J Neurophysiol 121: 1059–1077, 2019.
First published January 30, 2019; doi:10.1152/jn.00456.2018.

RESEARCH ARTICLE | *Sensory Processing*

Modeling diverse responses to filled and outline shapes in macaque V4

Dina V. Popovkina, Wyeth Bair, and Anitha Pasupathy
Department of Biological Structure, Washington National Primate Research Center, University of Washington, Seattle, Washington

Submitted 6 July 2018; accepted in final form 25 January 2019

Popovkina DV, Bair W, Pasupathy A. Modeling diverse responses to filled and outline shapes in macaque V4. *J Neurophysiol* 121: 1059–1077, 2019. First published January 30, 2019; doi: 10.1152/jn.00456.2018.—Visual area V4 is an important midlevel cortical processing stage that subserves object recognition in primates. Studies investigating shape coding in V4 have largely probed neuronal responses with filled shapes, i.e., shapes defined by both a boundary and an interior fill. As a result, we do not know whether form-selective V4 responses are dictated by boundary features alone or if interior fill is also important. We studied 43 V4 neurons in two male macaque monkeys (*Macaca mulatta*) with a set of 362 filled shapes and their corresponding outlines to determine how interior fill modulates neuronal responses in shape-selective neurons. Only a minority of neurons exhibited similar response strength and shape preferences for filled and outline stimuli. A majority responded preferentially to one stimulus category (either filled or outline shapes) and poorly to the other. Our findings are inconsistent with predictions of the hierarchical-max (HMax) V4 model that builds form selectivity from oriented boundary features and takes little account of attributes related to object surface, such as the phase of the boundary edge. We modified the V4 HMax model to include sensitivity to interior fill by either removing phase-pooling or introducing unoriented units at the V1 level; both modifications better explained our data without increasing the number of free parameters. Overall, our results suggest that boundary orientation and interior surface information are both maintained until at least the midlevel visual representation, consistent with the idea that object fill is important for recognition and perception in natural vision.

NEW & NOTEWORTHY The shape of an object's boundary is critical for identification; consistent with this idea, models of object recognition predict that filled and outline versions of a shape are encoded similarly. We report that many neurons in a midlevel visual cortical area respond differently to filled and outline shapes and modify a biologically plausible model to account for our data. Our results suggest that representations of boundary shape and surface fill are interrelated in visual cortex.

computational model; cortical visual processing; form selectivity; macaque visual cortex; ventral pathway

INTRODUCTION

Humans can recognize objects from line drawings as quickly and accurately as from full-color photographs (Biederman and Ju 1988). This remarkable ability to determine object identity in the absence of surface characteristics is established early in life (Hochberg and Brooks 1962; Yonas and Arterberry 1994), has been exploited since the earliest cave paintings, and has also been observed in monkeys (Truppa et al. 2009; Zimmerman and Hochberg 1970). Invariant recognition of filled and outline forms could arise from similar representations for both types of stimuli (Biederman and Ju 1988; Sayim and Cavanagh 2011), but this would be inconsistent with the demonstration that object recognition in natural scenes becomes difficult if the available information is restricted to edges without surface features (Fig. 1; see also Wagemans et al. 2008). To begin to understand the neural representations that support these seemingly disparate perceptual observations, here we focus on cortical area V4, an important stage in the processing of visual form in primates (De Weerd et al. 1996; Gallant et al. 2000; Heywood and Cowey 1987; Merigan 1996; Merigan and Pham 1998; Schiller 1995; Schiller and Lee 1991), and ask whether filled shapes and corresponding outlines are similarly encoded in V4 responses.

