

Neural Coding for Shape and Texture in Macaque Area V4

✉ Taekjun Kim,^{1,2} Wyeth Bair,^{1,2} and **Anitha Pasupathy^{1,2}**

¹Department of Biological Structure, and ²Washington National Primate Research Center, University of Washington, Seattle, Washington 98195

The distinct visual sensations of shape and texture have been studied separately in cortex; therefore, it remains unknown whether separate neuronal populations encode each of these properties or one population carries a joint encoding. We directly compared shape and texture selectivity of individual V4 neurons in awake macaques (1 male, 1 female) and found that V4 neurons lie along a continuum from strong tuning for boundary curvature of shapes to strong tuning for perceptual dimensions of texture. Among neurons tuned to both attributes, tuning for shape and texture were largely separable, with the latter delayed by ~30 ms. We also found that shape stimuli typically evoked stronger, more selective responses than did texture patches, regardless of whether the latter were contained within or extended beyond the receptive field. These results suggest that there are separate specializations in mid-level cortical processing for visual attributes of shape and texture.

Key words: monkey; object recognition; rhesus macaque; shape; texture; ventral visual pathway

Significance Statement

Object recognition depends on our ability to see both the shape of the boundaries of objects and properties of their surfaces. However, neuroscientists have never before examined how shape and texture are linked together in mid-level visual cortex. In this study, we used systematically designed sets of simple shapes and texture patches to probe the responses of individual neurons in the primate visual cortex. Our results provide the first evidence that some cortical neurons specialize in processing shape whereas others specialize in processing textures. Most neurons lie between the ends of this continuum, and in these neurons we find that shape and texture encoding are largely independent.

Introduction

Area V4 is an important intermediate stage in the ventral visual pathway specialized for object recognition. Many studies have reported that V4 neurons are selective for the shape of visual stimuli (Desimone and Schein, 1987; Kobatake and Tanaka, 1994; Gallant et al., 1996; Pasupathy and Connor, 2002; Nandy et al., 2013; El-Shamayleh and Pasupathy, 2016), and many others have reported V4 selectivity for surface properties (e.g., color, brightness, and texture) (Zeki, 1973; Schein and Desimone, 1990; Heywood et al., 1992; Conway and Tsao, 2006; Arcizet et al., 2008; Bushnell et al., 2011; Namima et al., 2014; Okazawa et al.,

2015). But because most previous studies have focused exclusively on the encoding of either shape or surface characteristics, we know little about how both types of information are multiplexed in the responses of individual neurons.

We investigated how form and texture information are jointly encoded in primate V4, a question that has received strikingly little attention anywhere in visual cortex compared with, for example, the multiplexing of form and color signals (Livingstone and Hubel, 1988; Johnson et al., 2001, 2008; Lennie and Movshon, 2005; Conway et al., 2007; Shapley and Hawken, 2011; Bushnell and Pasupathy, 2012). Nevertheless, this is important to address because it has been theorized that early and mid-level stages of the ventral visual pathway are specialized for encoding textures rather than the boundaries of objects (Adelson, 2001; Movshon and Simoncelli, 2014): in the terminology of Adelson (2001), the encoding of “stuff” rather than “things.” This assertion is based on the argument that much of the visual world is made of stuff, even the surface of things are made of stuff, and that selectivity for local orientation and spatial frequency (SF) in V1 and selectivity for texture in V2 and V4 may all be interpreted as encoding the stuff in an image (Adelson, 2001; Movshon and Simoncelli, 2014; Ziomba and Freeman, 2015). Indeed, bottom-up shape selectivity could arise as a consequence of a statistical representation of natural scenes (Movshon and Simoncelli, 2014; Ziomba and Freeman,

Received Dec. 6, 2018; revised March 19, 2019; accepted April 1, 2019.

Author contributions: T.K., W.B., and A.P. designed research; T.K. performed research; T.K. analyzed data; T.K. wrote the first draft of the paper; T.K., W.B., and A.P. edited the paper; T.K., W.B., and A.P. wrote the paper.

This work was supported by National Eye Institute Grant R01 EY018839 to A.P., National Science Foundation Collaborative Research in Computational Neuroscience Grant IIS-1309725 and National Eye Institute Grant R01 EY029997 to A.P. and W.B., National Eye Institute Grant R01 EY027023 to W.B., National Eye Institute Center Core Grant for Vision Research P30 EY017130 to the University of Washington, and National Institutes of Health/Office of Research Infrastructure Programs Grant P51 OD010425 to the Washington National Primate Research Center. We thank all laboratory members and Dr. Dina Popovkina for helpful discussions and comments on the manuscript; and Amber Fyall for assistance with animal training.

The authors declare no competing financial interests.

Correspondence should be addressed to Anitha Pasupathy at pasupat@u.washington.edu.

<https://doi.org/10.1523/JNEUROSCI.3073-18.2019>

Copyright © 2019 the authors

2015), allowing no mechanistic basis for a distinction between shape and texture selectivity, at least through mid-level processing. In this construct, recognition of objects may be based on distinct form selectivity that emerges only in the highest stages of the ventral pathway. While attractive, this idea remains conjecture in the absence of a direct comparison between the responses of individual neurons to shapes and textures.

To determine whether, and how, stimulus information related to form and texture are multiplexed by individual neurons, we directly compared the responses of V4 neurons with a variety of shape and texture stimuli. Our stimulus set included a subset of shape stimuli previously used to characterize tuning for boundary curvature in V4 (Pasupathy and Connor, 2001) and a custom-designed set of texture stimuli inspired by human texture perception (Tamura et al., 1978; Liu and Picard, 1996; Rao and Lohse, 1996). We also studied the responses to combination stimuli in which a texture was painted on the surface of preferred and nonpreferred shapes, the latter being determined for each neuron. To determine whether simple image elements, in terms of local orientation and SF information, can explain V4 responses to both shapes and textures, we evaluated whether a hierarchical max (HMax) model of V4 responses (Serre et al., 2005; Cadieu et al., 2007), built by pooling phase-invariant, oriented units, can provide a good fit for the observed data. Our results provide key insights into the differential encoding of shape and texture in V4.

Materials and Methods

Animal preparation

Two macaque monkeys (*Macaca mulatta*; M1: 9 years old, male; M2: 10 years old, female) participated in this study. All animal procedures conformed to National Institutes of Health guidelines and were approved by the Institutional Animal Care and Use Committee at the University of Washington. Each monkey underwent a surgery for headpost implantation followed by several months of training to perform a fixation task during a receptive field (RF) mapping procedure and visual stimulation for the main experiment (details below). A V4 recording chamber was placed over the left prelunate gyrus on the basis of a preoperative structural MRI scan. We made a small (10–15 mm) craniotomy over V4, 1–2 days before the start of recording.

Data collection

In each recording session, one tungsten microelectrode (FHC) was advanced perpendicular to the brain surface until a well-isolated single-unit signal was obtained. Neural signals were amplified and filtered (band-passed signal between 150 Hz and 8 kHz) by a data acquisition system, MAP software (Plexon, RRID:SCR_003170). Time stamps of spiking activity, eye position (EyeLink, SR Research, RRID:SCR_009602), and stimulus events (based on photodiode signal) were stored at 1 kHz sampling rate for off-line analysis (Offline sorter, Plexon, RRID:SCR_000012; MATLAB, The MathWorks, RRID:SCR_001622).

Once a well-isolated single unit was identified, RF location was determined by a hand-mapping procedure. To avoid biased sampling of shape-selective or texture selective neurons, we used a variety of visual stimuli for initial hand-mapping of RFs, including 2D shapes with different boundary curvature functions (Pasupathy and Connor, 2001) and sinusoidal, hyperbolic, and polar gratings (Gallant et al., 1993). After the RF mapping procedure, we conducted the main experiment with a standard set of shape and texture stimuli (described next). During RF mapping and the main experiment, monkeys held their gaze within the fixation window (1° radius) while 4–5 visual stimuli were presented sequentially within the RF of the cell under study. Each stimulus was presented for a 300 ms duration preceded by a 300 ms blank interstimulus interval. Each stimulus was repeated in random order, multiple times. Only cells with at least six repetitions for each stimulus condition were included in the data analysis.

Visual stimuli

The main experiment included 394 stimulus conditions: 225 shapes; 168 textures; 1 blank condition for assessing baseline activity. Details of the shape and texture stimulus groups are described below. The size of visual stimuli was scaled with the RF eccentricity so that all parts of the shape stimulus were within the estimated RF region (see Fig. 1C), where the estimated diameter was $1.0^\circ + 0.625 \times$ RF eccentricity (based on Gattass et al., 1988). All shape and texture images were adjusted to have the same mean RGB pixel values, [100, 100, 100], equal to ~ 16 cd/m 2 , and were presented against a gray background (8 cd/m 2).

2D shape stimuli. We used a set of 30 closed shapes to probe shape tuning. This is a subset of the standard set of shapes constructed by Pasupathy and Connor (2001) based on a systematic combination of convex and concave boundary fragments (see Fig. 1A). Each stimulus was presented at 1, 2, 4, or 8 orientations (in 45° increments) depending on rotational symmetry, and the circle stimulus was presented at three luminance contrasts (1, 16, 46 cd/m 2), for a total of 225 shapes.

Texture patches. To choose a tractable set of texture stimuli that span a broad range of perceptual qualities of texture, we worked within a 3D space defined by axes that are widely considered to be relevant for human texture perception: coarse versus fine, directional versus nondirectional, and regular versus irregular (Tamura et al., 1978; Liu and Picard, 1996; Rao and Lohse, 1996). We devised simple methods to quantify the degree of coarseness, directionality, and regularity in a given image, represented many candidate textures in this space and then chose a subset that sampled the space along all three dimensions.

We defined coarseness by the method of Rosenfeld and colleagues (Hayes et al., 1974), which measures the size of the elements forming the texture. At every point of texture image, we computed the average pixel value over different sized neighborhoods ($2^k \times 2^k$, where $k = 0, 1, 2, \dots, 6$). We then computed the difference of these values between neighboring pairs of points in the vertical and horizontal directions of the texture image, with nonoverlapping neighborhoods. The value of k that yielded the biggest difference in either direction provides a measure of the size of the texture element at that point. The average of k values over the entire image was taken as a coarseness measure. A small k implies a fine image while a large k implies a coarse image.

To compute the level of directionality of an image, we first performed a 2D Fourier transform on the texture images to get a magnitude map $F(sf, \theta)$, where sf and θ indicate the SF and angle, respectively, in polar coordinates. We divided this map into 8 orientation bands (22.5° each) and computed the average magnitude in each band. The summed magnitude in the top two orientation bands normalized (divided) by the overall magnitude across all bands was our measure of directionality. This metric provides a measure of oriented energy in an image independent of the specific orientation of the image.

Our regularity index was a quantitative measure of nonrandom repetitive pattern in a texture image. This measure is given by the highest peak prominence in the 2D autocorrelation map of an image. The 2D autocorrelation map, $\rho(x, y)$ is obtained with the following equation:

$$\rho(x, y) = \sum_{m=1}^M \sum_{n=1}^N I(m, n) \times I(m + x, n + y) / \sum_{m=1}^M \sum_{n=1}^N I^2(m, n)$$

where $I(m, n)$ indicates $M \times N$ texture image.

