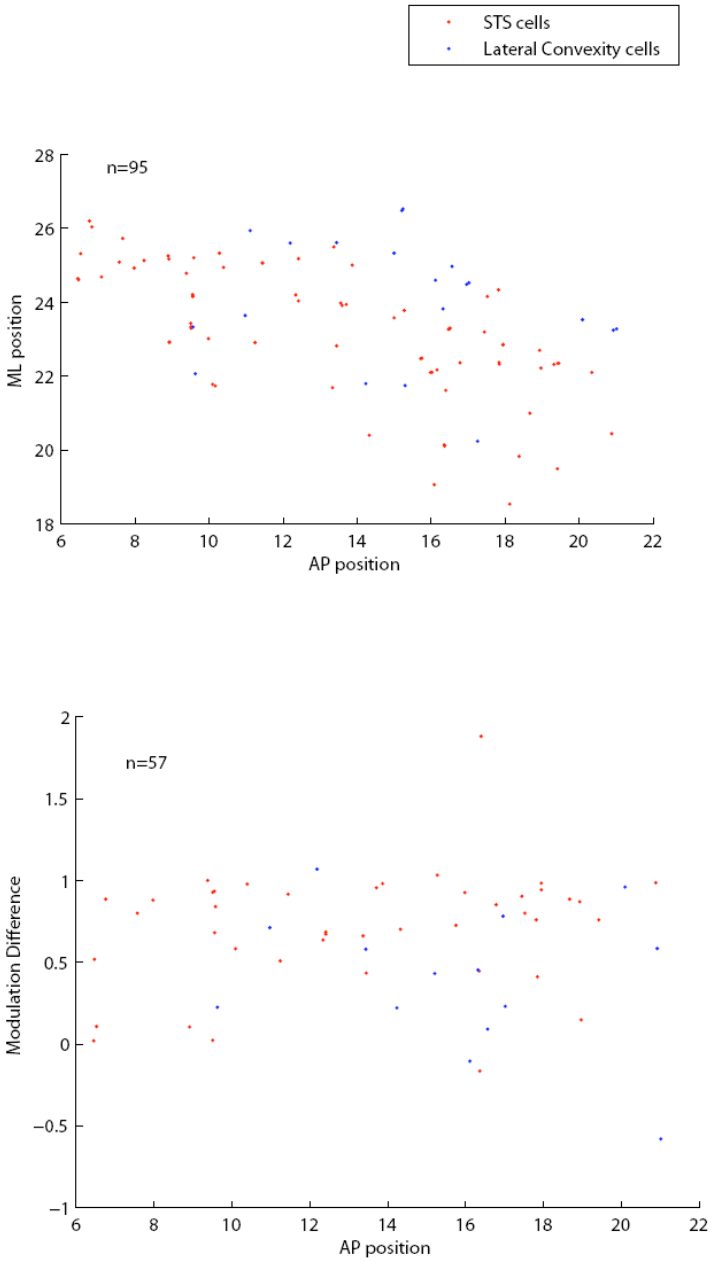
These correlation patterns were used to create extended surface shape models in which each Gaussian was grown into a cluster of Gaussians describing a larger shape configuration. For any given stimulus, the subunit response predicted by the cluster was the sum of predicted responses across all the component Gaussians:

$$Rc = \sum_{m=1}^{\#cluster\ components} Rm$$

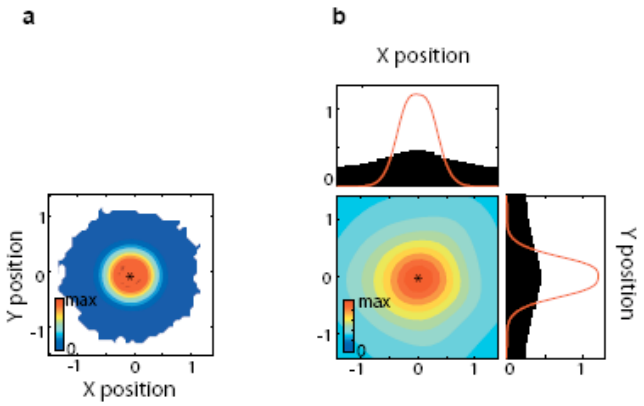
Gaussian functions were added to the cluster iteratively, one for each 7D bin, in order of increasing spatial distance, for bins with correlation values exceeding 0.85. The Gaussian peak in each bin was defined by the average value (in each dimension) across all stimulus samples in that bin. The original standard deviation value was used for all component Gaussians. At each step, the weights for the individual linear and nonlinear terms and the overall amplitude were refitted. This iterative accretion process was terminated when the percent response variance explained by the model fell from the original value by more than 5% of total response variance. In practice, the number of Gaussians in each cluster varied from 0-30. These extended models sometimes included a large number of added parameters, but none of these added parameters were fitted. Correlation between predicted and observed responses and thus explained variance dropped as Gaussians were added to the clusters, demonstrating that shape tuning was spatially restricted and not global. (**a**) Distributions of correlations between predicted and observed responses for the original 2-Gaussian models (black), for cluster models constrained to a drop of no more than 5% explained variance (red), and cluster models allowed to drop 10% or more in explained variance (blue). (**b**) Surface area coverage for the original 2-Gaussian models. For each neuron, the fractional surface area within the 1 standard deviation boundary of either tuning region was averaged across stimuli. The mean fractional coverage is 0.17. (**c**) Surface area coverage for the extended cluster models constrained to 5% drop in explained variance. The mean fractional coverage is 0.23. (**d**) Surface area coverage for cluster models with 10% or more drop in explained variance. The mean fractional coverage is 0.46.