Past studies have demonstrated that many V4 neurons are responsive to the shape of the boundaries of simple visual stimuli (Kobatake and Tanaka 1994; Pasupathy and Connor 1999, 2001). Such tuning for boundary conformation has been demonstrated largely with stimuli defined by both an outline and an interior fill. If the responses of shape-selective V4 neurons were in fact dictated primarily by the shape of the boundary, then a circle and a disk should be encoded similarly. For example, one might expect a shape-selective neuron to maintain the same preferences for filled shapes and for shapes presented as outlines (Fig. 2, *A* and *B*, respectively). Interestingly, this is the prediction of models of form processing that use phase-invariant, oriented units to build shape templates. In particular, the hierarchical-max (HMax) model of Cadieu et al. (Cadieu et al. 2007; Serre et al. 2005) relies heavily on phase-invariant oriented channels, which roughly approximate the selectivity of complex cells in macaque primary visual cortex (V1) (De Valois et al. 1982; Foster et al. 1985), and should thus respond similarly to filled and outline shapes.

To determine whether neurons in V4 maintain their shape selectivity to filled and outline stimuli, and to understand whether their responses are consistent with prevailing models of shape selectivity, we studied the responses of V4 neurons in two fixating macaque monkeys to a set of shape stimuli used previously to characterize V4 boundary curvature selectivity (Pasupathy and Connor 2001). We found that few neurons responded similarly to filled and outline stimuli and the ma-

Address for reprint requests and other correspondence: D. V. Popovkina, Univ. of Washington, 1959 NE Pacific St., Box 357420, HSB G-520, Seattle, WA 98195 (e-mail: dina4@uw.edu).

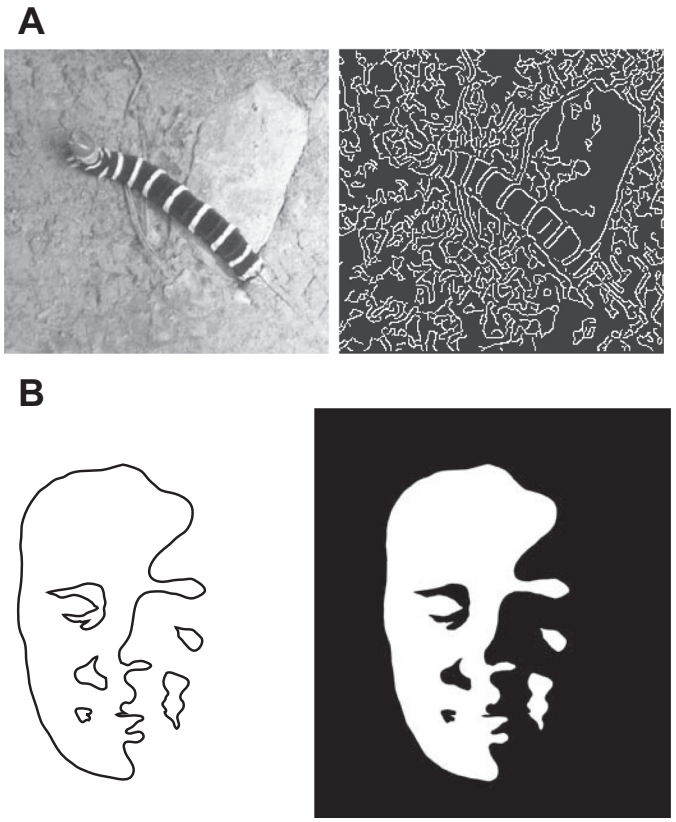


Fig. 1. Importance of surface features in recognition. *A*: an example natural image (*left*) and the component edges extracted with a Canny edge detector (*right*) illustrate the difficulty in parsing and recognition of a scene based on boundary information alone. *B*: the Mooney face percept is stronger when elements have surface fill (*right*), compared with the boundary edges alone (*left*). Inspired by Cavanagh (1991). This figure was originally published in Pasupathy et al. (2018) and has been reproduced by permission of Oxford University Press.

jority responded differentially to the two classes of stimuli, demonstrating that V4 responses are influenced by both boundary features and interior fill. We show that our results are inconsistent with the predictions of the HMax model. To better explain how V4 responses arise, we tested two strategies for retaining surface information within the model: eliminating phase-pooling and including unoriented channels that conveyed sensitivity to surface fill. Both modifications produced model V4 units that successfully captured our observed data.