To build our texture stimulus set, we first calculated the three indices for a set of 112 textures in the Brodatz texture dataset (a commonly used texture dataset) (Brodatz, 1966), and 32 textures from a commercial library (www.textures.com). Raw scores were transformed into z scores, and each image was placed in one of two groups along each dimension: that is, fine or coarse, regular or irregular, and directional or nondirectional, based on the sign of the z scored value along that dimension. We then constructed 8 texture categories corresponding to the two possible groups along each dimension. Our texture stimulus set (see Fig. 1B) includes 2–3 textures for each of the 8 categories. Each texture was represented through a circular aperture. To dissociate selectivity for a specific instantiation of a texture (i.e., phase of orientation) from higher-order tuning for texture statistics, we presented each texture at 4 orientations in 45° increments. We also presented each texture at two sizes:

one that matched the size of the large circle (see Fig. 1A, first shape) and was completely contained within the estimated RF size and a second that was twice as large. The two scale conditions were achieved by applying two different sized circular masks to the same texture stimulus. Therefore, textures shown through small and large apertures are identical within the RF, but textures with large aperture have additional information in the RF surround (see Fig. 1C). Total number of texture stimuli was 168 (21 textures \times 4 orientations \times 2 apertures); this includes 84 small and 84 large aperture textures. In addition to shape stimuli and texture patches, we also interleaved 40 natural scenes, but responses for these are not analyzed here.

Textures through shape apertures. We conducted control experiments to determine whether neuronal responses are jointly modulated both by texture and shape attributes of a stimulus. In 43 neurons, we studied the response to 30 (3×10) stimuli constructed by presenting 10 textures through three different shape boundaries. One of these shapes was always a circle. The other two were customized to each neuron and included one shape that evoked strong responses and another that evoked weak responses. We used the same 10 nondirectional textures (see Fig. 1B, bottom two rows) for all neurons.

Data analysis

Quantification of neural response magnitude. For each stimulus, we computed average response rate by counting spikes within a window from 50 to 400 ms after each stimulus onset to allow for onset and offset response latency of V4 neurons (Zamarashkina et al., 2017).

Permutation test. We asked whether the best response to a shape stimulus was larger than the best response to a texture patch within the RF (small aperture texture) (see Fig. 2B). To determine whether the higher best response for shape stimuli across V4 neurons could be simply due to the larger set of shape stimuli (shape = 225; small aperture texture = 84), we conducted a permutation test. For each permutation, we randomly reassigned the shape and texture responses and computed the difference between the best shape versus texture responses. We repeated this process 10,000 times. The one-sided p value of the test calculated as the proportion of sampled permutations where the difference in best response values was greater than or equal to the observed difference. We conducted a similar analysis comparing best shape response to best large aperture texture response.

HMax model simulation. The difference in responses between shape and texture stimuli could result from differences between the two sets of stimuli in terms of simple image features, specifically orientation and SF content, and the conjunctions of orientations. If this was the case, the HMax model, previously shown to provide a good description of V4 responses to shape stimuli (Riesenhuber and Poggio, 1999; Serre et al., 2005; Cadieu et al., 2007) in terms of orientation conjunctions, should explain overall trends in shape and texture selectivity observed in our V4 data.

Briefly, the HMax model consists of four (S1, C1, S2, C2) layers. The selectivity (i.e., template matching) and invariance (i.e., max pooling) operations are performed in alternating layers. S1 units correspond to simple cells in V1. They are designed to have Gabor RF profiles with six different sizes and four orientations, and their responses are determined by normalized dot product of the Gabor filter and the image patch within the RF. Outputs from S1 units are then max pooled to build C1 units with larger orientation-tuned RFs. S2 units receive inputs from a combination of C1 units with various sizes, positions, and orientations. Thus, S2 units build selectivity for complex patterns by pooling signals from a variety of orientations. Last, the same max pooling operation is repeated between S2 and C2 layers. The response of S2 unit is given by the following:

$$r = g\left(\frac{\sum_i w_i x_i}{\sum_i x_i^2 + k}\right)$$

where x_i and w_i are the response and the synaptic weight of i th C1 unit. The constant k (0.0001) prevents division by zero, and $g(u)$ is the sigmoid function to implement an inhibitory mechanism and defined by the following:

$$g(u) = \frac{s}{1 + \exp(-\alpha(u - \beta))}$$

where α , β , and s are free parameters to determine the shape of the sigmoid function. Finding a HMax model instantiation that best describes the responses of a V4 neuron amounts to finding the weights, w_i , from a set of C1 to S2 units that best describes the response. To avoid overfitting, we restricted the number of C1 units to 13, the median value of the optimal number of C1 units determined by a cross-validation procedure in the original paper (Cadieu et al., 2007). Thus, 16 parameters (synaptic weights for 13 C1 units and 3 sigmoid parameters) were adjusted to simulate individual neural responses (for further details, see Cadieu et al., 2007). We evaluated the model fitting performance (correlation coefficients on the training and test sets) by the median of 10-fold cross-validation results and found that shape responses of most neurons were successfully predicted by the HMax model (see Results). The HMax model fits to texture data were based only on the responses to small texture stimuli that were confined to the RF.

APC model for curvature selectivity. To quantify curvature selectivity of each neuron, Pasupathy and Connor (2001) proposed the angular position and curvature (APC) model in which neural responses are fit with a 2D Gaussian function in the plane of boundary curvature and object-centered angular position. In this space, curvature values ranged from −0.3 (shallow concave) to 1.0 (sharp convex), and angular position progressed from right (0°) in a counterclockwise direction: i.e., right, top (90°), left (180°), and bottom (270°). Each shape in our stimulus set included multiple convex and concave features, and the neuron's response were modeled as the maximum response predicted across all component features. Thus, the predicted response r is given by the following:

$$r = \max_p \left[k \cdot \prod_{i=1}^2 e^{-(X_{ip} - \mu_i)^2 / 2 \cdot \sigma_i^2} \right]$$

where μ and σ indicate the mean and SD of the curvature and angular position dimensions (indexed by i), k represents the amplitude of the 2D Gaussian, and X_{ip} represents the curvature ($i = 1$) and angular position ($i = 2$) value of a specific feature p . Therefore, this model allows us to see how response of a single unit is well explained by the preference to a particular boundary feature.

Regression model for texture selectivity. To ask how the perceptual dimension (coarseness, directionality, regularity, and contrast) modulate neural responses to texture, we conducted linear regression analyses with these four dimensions as independent variables. The formula is as follows:

$$Z_{\text{response}} = \beta_1 \cdot Z_{\text{Coarseness}} + \beta_2 \cdot Z_{\text{Directionality}} + \beta_3 \cdot Z_{\text{Regularity}} + \beta_4 \cdot Z_{\text{Contrast}}$$

where both independent and dependent variables are standardized so that their means are equal to 0 and SDs are equal to 1. Here, we additionally defined a contrast index that was not used in the texture selection procedure. The contrast index, I_{Contrast} , is given by the following:

$$I_{\text{Contrast}} = \sigma / (\alpha_4)^n$$

where σ and α_4 are the SD and kurtosis of the gray level distribution of an image and $n = 1/4$ was fixed (Tamura et al., 1978).

Separability of tuning for shape and texture. Based on the control data where we varied shape and texture simultaneously, we assessed whether tuning for shape and texture could be described as mathematically separable (i.e., whether the shape preference was consistent under different texture conditions and vice versa). First, we estimated the one-dimensional shape and texture tuning functions by averaging over the texture and shape dimensions, respectively. The product of these functions represented the predicted responses for stimuli defined by both shape and texture features. Separability was quantified for each neuron

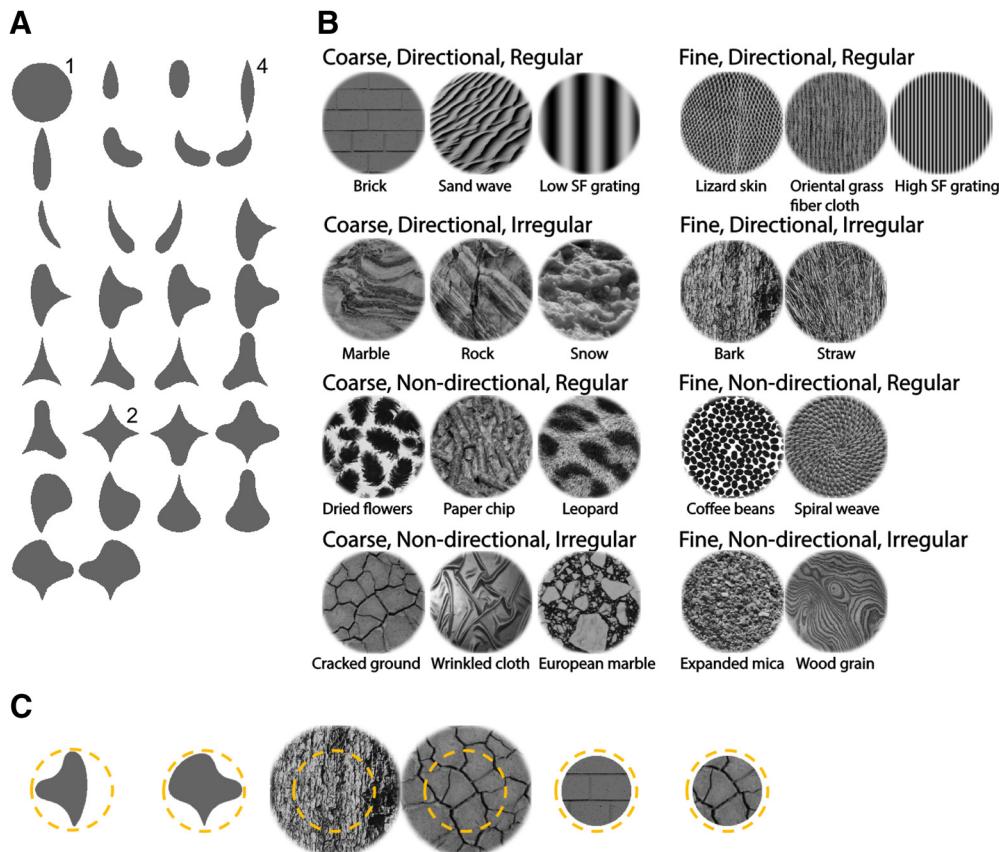


Figure 1. Visual stimuli. **A**, Shape stimuli. We used a subset (30 of 51) of the 2D shapes developed by Pasupathy and Connor (2001) to study how boundary conformation influences V4 responses. Most shapes were presented at 8 rotations (45° increments); a few shapes (those identified with a superscript) were presented at fewer rotations (1, 2, or 4, as noted in figure) due to rotational symmetry. The circle was presented at three luminance contrast levels (1, 16, 46 cd/m²) relative to the background (8 cd/m²), for a total of 225 shape stimuli. **B**, Texture stimuli. We constructed eight (2^3) texture categories based on three dimensions that influence human texture perception (coarse vs fine, directional vs nondirectional, regular vs irregular), and selected 2–3 representative textures for each category (see Materials and Methods). Each texture was presented through a circular aperture of two sizes and at four orientations at 45° increments, for a total of 168 texture stimuli. **C**, Examples of shape, large aperture texture, and small aperture texture stimuli. All parts of the shape stimulus were within the estimated RF region (yellow dotted line). The diameter of the large aperture texture stimulus was twice that of the estimated RF. Small aperture texture condition was created by applying a RF sized circular aperture to large aperture texture.

by computing the correlation coefficient between measured and predicted responses. For comparison, we also considered an additive model where the predicted responses were given by the sum of the components of shape and texture tuning functions.