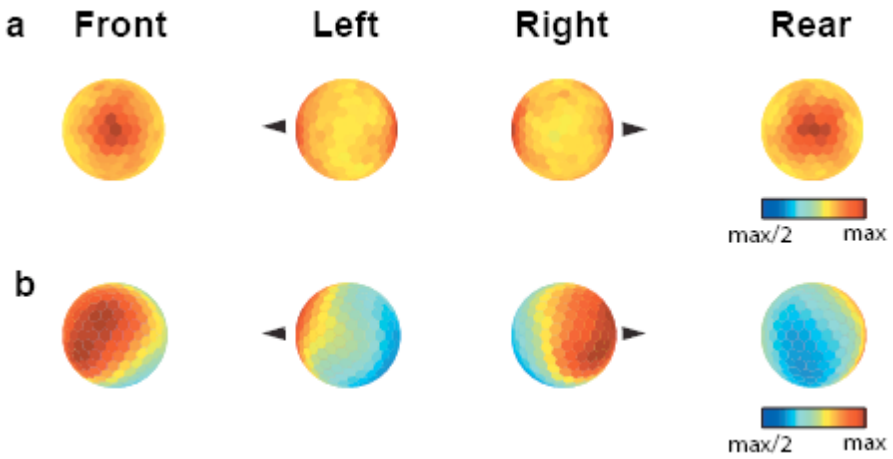
Supplementary Figure 14 | Distribution of distances between surface fragments in the 2-Gaussian models. For each model, the Euclidean distance between Gaussian means in the 3D position space was calculated in terms of standard deviation units (left panel) and average object size (maximum length in any dimension). For most models, the Gaussian tuning means were separated in 3D position by more than 2 standard deviations and more than half the average object size.



Supplementary Figure 15 | Anatomical distribution of 3D tuning. Top: Distribution of medio-lateral and antero-posterior stereotaxic positions of 95 neurons in the ventral bank of the superior temporal sulcus (red) and the inferotemporal gyrus (blue). Bottom: Distribution of differences between 2D and 3D modulation strength for 57 neurons tested with and without depth cues. Positive values correspond to stronger modulation with 3D cues present. The mean modulation difference was significantly higher ($p = 0.028$) for sulcus neurons (red, 0.71) than for gyrus neurons (blue, 0.40). This difference is consistent with previous results suggesting stronger representation of depth in the sulcus (see main text, reference 46).



Supplementary Figure 16 | Position distributions. (**a**) The normalized distribution across all stimuli of surface point positions in the x/y plane. (**b**) The normalized distribution of Gaussian peak positions in the x/y plane is broader than the stimulus distribution, as expected since stimuli were constrained to overlap the fixation point.



Supplementary Figure 17 | Orientation distributions. (**a**) The normalized distribution of surface normal orientations across all 1st generation stimuli for 95 cells is plotted on a spherical surface shown from 4 viewpoints. For 2D object boundaries in the image plane, the domain of contour orientations is circular. For 3D object surfaces, the domain of surface fragment orientations is spherical. Here, the color at each point on the spherical surface indicates the proportion of surface fragments sharing the same orientation in the purely random stimuli that made up the 1st generations in all experiments. There is a slight bias toward normals aligned with the z axis (pointing toward or away from the viewer; dark red). The spherical plot is shown from four viewpoints to make the entire domain visible. In the left and right views, the black arrowhead points toward the viewer. (**b**) The normalized distribution of Gaussian peak surface normal orientations, for the extended models that include Gaussian tuning regions highly correlated with the fitted tuning functions (see Methods). There is an expected bias toward visible surface orientations in the front-facing hemisphere of this domain. There is also a bias toward surface normals pointing more toward the viewer's left, which is characteristic of object surfaces in the left half of the object image. This could reflect a contralateral processing bias, since all the neurons in our sample were

recorded from right hemispheres. There was no apparent bias toward surface normal orientations in the image plane (which would appear as a ring bisecting the left and right views) that would correspond to surface fragments forming the 2D occlusion boundary of the object. The continuity of these orientations with the rest of the 3D orientation domain suggests that tuning for 2D boundary shape constitutes a subspace of 3D surface shape sensitivity.