METHODS

Surgical Methods

We implanted two adult male macaque monkeys (*Macaca mulatta*) with custom head posts and recording chambers positioned over dorsal area V4 (left hemisphere). The placement of the chamber over the prelunate gyrus and subsequent craniotomy were guided by structural MRI (for details, see Bushnell et al. 2011a). All animal procedures conformed to National Institutes of Health guidelines and were reviewed and approved by the Institutional Animal Care and Use Committee at the University of Washington.

Visual Stimulus Presentation

Stimuli were presented against a uniform gray background (luminance 5.4 cd/m^2) with a spectrally calibrated (PR650; PhotoResearch) CRT monitor with pixel resolution 1,600 × 1,200 and screen size 16.1 × 11.8 in., positioned 45.4 or 56 cm away (*monkeys M1* and *M2*, respectively). The animals fixated a small white dot in the center of the screen within a window of radius 0.75–1°. Eye position was monitored with an infrared eye tracking system (EyeLink 1000; SR Research) and coordinated with stimulus presentation using custom software based on Pype (Mazer 2013). Animals were rewarded with drops of juice for successful fixation for the entire duration of a trial (typically 2–2.5 s).

Experimental Design

Stimuli. We used a set of 51 shapes (Fig. 2*A*) presented at eight rotations at 45° increments to assess neuronal shape selectivity. The design of these stimuli is described in detail elsewhere (Pasupathy and Connor 2001). To minimize repeated presentations at the same spatial location within the receptive field (RF), stimuli were minimally displaced from center of mass before rotation (average displacement was 7% of RF diameter; across the range of recorded RFs, the average displacement was ~0.18°). Stimuli with rotational symmetry (e.g., circles) were treated as identical across congruent rotations (i.e., responses were averaged), yielding a total of 362 unique stimuli in the

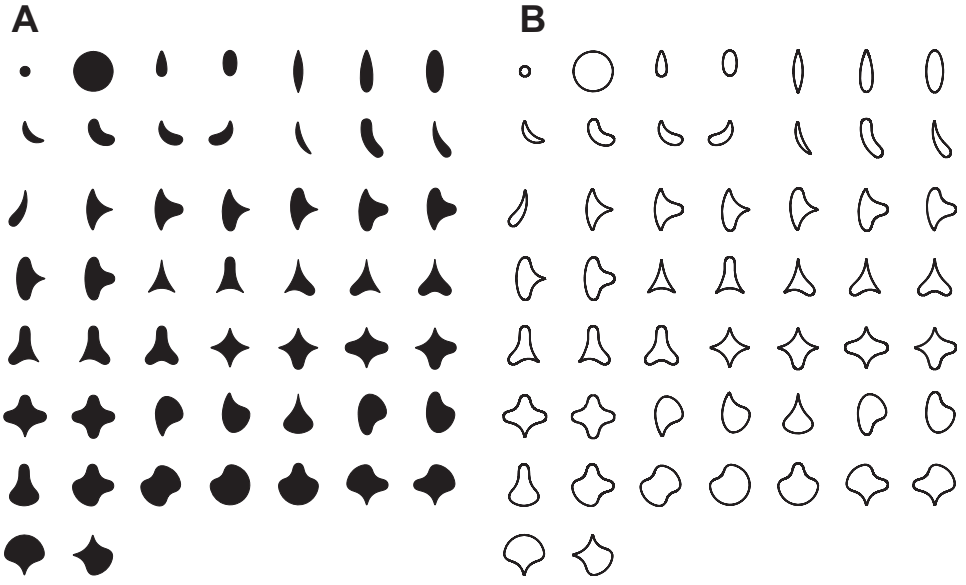


Fig. 2. Stimulus set. Filled (*A*) and outline (*B*) stimuli used to study responses of V4 neurons. *A*: stimuli were presented in up to 8 rotations (fewer in the case of rotational symmetry) in 45° increments. All stimuli were scaled to fit within the receptive field (RF) and were presented in a preferred color and luminance for the neuron under study. *B*: for outline stimuli, line widths ranged from 2 to 4 pixels based on RF eccentricity (see METHODS for details).