Effect sizes of shape and texture. Based on the control data where we varied shape and texture simultaneously, we performed two-way ANOVA to calculate the effect sizes of shape and texture variables. The effect size, η^2 , is defined as the proportion of total variance that is attributable to an effect of interest (Cohen, 1973). Its formula is as follows:

$$\eta^2 = \frac{SS_{\text{Effect}}}{SS_{\text{Total}}}$$

where SS represents the sum of squares used in a two-way ANOVA.

Onset of shape- and texture-dependent response modulation. To compare the time course of shape and texture-dependent response modulation, peri-stimulus time histograms (PSTHs) were constructed by convolving spike trains with a Gaussian kernel ($\sigma = 5$ ms). For each stimulus group (i.e., shapes, large and small aperture textures), Mann–Whitney U test within a 20 ms sliding window (moving in 1 ms steps) was conducted to determine whether mean spike count from top 50% preferred stimuli significantly deviated from that for bottom 50% nonpreferred stimuli ($p < 0.01$). The onset time for shape or texture selectivity was determined as the earliest time when Mann–Whitney U test results were significant in 30 consecutive windows.

Experimental design and statistical analysis

Details of experimental procedure and visual stimuli are described above (see Data collection, Visual stimuli). For all statistical tests presented

here, independent group comparisons were performed using non-parametric Mann–Whitney U test. And paired-data comparisons were performed with the Wilcoxon Signed Rank test. The strength of linear relationship between pairs of variables was assessed by Pearson's correlation coefficient. A p value <0.05 was considered significant.

Data and software availability

The data and analysis code that support the findings of this study are available from the corresponding author upon reasonable request.

Results

We studied the responses of 101 single units in two macaque monkeys (M1: 47 cells; M2: 54 cells) to a variety of shapes and textures as shown in Figure 1. The shapes (Fig. 1A) are a subset of those used previously to systematically characterize tuning for boundary curvature in V4 (Pasupathy and Connor, 2001), whereas the textures (Fig. 1B) span three fundamental dimensions for texture perception: regularity, coarseness, and directionality (Tamura et al., 1978; Liu and Picard, 1996; Rao and Lohse, 1996).

We found that many neurons were driven well by our shape and texture stimuli, but most showed a significant bias for one or the other stimulus set. For example, Neuron 1 (Fig. 2A, left-most) fired up to 32 Hz for shape stimuli (green; maximal response is normalized to 1), whereas no texture stimulus in our set, small or large (red and blue, respectively), caused the neuron to fire $>20\%$ of the maximum shape response.

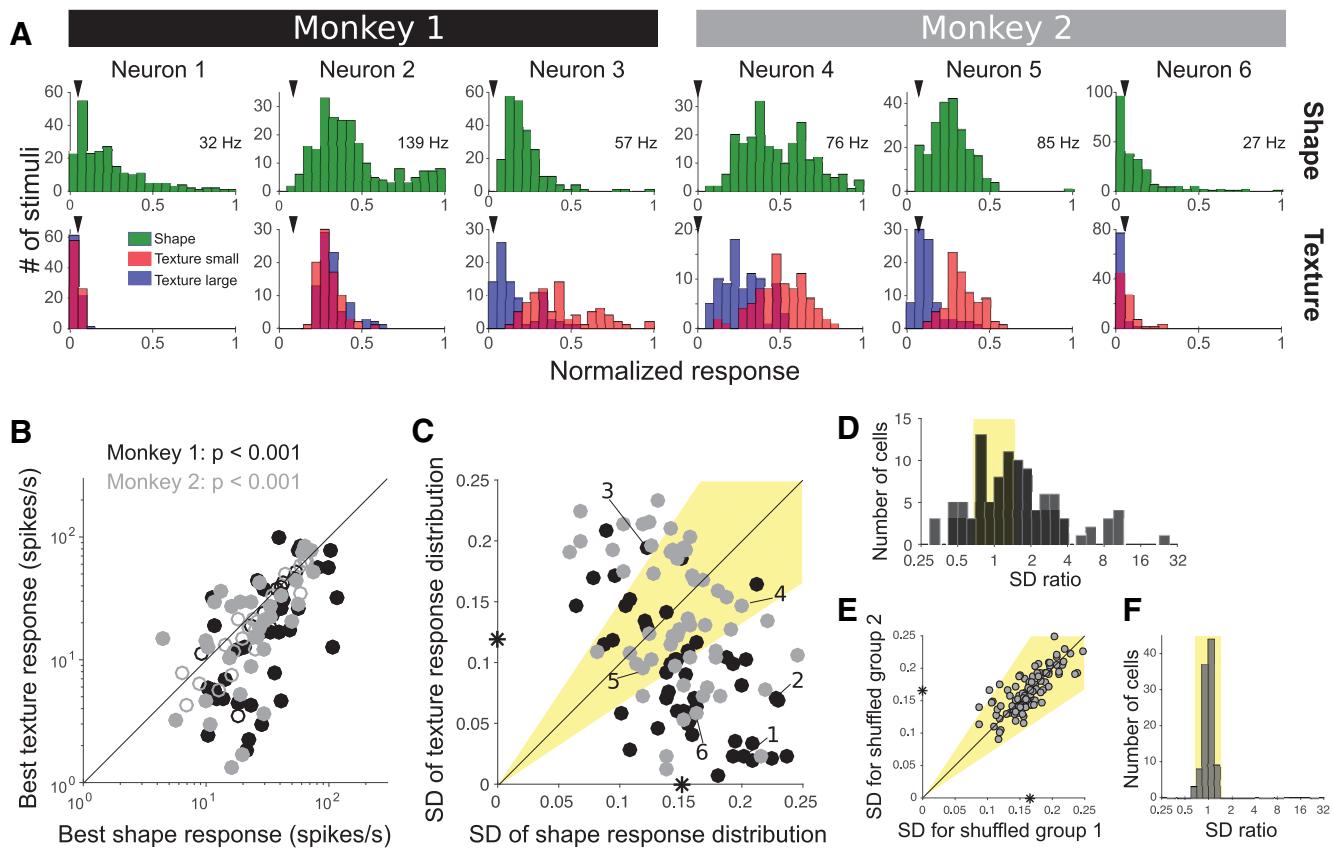


Figure 2. Example neurons and population results. **A**, Response frequency histogram for shape (top row) and texture (bottom row) stimuli for 6 example neurons (columns). Red and blue histograms represent responses to small and large aperture textures, respectively. Responses for each neuron were normalized to the maximum across all shape and texture stimuli (maximum values are shown for each neuron). Triangles represent the background responses (no visual stimulus). **B**, Maximum response across all shape stimuli (*x* axis) is plotted against the maximum response across all small aperture texture stimuli (*y* axis) for each neuron from Monkey 1 (black) and Monkey 2 (gray). Filled symbols represent neurons with a statistically significant difference between the strongest shape and texture response, assessed with a permutation test (see Materials and Methods). In both monkeys, the maximum response across shape stimuli was typically greater than the maximum response across texture stimuli. **C**, SD of the response frequency histogram for shape (*x* axis) is plotted against that for texture stimuli (*y* axis). Asterisks indicate mean value. Yellow highlight represents region where SD ratio for shape versus texture lies between $2/3$ and $3/2$. Points corresponding to examples in **A** are identified. **D**, Histogram of the SD ratio: $SD_{shape}/SD_{texture}$. Yellow highlight as in **C**. **E**, **F**, SD values for shuffled shape and texture responses. Shape and texture responses for each neuron were shuffled and SDs were recomputed for two randomly divided groups: Group 1 ($N = 225$, number of shapes) and Group 2 ($N = 84$, number of textures). **F**, Gray bars represent SD ratios computed from **E**. This process was repeated 10,000 times and width of the distribution from the observed data, quantified by the interquartile range of log (SD ratio), was always at least 5 times broader than that from the shuffled data.

Neurons 2 and 6 (Fig. 2A) also showed stronger and more broadly distributed responses to shapes than to texture. In other cases, however, shape and texture response distributions were similar (e.g., Neurons 4 and 5) both in terms of peak response and width of the distribution, or the texture response distribution was broader than that for shapes (Neuron 3). For all six examples, the maximum and the range of responses to textures that extend beyond the RF (large aperture, blue) were similar or smaller than those for small aperture textures (red), likely due to surround suppression.

To summarize these observations across the population, we characterized the response frequency histograms for shape and texture using two simple metrics each: the maximum response magnitude and the SD of the normalized responses. Across the population, we found that the peak response across shape stimuli was positively correlated with the peak response across texture stimuli (Fig. 2B; $r = 0.65$, $p < 0.001$ for best response). This was also true when mean responses were considered ($r = 0.57$, $p < 0.001$ for normalized mean response). But the strongest neural response for 2D shape stimuli was significantly greater than that for texture stimuli confined to the RF (Wilcoxon Signed Rank test: $p < 0.001$ in Monkey 1, $p = 0.009$ in Monkey 2; Fig. 2B). To

rule out the possibility that this was simply because our shape set was larger (225 shapes vs 84 small-aperture textures), we examined statistical significance of individual neurons using a permutation test (see Materials and Methods). Figure 2B (filled symbols) identify neurons for which there was a statistically significant ($p < 0.05$) difference between the best shape and best texture responses. For 52 of 101 neurons across our population, responses to the preferred shape were significantly greater than those to the preferred texture, factoring out stimulus set size. On the other hand, only 19 neurons had responses to the best texture stimulus that significantly exceeded the best shape response.

To compare the relative spread of responses for shapes and textures, we first normalized the responses to discount the influence of peak firing rates and then computed the SD for each response distribution (Fig. 2C). As with the example neurons, we found that some neurons exhibited a broader range of responses for shape stimuli (Fig. 2C, points below the diagonal), whereas others exhibited a broader range for texture stimuli (Fig. 2C, above the diagonal); still other neurons lie along the diagonal because both stimulus classes evoked similar ranges of responses. The ratio of the SDs for shape versus texture responses (Fig. 2D), which is the same regardless of whether responses are normal-

ized, provides an intuitive visualization of the relative spread of responses (Fig. 2D, histogram). Across our population, the ratio of SDs span a broad range representing the continuum from texture selective (values $\ll 1$) to shape-selective neurons (values $\gg 1$). The breadth of this SD ratio histogram is significantly and substantially larger than expected by chance, which is largely confined to the yellow region ($2/3 < \text{SD ratio} < 3/2$) indicated in Figure 2E, F. The ratio of SDs for shape versus texture was also skewed: there was a small group of highly shape-selective neurons (SD ratio > 2) that was not matched by a similar subpopulation for textures. Finally, we also found that the average SD for shape was greater than that for texture (Fig. 2C, asterisks, shape = 0.15, texture = 0.12; Mann–Whitney U test, $p < 0.001$), consistent with firing rates being greater for shape (Fig. 2B). Overall, these results support the hypothesis that there exists a continuum of neurons in V4 ranging from those specialized to encode texture to those specialized to encode shape.

In our study, the shape and texture stimuli were equated for mean luminance, so the observations in Figure 2C cannot be due to a simple difference in overall luminance. To determine whether the spread of SD ratios observed in Figure 2C was due to differences between our particular shape and texture stimulus sets in terms of fundamental spatial features, such as orientation conjunctions or more generally the SF \times orientation content, we fit each neuron's shape responses to the HMax model of V4 shape selectivity proposed by Cadieu et al. (2007), and then compared the predicted SD for shape and texture responses across the set of best-fitting models. The HMax model, which has been previously shown to provide a good fit for V4 shape responses (Cadieu et al., 2007; Wei and Dong, 2015), builds shape selectivity by pooling the output of oriented filters (see Materials and Methods) and can thus provide a good fit for responses dictated by combinations of orientation at different spatial frequencies and relative locations. For each V4 neuron, we identified the HMax model that provided the best fit to the observed shape responses. Figure 3A–C shows the results for an example neuron that responded strongly to a variety of shapes, all of which included a sharp convexity to the left. The locations and orientations of C1 subunits (Fig. 3B, ellipses) are consistent with this shape preference. Shape responses predicted by the best-fitting HMax model were strongly correlated with the measured responses ($r = 0.94$ on the training set and 0.92 on the test set), and the response range across shapes for the model and neuron were comparable (SD for the observed shape responses = 0.24; SD for predicted shape responses = 0.22). This model predicts a broad range of responses to texture stimuli (Fig. 3C, red dots, SD = 0.21), but the observed texture responses were weak and spanned a narrow range (SD = 0.03). A second example is illustrated in Figure 3D–F for a neuron that was selective for shapes having a concave curvature at the bottom. Again, the best fitting HMax model provided a good fit for shape responses ($r = 0.85$ on the training set and 0.76 on the test set; SDs are 0.16 and 0.15 for observed and predicted responses, respectively), but a very poor fit for texture responses ($r = 0.12$; SD = 0.06 and SD = 0.13 for observed and predicted responses, respectively).

The results across all neurons were consistent with the examples in Figure 3. The best fitting HMax models, optimized based on shape responses, provided a good fit for shape (median $r = 0.75$ on the training set and 0.61 on the test set) but a poor fit for the texture responses (median $r = 0.02$) (Fig. 4A, B). These best-fitting models also predicted that response ranges (SD values) should be similar for shape and texture stimuli for most neurons (Fig. 4C). This resulted in a narrow spread of SD ratios (Fig. 4D)

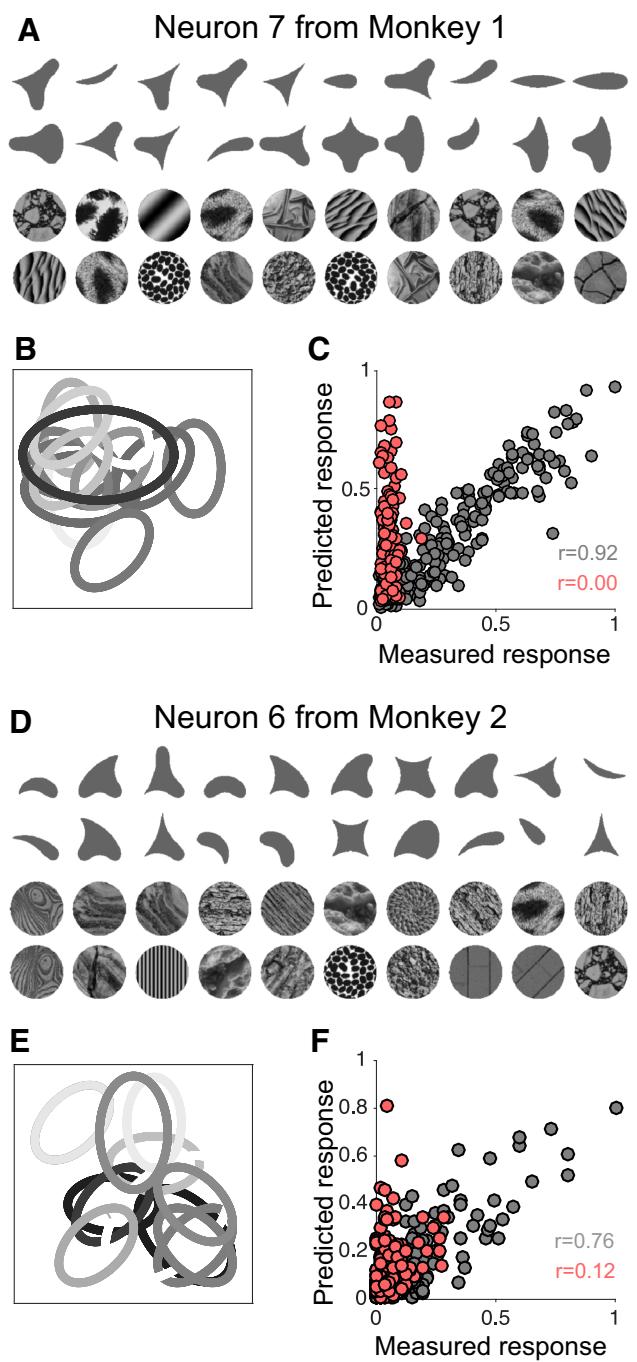


Figure 3. HMax model prediction of responses to texture stimuli. **A**, The top 20 preferred shape and texture stimuli of an example neuron (#7). **B**, Shape template for the S2 unit corresponding to the best fitting HMax model based on the responses to shape stimuli. Each ellipse indicates position, orientation, and size of complex-cell like subunit (C1 unit). Grayscale represents weighting strength with darker color denoting stronger weight. **C**, Predicted responses (y axis) based on the best HMax model fit (shown in **B**) for shape (gray) and texture (red) stimuli are plotted against measured responses (x axis). For this neuron, the HMax model provided an excellent fit for shape responses, but not for texture responses. Predicted texture responses showed a much broader range than the observed data. **D–F**, The results from another example neuron. The same conventions as in **A–C**.

compared with the V4 data (Fig. 2D). Specifically, few points fell outside of the yellow region. Thus, the HMax model, which predicts V4 shape responses based on a combination of inputs varying in orientation, scale, and location, predicts comparable response ranges for our shape and texture stimuli. Results were

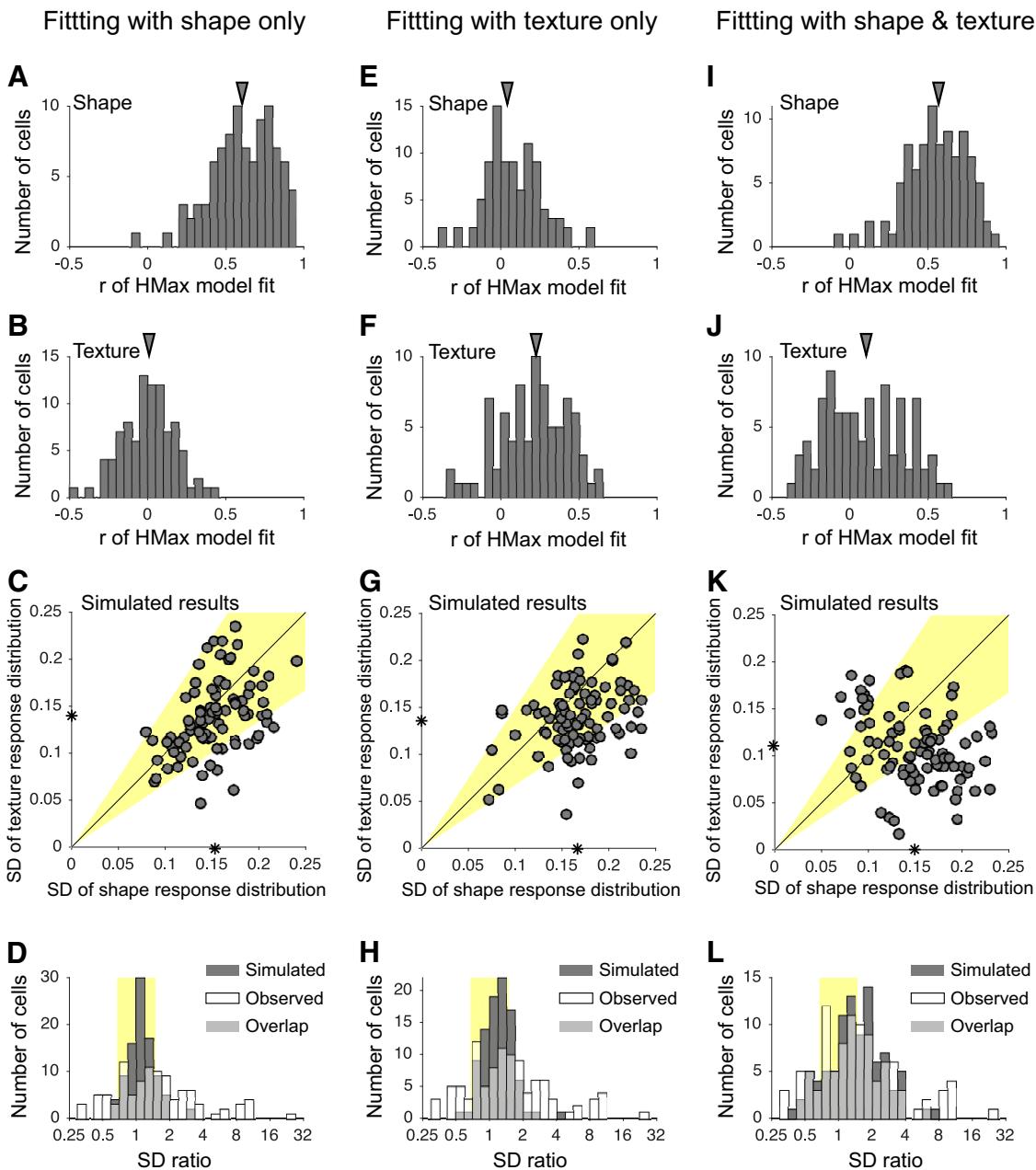


Figure 4. SD ratios from HMax model prediction. **A–D**, Population results for HMax models optimized based on shape responses only. Model goodness of fit for shape (**A**) and texture (**B**) responses across all neurons. Goodness of fit was determined as the median correlation coefficient (r) of 10-fold cross-validation test sets. Triangles represent median. HMax models provided a good fit for shape responses (median $r = 0.61$) but a poor fit for the texture responses (median $r = 0.02$). Predicted response ranges (SD values) for shape and texture stimuli were similar (**C**), and the SD ratios (gray bars in **D**) spanned a narrow range. SD ratios from the observed data (as in Fig. 2D) are overlaid in white (**D**) for comparison. **D**, Light gray bars represent overlap between gray and white distributions. White bars are all the same in **D, H, L**. Yellow shaded area as in Figure 2. **E–H**, HMax model results optimized based on texture responses only. HMax models provided a poor fit for both shape (median $r = 0.07$) and texture (median $r = 0.23$) responses. Model results in terms of SD values and ratios (**G, H**) were similar to those in **C, D, I–L**, HMax models optimized simultaneously based on shape and texture responses. Models provided a good fit for shape responses (median $r = 0.56$) but a poor fit for texture responses (median $r = 0.10$). In this case, the distribution of SD ratios were similar to V4 data (compare gray and white bars in **L**), but the SDs for predicted texture responses were unlike observed data: note the lack of low (<0.05) and high (>0.2) SDs for texture in **K** compared with Figure 2C. Asterisk indicate mean value.

similar when we optimized HMax models with texture data instead: fits to texture data were marginally better than before (median $r = 0.23$; Fig. 4F) and those to shape data were worse (median $r = 0.07$; Fig. 4E), but the range of responses were similar for shape and texture stimuli (Fig. 4G,H), as in Figure 4C. Finally, when we optimized HMax models based on both shape and texture responses simultaneously (Fig. 4I–L), the median goodness of fit was substantially lower for textures (median $r = 0.24$ training set, 0.10 test set) than for shapes ($r = 0.71$ training

set, 0.56 test set), but the resulting best-fit models began to capture the dissociation between the range of responses for shape and texture evident in our V4 data (Fig. 4K,L).

It is possible that the inability of the HMax model to provide a good fit for V4 responses to texture stimuli was due to the limited SF bandwidth of the original HMax model or the fitting constraints we imposed in terms of the fixed number of subunits to model data. To consider these possibilities further, we first tested whether upsampling or downsampling the texture stimuli would

improve fits. The HMax model has a total input field of 180×180 pixels, and the basic tile sizes at the S1 (Gabor) layer range from 32 to 60 pixel square fields (Cadieu et al., 2007). Results were very similar when the texture stimuli were presented within a 128×128 region corresponding to the V4 RF (as in the simulations above) or downsampled to 64×64 pixels or upsampled to 256×256 pixels (with truncation beyond the HMax input field). For all tested neurons ($N = 101$), median r values on the test set for texture responses were 0.28, 0.23, and 0.15 for the smallest to largest patch size, respectively. For comparison, that for the shape data was 0.56, 0.61, and 0.52, respectively. In other words, scaling the SF content of texture patches (or equivalently scaling the SF range of the model) did not produce better fits. Fits to texture data were not improved when we included smaller spatial scale (e.g., S1 units of 20 pixel square field) to increase the HMax model's SF bandwidth. Here again, the median r across the test set was 0.19 for all tested neurons ($N = 101$). It is possible that a greater diversity of V1 basis filters (e.g., Victor et al., 2006) would improve the performance of the HMax model, but this was not assessed. Last, rather than restricting the fits to 13 subunits, we optimized the number of subunits for each neuron by cross-validation. Specifically, for each neuron, we identified the number of subunits beyond which adding subunits failed to improve performance on the test set. In this case, across all tested neurons ($N = 101$), the mean number of subunits was 11, and the median r of HMax texture fit was slightly higher (0.31 vs 0.23).

In summary, these results indicate that a model that successfully explains V4 shape responses does not also do a good job of predicting texture responses, and vice versa. This implies that the broad range of SD ratios observed in V4 are not a trivial consequence of our stimulus choice and that shape and texture selectivity may arise from separate mechanisms (see Discussion). Second, while the HMax model provides a reasonable fit for shape responses in many neurons, it does a poor job of capturing texture selectivity in V4 suggesting that a straightforward combination of orientations and SFs, cannot explain both the shape and texture responses of V4 neurons. Next, we quantify tuning for shape and texture stimuli in terms of tuning for boundary curvature and the perceptual dimensions of texture.

Tuning for boundary curvature and perceptual dimensions of texture

We found that many of the neurons that exhibited a broad range of responses to shape stimuli (those in the bottom right of the scatter in Fig. 2C) also exhibited strong tuning for boundary curvature. Figure 5 shows two such example neurons. For Neuron 2, the preferred shapes (those eliciting the strongest responses; Fig. 5A, top), all had a broad concavity to the top of the shape. In contrast, the nonpreferred shapes (those eliciting the lowest responses; Fig. 5A, bottom) lacked such a feature and often had a sharp point or convex curvature at the top of the shape. To quantify such curvature preference, we fit the angular position and curvature (APC) model (Pasupathy and Connor, 2001) (see Materials and Methods), the parameters of which indicated a preference for a concavity (curvature = -0.27) pointing at 91.7° counterclockwise from rightward (thus, upward; Fig. 5, for details of fit, see legend) and produced predicted responses that were strongly correlated with the measured responses ($r = 0.74$, $p < 0.001$; Fig. 5C). Figure 5D–F shows results from an example neuron from the second monkey. This neuron responded preferentially to shapes with a concave curvature to the top-right of the shape. The curvature preference was well described by the

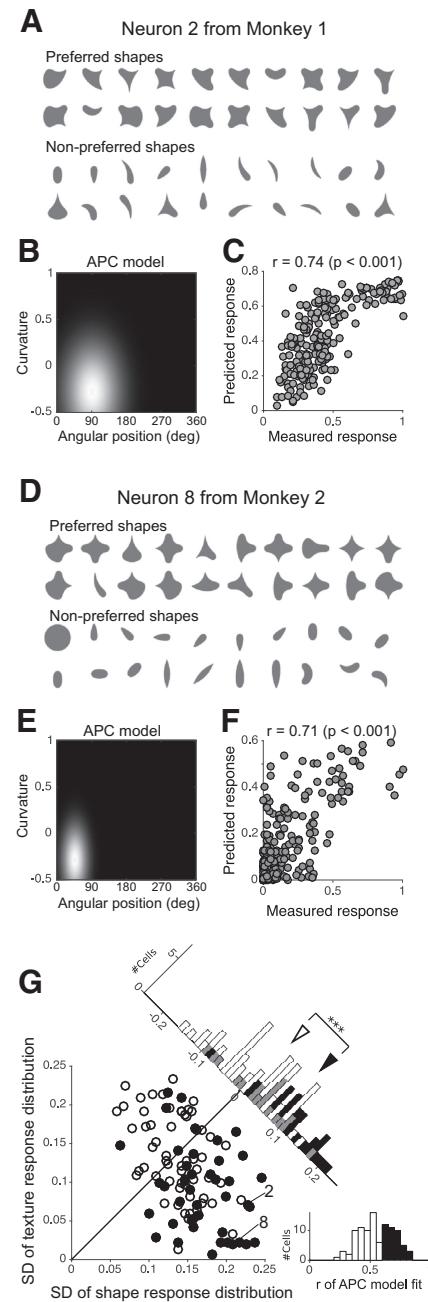


Figure 5. Tuning for boundary curvature in shape-selective neurons. **A**, Shape stimuli that evoked the strongest (preferred) and weakest (nonpreferred) responses from an example neuron (#2; also in Fig. 2). For this neuron, shapes evoked a broader range of responses than textures: SD for shape = 0.23; SD for texture = 0.07. **B**, Responses to shape stimuli were best explained by a 2D Gaussian APC model with a peak at a curvature of -0.27 at 90° , reflecting the preference for concave curvature to the top of the shape. **C**, Responses predicted by the best fit APC model (y axis) are well correlated with the observed responses. **D–F**, Results from a second example neuron (#8). The same conventions as in **A–C**. This neuron responded strongly to shapes with a concave contour at top right of the shape (45°). SD for shape = 0.22; SD for texture = 0.02. **G**, Neurons whose responses are well predicted by the APC model (filled symbols, APC model goodness of fit > 0.6) are identified on a scatter plot of response range for shape and texture stimuli (same as Fig. 2C). This included 42 of 101 neurons across our dataset (right, histogram). Top right, Histogram represents the distribution of SD differences (shape – texture) for highly shape-selective (black bars) and other neurons (white). Mean SD shape minus SD texture was significantly different for the two groups of neurons (Mann–Whitney U test, $p < 0.001$). Triangles represent median values of the distributions. Gray bars represent overlap between two distributions. Data points corresponding to the example neurons in **A–C** (#2), and **D–F** (#8) are identified. Asterisks indicate that difference between two distributions are statistically significant at the level of $p < 0.001$ (Mann–Whitney U -test).

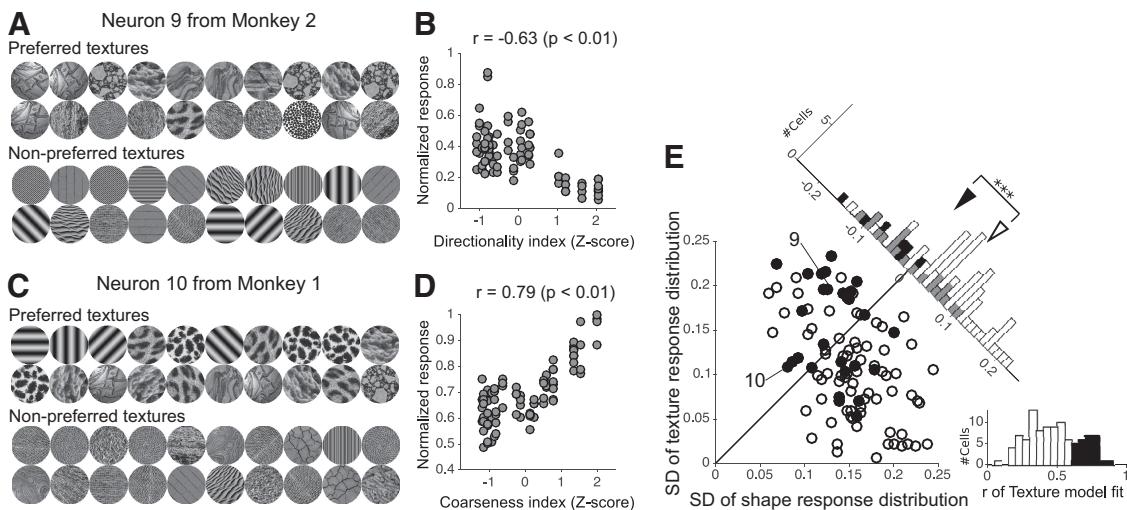


Figure 6. Tuning for perceptual dimension of texture. **A**, Texture stimuli that evoked the strongest (preferred) and weakest (nonpreferred) responses from an example neuron (#9). The nonpreferred textures are directional, oriented at different directions, unlike the preferred stimuli, which tend to be nondirectional for this neuron. SD for shape = 0.13; SD for texture = 0.20. **B**, Neural responses for all texture stimuli (*y* axis) plotted as a function of the directionality index (*x* axis) shows a statistically significant ($p < 0.001$) negative correlation. **C, D**, Example neuron (#10) that responded strongly to coarse rather than fine textures. SD for shape = 0.08; SD for texture = 0.11. **E**, Neurons whose responses are well predicted by the texture model (see Materials and Methods; filled symbols, texture model goodness of fit > 0.6) are identified on the scatter plot of response range for shape and texture stimuli. This included 27 of 101 neurons across our dataset (right, histogram). These texture-selective neurons (filled circles) and the other neurons (open circles) showed a significant difference in distribution of shape SD minus texture SD (Mann–Whitney *U* test, $p < 0.001$; see top right, histogram). Triangles represent median values of the distributions. There was limited overlap ($n = 6$) between neurons with APC model goodness of fit > 0.6 and those with texture model goodness of fit > 0.6 (compare filled symbols in Fig. 5G vs Fig. 6E). Data points corresponding to the example neurons in **A** and **B** (#9), and **C** and **D** (#10) are identified.

best-fit APC model (Fig. 5E), and there was a strong correlation between observed and predicted responses ($r = 0.71$; $p < 0.001$; Fig. 5F).

In Figure 5G, we highlight those neurons that are highly shape-selective (best-fit by APC model, $r > 0.6$, 42 of 101; filled circles) on the continuum of the response range (SD) scatter plot. Highly shape-selective neurons tended to be overrepresented in the lower half where SD for shape $>$ SD for texture: there was a significant and substantial difference (Mann–Whitney *U* test, $p < 0.001$) between highly shape-selective (filled circles) and other neurons (open circles) in terms of their shape SD minus texture SD (Fig. 5G, histogram on diagonal). By demonstrating that the responses of a subpopulation of V4 neurons with a larger dynamic range for shapes versus textures can be explained by tuning for boundary features, we give credence to the idea that neurons with a larger dynamic range for shapes versus textures do have a greater tendency to be involved in coding an aspect of boundary form.

In contrast to the shape-selective neurons discussed above, other neurons that exhibited a larger range of responses to texture stimuli were tuned along one or more of the perceptual dimensions of texture that we varied. For example, the preferred textures of Neuron 9 (Fig. 6A, top) all lacked parallel-oriented elements, whereas the nonpreferred textures (Fig. 6A, bottom) all included them (regardless of orientations), suggesting a selectivity for textures that score low on the axis of directionality. Indeed, there was a strong negative correlation between directionality and neuronal response ($r = -0.63$, $p < 0.001$; Fig. 6B). Figure 6C, D shows results from a neuron from the second monkey that had a strong preference for coarse textures.

We determined which neurons were quantitatively well fit by a texture model using linear regression along the perceptual dimensions of directionality, coarseness, regularity, and contrast (see Materials and Methods). In Figure 6E, we identify those texture-selective neurons (filled symbols) among all neurons on

the axes of response range (SD) for the shape and texture stimulus sets. Neurons with responses that were well predicted by the texture model ($r > 0.6$, filled symbols, 27 of 101), tended to have a larger dynamic range for texture stimuli than did other neurons (open symbols, $r < 0.6$): there was a statistically significant difference in shape SD minus texture SD between these two groups of neurons (Mann–Whitney *U* test, $p < 0.001$; histogram on diagonal). There was minimal overlap between neurons tuned to texture dimensions and those tuned to curvature (Fig. 5G vs 6E; 42 vs 27; 6 neurons overlapped). This is consistent with the idea that many individual neurons may be specialized to encode either shape or texture. In Figure 6E, texture selectivity was evaluated based on neural responses to large-aperture textures, but we verified that results for large and small apertures were similar: weights for each independent variable (i.e., perceptual texture dimension) showed strong and statistically significant positive correlation (*r* values for coarseness, directionality, regularity, and contrast were 0.65, 0.68, 0.70, and 0.69, respectively).

Separable tuning for shape and texture

To determine whether shape tuning was consistent across different surface textures, for each neuron, we chose 2D shapes that evoked a strong, a moderate, and a weak response. We then presented the 10 nondirectional texture stimuli (Fig. 1B, bottom two rows) within each of these three shape apertures. If neuronal selectivity for shape and texture information are largely independent, texture tuning should be similar regardless of the shape aperture, and shape preference should be preserved regardless of texture. Figure 7 shows the responses of six neurons to the shape–texture combination stimuli. Response patterns are quite different across these neurons. In some cases (Fig. 7A), the response modulation across the three shapes was stronger than modulation across the 10 textures. In other cases, texture had the larger influence on the neuronal responses (Fig. 7B). And in still others, both shape and texture modulate responses (Fig. 7C–F). For 43

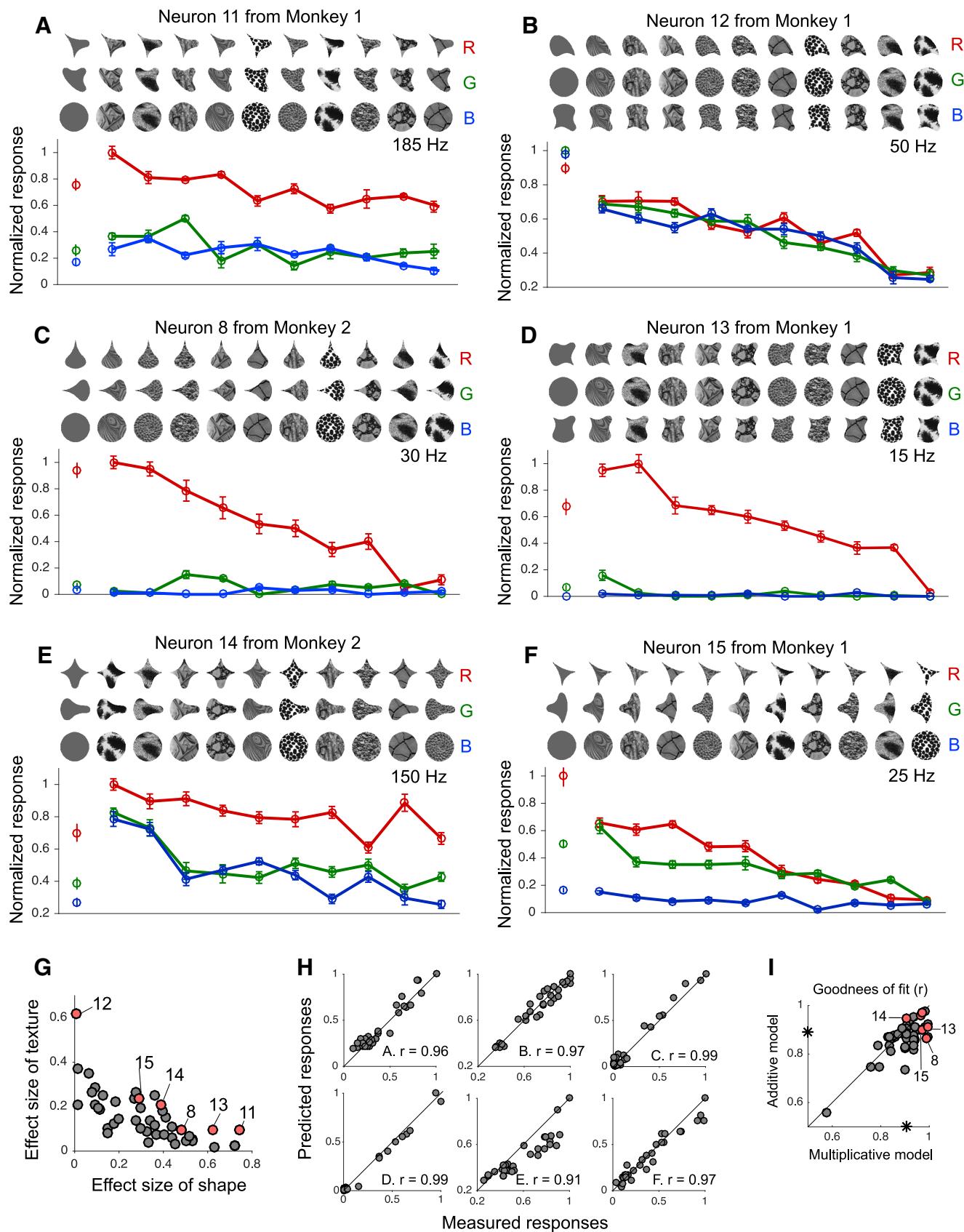


Figure 7. Joint coding of shape and texture. **A**, Responses of an example neuron (#11) to 10 nondirectional textures (x axis) presented through three different shapes apertures (line colors). Responses to the three shapes presented with a uniform luminance contrast are also shown (leftmost symbols) for comparison. Error bars indicate ± 1 SEM. This neuron exhibited a broader range of shape responses than texture responses (SD for shape = 0.19, SD for texture = 0.10), but overall, shape preference was largely preserved across textures. **B**, Example neuron (#12) with a strong preference for texture but not shape (SD for shape = 0.07, SD for texture = 0.16). All details as in **A**. **C**, Neuron 8 showed selectivity along shape and texture (Figure legend continues).

neurons in which the control experiment was conducted, we calculated the effect size (see Materials and Methods) of shape and texture variables, respectively (Fig. 7G). Neurons tuned along both dimensions have a large effect size along both dimensions. In all cases, tuning for shape and texture tuning were largely independent. To quantify the independence of shape and texture tuning, we evaluated whether responses to shape-texture combination stimuli can be predicted by the product of the tuning for shape and texture (see Materials and Methods). The correlation coefficients, r between model and data, are shown for the six example neurons (Fig. 7H) and the mean r was 0.90 ± 0.08 ($N = 43$; Fig. 7I, x axis), indicating that the multiplicative model captures $\sim 81\%$ of the variance in the data. In most cells, the multiplicative model (Fig. 7I, x axis) gave a better prediction than an additive model did (Fig. 7I, y axis) but either would imply separable tuning.

Different time courses for shape and texture processing

To compare the time course of shape and texture selectivity, for each neuron, we constructed average PSTHs based on the top 50% (preferred) and bottom 50% shape and texture stimuli (see Materials and Methods). For neuron 16 (Fig. 8A), responses to preferred and nonpreferred shapes (left) diverged soon after response onset (~ 50 ms from the stimulus onset), and the difference was sustained throughout the stimulus presentation period. However, for texture stimuli (middle and right panels), preferred and nonpreferred PSTHs diverged only ~ 100 ms from the stimulus onset. This delayed separation was observed for both large and small aperture textures. Figure 8B shows an example neuron from Monkey 2 showing a similar pattern of results.

We found a similar trend across the population. Figure 8C shows a scatter plot of the time of onset of texture selectivity (see Materials and Methods) (y axis) versus shape selectivity (x axis). Across the population, shape selectivity emerged early (mean onset time was 55 ms; SD = 16 ms), whereas texture selectivity was delayed by ~ 30 ms. Mean onset times for texture selectivity were similar regardless of the size of aperture (Fig. 8C,D): 84 and 85 ms for texture selectivity based on large (filled symbols) and small (open symbols) aperture textures, respectively. In both monkeys, the differences between onset times for shape and texture selectivity were statistically significant (Wilcoxon Signed Rank test, $p < 0.001$). These results indicate that the encoding of boundary shape and surface texture may occur with different temporal dynamics.

It may be argued that the delayed onset of texture processing could be associated with weaker response modulation to textures compared with shapes. To address this concern, we divided our data into three groups depending on the ratios of the SDs for

←

(Figure legend continued.) dimensions. Preference for fine textures was observable only within the preferred shape boundary. **D–F**, Additional example neurons (#13, #14, #15) that exhibited joint tuning for shape and texture. **G**, Effect size (see Materials and Methods) of texture was compared with that of shape for each of 43 neurons subjected to the control experiment. Data points corresponding to the example neurons (**A–F**) are identified. **H**, To quantify the independence of shape and texture tuning, we evaluated whether responses to shape-texture combination stimuli can be predicted by the product of the responses to shape and texture. Scatter plots show observed responses versus those predicted by a multiplicative model (see Materials and Methods) for neurons in **A–F**. **I**, Comparison between multiplicative (x axis) model and additive (y axis) model. Goodness-of-fit (r) values were quantified by the correlation coefficient between observed and predicted responses across all neurons ($n = 43$). Multiplicative model (median $r = 0.91$) generally provides a better fit than an additive model (median $r = 0.86$). Asterisk indicates median value.

shape versus texture responses (Fig. 2D), and then compared mean latency for shape and texture processing in each group of neurons: Group 1 ($N = 15$; SD ratio < 0.66), Group 2 ($N = 42$; $0.66 < \text{SD ratio} < 1.5$; Fig. 2D, yellow shaded area), Group 3 ($N = 44$; SD ratio > 1.5). Shape selectivity emerged earlier than texture selectivity for all three groups. The difference was the largest for Group 3 (54 ms for shape, 96 ms for large aperture texture, 92.69 ms for small aperture texture), which included neurons that exhibit a greater range of responses for shape. But delayed onset of texture selectivity was also observed for Group 2 (55 ms for shape, 82 ms for large aperture texture, 85 ms for small aperture texture) and Group 1 (59 ms for shape, 77 ms for large aperture texture, 79 ms for small aperture texture), which included neurons more selective for texture. Therefore, it is safe to conclude that texture information processing is delayed than shape information processing in area V4.

Discussion

We used systematically designed stimuli to compare the responses of V4 neurons to shapes and textures. Our results reveal four novel and potentially fundamental properties of form-texture encoding in visual cortex. First, individual V4 neurons exhibit joint, separable tuning for shape and texture. Second, neurons span a continuum from strongly shape-selective to strongly texture-selective, but overall V4 responses were more strongly modulated by object boundary features rather than by texture. Third, many V4 neurons were highly selective along the texture dimensions of coarseness, regularity, and directionality thought to be important for texture perception in human subjects. Finally, shape and texture information processing followed different temporal dynamics: texture selectivity emerged significantly later than shape. These results argue for an important role for area V4 in the emergence of object-based structural codes from surface characteristics-based representations in earlier stages of the ventral pathway.

Encoding “things” and “stuff” in area V4

Previous studies in V4 demonstrated selectivity for object features (Kobatake and Tanaka, 1994; Pasupathy and Connor, 1999) and for texture (Arcizet et al., 2008; Okazawa et al., 2015). But because most studies focus either on the encoding of “things” or “stuff,” we do not know whether different subgroups of V4 neurons are selective for shape and texture, or whether the same neurons carry information about both. Specifically, Ziembka and Freeman (2015) have argued theoretically that tuning for texture in terms of higher-order image statistics could produce shape selectivity as a byproduct. But our results, which document joint, separable V4 tuning for boundary shape and texture, argue against this possibility. Because image statistics would change substantially depending on the texture painted on the surface of a shape, any shape preference that is based on tuning for higher-order image statistics would be highly dependent on surface texture attributes of the stimulus. Therefore, the texture-invariant shape tuning that we document here cannot be based on tuning for higher-order image statistics. Because textures may be envisioned as being composed of small shape elements (Julesz, 1981; Galun et al., 2003), tuning for shape versus texture may be based on preference for scale: neurons that prefer small shape elements may exhibit texture tuning while neurons that prefer large stimuli may exhibit shape tuning. However, because many neurons in our dataset exhibit joint tuning at multiple scales, for example, simultaneous selectivity for shape and for fine (as opposed to

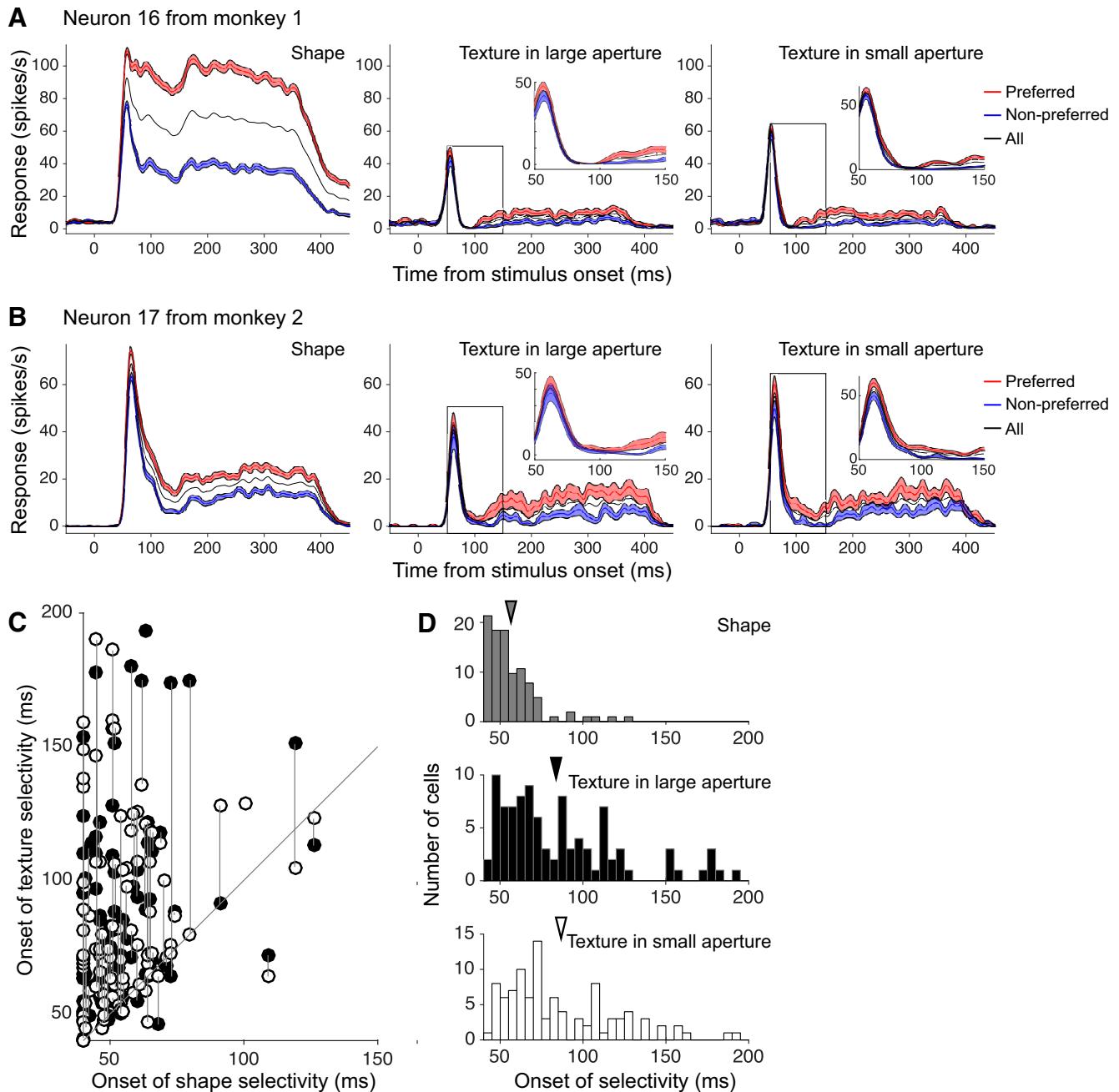


Figure 8. Temporal dynamics of shape and texture selectivity. **A**, PSTHs for shape (left), large and small aperture textures (middle and right, respectively) are shown for preferred (red; top 50% of stimuli based on spike counts between 50 and 400 ms after stimulus onset; see Materials and Methods), nonpreferred (blue; bottom 50%), and all stimuli (black) for one example neuron. Shaded area represents ± 1 SEM. For shape stimuli, difference in responses between preferred and nonpreferred stimuli emerged soon (50 ms) after response onset. For textures, statistically significant difference emerged at 100 ms after stimulus onset. **B**, Second example neuron showing delayed emergence of texture selectivity (shape-dependent modulation ≥ 51 ms; texture-dependent modulation ≥ 88 ms; earlier onset was determined for small aperture texture condition). **C**, Across all neurons, onset of shape selectivity (*x* axis) is plotted against onset of texture selectivity (*y* axis). Filled and open symbols represent large and small aperture conditions, respectively. Data points from the same neuron are connected by a vertical line. In a few neurons (data points without vertical line), onset of texture selectivity could not be defined for one of the aperture conditions due to weak responses. Most data points lie above the diagonal line, indicating that texture information is processed later than shape information. **D**, Marginal histograms for onset times for shape (gray), large aperture texture (black), and small aperture texture (white). Triangles represent the mean of each distribution (shape = 55.72 ms, large aperture texture = 85.53 ms, small aperture texture = 85.78 ms). On average, onset of shape selectivity was ~ 30 ms faster than onset of texture selectivity.

coarse) textures, stimulus scale alone cannot explain V4 encoding of shapes and textures.

Instead, our results support the hypothesis that the responses of individual V4 neurons are informed by two largely separate and independent computations that inform shape and texture selectivity, respectively. Recent studies suggest that texture selectivity may be based on computing high-order image statistics

from the visual image (Freeman et al., 2013; Okazawa et al., 2015), whereas shape selectivity may be based on the structure of larger scale contrast boundaries within the RF (Popovkina et al., 2019). Modeling of neural activity with deep convolutional neural networks (CNNs) may provide additional clues into how the brain builds a complex object recognition system from simple earlier representations (Cadieu et al., 2014; Yamins et al., 2014).

For example, a recent study that probed a CNN with shape stimuli like those used to study V4 neurons reported that units in the middle layers of the CNN provide the best-known image-computable model for V4-like translation-invariant boundary curvature selectivity (Pospisil et al., 2018). Another study (Geirhos et al., 2018) reported that, unlike human observers, ImageNet-trained CNNs tend to classify objects according to local textures rather than shapes, and this texture bias can be overcome and changed toward a shape bias if trained on a suitable dataset. Future work should seek to understand whether and how shape and texture information is jointly encoded in CNNs.

Our results suggest that V4 includes a form-texture continuum between two conceptually distinct endpoints, similar to that in V1 for simple and complex cells (Hubel and Wiesel, 1968), or in MT for pattern and component cells (Movshon et al., 1985). In V4, “form cells” exhibit strong response modulation for shapes but weak modulation for texture, whereas “texture cells” show strong tuning for texture but weak tuning for shape. As in analogous continua in V1 and MT, most neurons lie in between and exhibit moderate shape and texture tuning. Overall, our results support the idea that partially overlapping subsets of V4 neurons contribute to the encoding of shape and texture in visual scenes.

Emergence of object representation in V4

One of the major challenges in natural vision is the segmentation of objects from surround texture, a process critical for successful object recognition (Thielscher and Neumann, 2003; Grigorescu et al., 2004). Past studies have argued that mechanisms of contextual modulation could facilitate object segmentation. For example, the influence of surround suppression on neuronal responses may be stronger for images that provide a homogeneous stimulation of the RF and surround (Coen-Cagli et al., 2015). In particular, iso-orientation surround suppression could suppress the encoding of uniform texture (Grigorescu et al., 2003; Wei et al., 2013; Schmid and Victor, 2014) and enhance the representation of object boundaries. Our discovery of stronger responses to shapes in V4 is consistent with this process called detexturization (Gheorghiu et al., 2014) and with psychophysical studies that argue for a primary role for boundary information in object recognition (Biederman and Ju, 1988; Davidoff and Ostergaard, 1988; Elder and Velisavljevic, 2009; Fu et al., 2016). Because neurons in V1 and V2 primarily encode surface characteristics (but see border-ownership coding in V2) (Zhou et al., 2000), a preference for encoding objects in V4 could reflect its fundamental role in the computation of object-based representations in the ventral stream. Future studies with more diverse stimulus sets will be required to determine whether our results hold up for more realistic objects where form and texture are rendered with 3D realism.

V4 selectivity for perceptual dimensions of texture

Several recent physiological studies have documented selectivity for naturalistic texture in V2 and V4, and have described such selectivity on the basis of higher-order image statistics (Freeman et al., 2013; Okazawa et al., 2015). Consistent with these previous studies, we too find selectivity for texture in V4 neurons. But because we quantified our texture stimuli in terms of regularity, coarseness, and directionality (dimensions critical for human texture perception), we provide the first documentation of V4 selectivity for perceptual dimensions of texture. These results are consistent with results from lesion studies in V4 demonstrating impaired texture segregation (Merigan, 2000; Allen et al., 2009),

and they support a prominent role for V4 in the perception of textures.

However, our texture dimensions also had several limitations. First, we matched mean luminance across textures, but not local contrast (see Materials and Methods). So, further studies are needed to determine whether texture selectivity depends on contrast. Second, we did not consider roughness (or gloss) as a texture dimension but previous studies in V4 (Arcizet et al., 2008) and IT cortex (Nishio et al., 2012) have reported selectivity for this stimulus attribute. Our dataset included neurons whose responses were well modulated by our texture stimuli, but not well described by the linear texture model used here (Fig. 6E, open circles above the diagonal line). So additional texture dimensions may need to be considered. Finally, for the texture stimuli, regularity was not entirely orthogonal to directionality and coarseness (strong directionality or low coarseness (fineness) was often correlated with regular texture), but it is not known whether this is a general relationship in natural textures.

Delayed emergence of texture selectivity

Previous psychophysical studies have argued that rapid natural scene categorization is primarily mediated by edge-based representations because surface information takes longer to influence categorization performance (Elder and Zucker, 1998; Fu et al., 2016). In the rodent somatosensory cortex, Isett et al. (2018) recently found that local geometry (shape) information was processed by instantaneous firing, whereas surface texture (roughness vs smoothness) was processed by a slower rate code. Our demonstration of delayed onset of texture selectivity in V4 is consistent with these previous studies.

Models of coarse-to-fine processing postulate that the visual system first processes low-SF content of the image that carries the “gist” of the scene and that higher SF content, which provides spatial detail, is processed more slowly (Oliva, 2005; Allen and Freeman, 2006; Hegdé, 2008). Other models propose that scene segmentation may be initiated from the detection of boundaries, then followed by filling-in between the edges (Lamme et al., 1999; Grossberg, 2003; Huang and Paradiso, 2008; Poort et al., 2012). Recent studies hypothesize that texture selectivity in V2 and V4 may be based on computing correlations in activity among neighboring neurons (Okazawa et al., 2015; Ziomba et al., 2016), which may depend on lateral cortical connections known to be slower than feedforward connections (Grinvald et al., 1994; Kim and Freeman, 2014).

In conclusion, our results begin to unravel how shape and texture information are multiplexed in individual V4 neurons across time, and across the V4 population, to underlie the perception of objects and surfaces.

References

- Adelson EH (2001) On seeing stuff: the perception of materials by humans and machines. *Proc SPIE* 4299:1–12.
- Allen EA, Freeman RD (2006) Dynamic spatial processing originates in early visual pathways. *J Neurosci* 26:11763–11774.
- Allen HA, Humphreys GW, Colin J, Neumann H (2009) Ventral extrastriate cortical areas are required for human visual texture segmentation. *J Vis* 9:2.
- Arcizet F, Jouffrais C, Girard P (2008) Natural textures classification in area V4 of the macaque monkey. *Exp Brain Res* 189:109–120.
- Biederman I, Ju G (1988) Surface versus edge-based determinants of visual recognition. *Cogn Psychol* 20:38–64.
- Brodatz P (1966) *Textures: a photographic album for artists and designers*. Minneola, NY: Dover.
- Bushnell BN, Pasupathy A (2012) Shape encoding consistency across colors in primate V4. *J Neurophysiol* 108:1299–1308.

- Bushnell BN, Harding PJ, Kosai Y, Bair W, Pasupathy A (2011) Equilumiance cells in visual cortical area V4. *J Neurosci* 31:12398–12412.
- Cadieu CF, Hong H, Yamins DL, Pinto N, Ardila D, Solomon EA, Majaj NJ, DiCarlo JJ (2014) Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput Biol* 10: e1003963.
- Cadieu C, Koth M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. *J Neurophysiol* 98:1733–1750.
- Coen-Cagli R, Kohn A, Schwartz O (2015) Flexible gating of contextual influences in natural vision. *Nat Neurosci* 18:1648–1655.
- Cohen J (1973) Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educ Psychol Meas* 33:107–112.
- Conway BR, Tsao DY (2006) Color architecture in alert macaque cortex revealed by fMRI. *Cereb Cortex* 16:1604–1613.
- Conway BR, Moeller S, Tsao DY (2007) Specialized color modules in macaque extrastriate cortex. *Neuron* 56:560–573.
- Davidoff JB, Ostergaard AL (1988) The role of colour in categorial judgements. *Q J Exp Psychol A* 40:533–544.
- Desimone R, Schein SJ (1987) Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol* 57:835–868.
- Elder JH, Velisavljevic L (2009) Cue dynamics underlying rapid detection of animals in natural scenes. *J Vis* 9:7.
- Elder JH, Zucker SW (1998) Evidence for boundary-specific grouping. *Vision Res* 38:143–152.
- El-Shamayleh Y, Pasupathy A (2016) Contour curvature as an invariant code for objects in visual area V4. *J Neurosci* 36:5532–5543.
- Freeman J, Ziomba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. *Nat Neurosci* 16:974–981.
- Fu Q, Liu YJ, Dienes Z, Wu J, Chen W, Fu X (2016) The role of edge-based and surface-based information in natural scene categorization: evidence from behavior and event-related potentials. *Conscious Cogn* 43:152–166.
- Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science* 259:100–103.
- Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC (1996) Neural responses to polar, hyperbolic, and cartesian grating in area V4 of the macaque monkey. *J Neurophysiol* 76:2718–2739.
- Galun M, Sharon E, Basri R, Brandt A (2003) Texture segmentation by multiscale aggregation of filter responses and shape elements. In: *Proceedings Ninth IEEE International Conference on Computer Vision*, Vol 1, pp 716–723. IEEE.
- Gattass R, Sousa AP, Gross CG (1988) Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8:1831–1845.
- Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, Brendel W (2018) ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv: 1811.12231*.
- Gheorghiu E, Kingdom FA, Petkov N (2014) Contextual modulation as de-texturizer. *Vision Res* 104:12–23.
- Grigorescu C, Petkov N, Westenberg MA (2003) Contour detection based on nonclassical receptive field inhibition. *IEEE Trans Image Process* 12: 729–739.
- Grigorescu C, Petkov N, Westenberg MA (2004) Contour and boundary detection improved by surround suppression of texture edges. *Image Vis Comput* 22:609–622.
- Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *J Neurosci* 14:2545–2568.
- Grossberg S (2003) Filling-In the Forms: Surface and Boundary Interactions in Visual Cortex. In: *Filling-In*, pp 13–37. New York, NY: Oxford University Press.
- Hayes KC, Shah AN, Rosenfeld A (1974) Texture coarseness: further experiments. *IEEE Trans Syst Man Cybern SMC-4*, pp 467–472.
- Hegdé J (2008) Time course of visual perception: coarse-to-fine processing and beyond. *Prog Neurobiol* 84:405–439.
- Heywood CA, Gadotti A, Cowey A (1992) Cortical area V4 and its role in the perception of color. *J Neurosci* 12:4056–4065.
- Huang X, Paradiso MA (2008) V1 response timing and surface filling-in. *J Neurophysiol* 100:539–547.
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215–243.
- Isett BR, Feasel SH, Lane MA, Feldman DE (2018) Slip-based coding of local shape and texture in mouse S1. *Neuron* 97:418–433.e5.
- Johnson EN, Hawken MJ, Shapley R (2001) The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat Neurosci* 4:409–416.
- Johnson EN, Hawken MJ, Shapley R (2008) The orientation selectivity of color-responsive neurons in macaque V1. *J Neurosci* 28:8096–8106.
- Julesz B (1981) Textons, the elements of texture perception, and their interactions. *Nature* 290:91–97.
- Kim T, Freeman RD (2014) Selective stimulation of neurons in visual cortex enables segregation of slow and fast connections. *Neuroscience* 274: 170–186.
- Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856–867.
- Lamme VA, Rodriguez-Rodriguez V, Spekreijse H (1999) Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cereb Cortex* 9:406–413.
- Lennie P, Movshon JA (2005) Coding of color and form in the geniculostriate visual pathway (invited review). *J Opt Soc Am A* 22:2013–2033.
- Liu F, Picard RW (1996) Periodicity, directionality, and randomness: wold features for image modeling and retrieval. *IEEE Trans Pattern Anal Mach Intell* 18:722–733.
- Livingstone M, Hubel D (1988) Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240:740–749.
- Merigan WH (2000) Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. *Vis Neurosci* 17: 949–958.
- Movshon JA, Simoncelli EP (2014) Representation of naturalistic image structure in the primate visual cortex. *Cold Spring Harb Symp Quant Biol* 79:115–122.
- Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving visual patterns. *Pontificia Academica Scripta Varia* 54:117–151.
- Namima T, Yasuda M, Banno T, Okazawa G, Komatsu H (2014) Effects of luminance contrast on the color selectivity of neurons in the macaque area V4 and inferior temporal cortex. *J Neurosci* 34:14934–14947.
- Nandy AS, Sharpee TO, Reynolds JH, Mitchell JF (2013) The fine structure of shape tuning in area V4. *Neuron* 78:1102–1115.
- Nishio A, Goda N, Komatsu H (2012) Neural selectivity and representation of gloss in the monkey inferior temporal cortex. *J Neurosci* 32:10780–10793.
- Okazawa G, Tajima S, Komatsu H (2015) Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc Natl Acad Sci U S A* 112:E351–E360.
- Oliva A (2005) Gist of the scene. In: *Encyclopedia of neurobiology of attention* (Itti L, Rees G, Tsotsos JK, eds), pp 251–256. San Diego: Elsevier.
- Pasupathy A, Connor CE (1999) Responses to contour features in macaque area V4. *J Neurophysiol* 82:2490–2502.
- Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86:2505–2519.
- Pasupathy A, Connor CE (2002) Population coding of shape in area V4. *Nat Neurosci* 5:1332–1338.
- Poort J, Raudies F, Wannig A, Lamme VA, Neumann H, Roelfsema PR (2012) The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* 75:143–156.
- Popovkina DV, Bair W, Pasupathy A (2019) Modelling diverse responses to filled and outline shapes in macaque V4. *J Neurophysiol* 121:1059–1077.
- Pospisil DA, Pasupathy A, Bair W (2018) “Artiphysiology” reveals V4-like shape tuning in a deep network trained for image classification. *Elife* 7:e38242.
- Rao AR, Lohse GL (1996) Towards a texture naming system: identifying relevant dimensions of texture. *Vision Res* 36:1649–1669.
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.
- Schein SJ, Desimone R (1990) Spectral properties of V4 neurons in the macaque. *J Neurosci* 10:3369–3389.
- Schmid AM, Victor JD (2014) Possible functions of contextual modulations and receptive field nonlinearities: pop-out and texture segmentation. *Vision Res* 104:57–67.

- Serre T, Khou M, Cadieu C, Knoblich U, Kreiman G, Poggio T (2005) A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. CBCL Paper #259/AI Memo #2005-036. Cambridge, MA: MIT.
- Shapley R, Hawken MJ (2011) Color in the cortex: single- and double-opponent cells. *Vision Res* 51:701–717.
- Tamura H, Mori S, Yamawaki T (1978) Textural features corresponding to visual perception. *IEEE Trans Syst Man Cybern* 8:460–473.
- Thielscher A, Neumann H (2003) Neural mechanisms of cortico-cortical interaction in texture boundary detection: a modeling approach. *Neuroscience* 122:921–939.
- Victor JD, Mechler F, Repucci MA, Purpura KP, Sharpee T (2006) Responses of V1 neurons to two-dimensional hermite functions. *J Neurophysiol* 95:379–400.
- Wei H, Dong Z (2015) V4 neural network model for shape-based feature extraction and object discrimination. *Cognit Comput* 7:753–762.
- Wei H, Lang B, Zuo Q (2013) Contour detection model with multi-scale integration based on non-classical receptive field. *Neurocomputing* 103:247–262.
- Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci U S A* 111:8619–8624.
- Zamarashkina P, Popovkina D, Pasupathy A (2017) Stimulus and task dependence of response latencies in primate area V4. *J Vis* 17:476.
- Zeki SM (1973) Colour coding in rhesus monkey prestriate cortex. *Brain Res* 53:422–427.
- Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. *J Neurosci* 20:6594–6611.
- Ziemba CM, Freeman J (2015) Representing “stuff” in visual cortex. *Proc Natl Acad Sci U S A* 112:942–943.
- Ziemba CM, Freeman J, Movshon JA, Simoncelli EP (2016) Selectivity and tolerance for visual texture in macaque V2. *Proc Natl Acad Sci U S A* 113:E3140–E3149.