set. The stimulus set was scaled such that all parts of all stimuli fell within 80% of the RF diameter, which was estimated as a function of eccentricity based on the data in Gattass et al. (1988).

Outline stimuli were constructed analogously to filled stimuli but with an interior fill equivalent to the background color (Fig. 2*B*). Outline width was set to 2, 3, or 4 pixels (0.05–0.1°), with thicker outlines for neurons with more eccentric RFs. We verified that outline stimuli in the periphery were psychophysically distinguishable by the animal. The two animals whose data are presented here (*M1* and *M2*) were trained to perform a delayed match-to-sample shape discrimination task with the filled versions of the stimuli. Before beginning data collection for the present study, we tested whether *M1* could perform this behavioral task with thin outline stimuli at typical recorded eccentricities. *M1* was able to achieve discrimination accuracy of 73%, averaged across four recorded sessions ($n = 489$ trials) without any additional training with outline stimuli. These sessions included different discriminanda (2 shapes tailored to the preferences of the neuron being recorded in each session), RF locations (1.9–4.7° eccentricity), and outline widths (2–4 pixels), all of which may have influenced task difficulty; nevertheless, performance accuracy was well above chance.

Each stimulus was presented for 300 ms; four or five stimuli were typically shown per trial, with a 200-ms blank interval preceding each stimulus. Filled and outline stimuli were randomly interleaved and presented multiple times. We include neurons for which we were able to repeat each stimulus at least three times (median: 5 repetitions). To evaluate whether differences in neuronal responses to filled and outline stimuli were attributable to overall stimulus intensity, we studied the responses to a subset of five or six outline shapes with two or three outline widths each, ranging from 1 to 6 pixels, for 20 neurons.

Data collection. Each day, we lowered a single tungsten microelectrode (FHC) into area V4 with a stepper motor drive (Gray Matter Research). We amplified and filtered the signals from the electrode and sorted waveforms (Plexon Systems) to identify single units. We include responses from 43 well-isolated neurons (7 from *M1* and 36 from *M2*) for which RF eccentricities ranged from 0.4° to 6.1°. For each neuron, we first assessed the spatial extent of the RF manually, following commonly used procedures. First, we handmapped the borders of the excitatory portion of the RF using diverse stimuli. To elicit a robust neuronal response, we varied the position, shape (bars, gratings, filled and outline stimuli from our set), size, orientation, color, and luminance contrast. Using the handmapped RF borders, we estimated the RF center location assuming a circular RF and reconfirmed this estimate manually. Finally, we estimated RF size based on the RF center location, using eccentricity-dependent scaling following Gattass et al. (1988). We then characterized each neuron's color/luminance preference either manually or via an automated protocol that sampled responses to 25 colors in CIE space at three or four luminance levels (see Bushnell et al. 2011b). All stimuli were subsequently presented in a color and a luminance that elicited a robust response from the neuron.

Analysis. Spiking responses were averaged across multiple presentations of the same stimulus and over the 300-ms duration of each stimulus presentation. Stimulus onset and offset times were detected with a photodiode.

We computed two metrics to compare the responses to filled and outline stimuli. First, to assess similarity of tuning, we computed Pearson's correlation coefficient, r, between responses to filled and outline stimuli for each neuron. Second, to compare the range of responses evoked by filled stimuli to that evoked by outline stimuli, we computed a relative response range metric (ρ) for each neuron. For this, we first computed the average absolute deviation (AAD) from the mean for responses to filled or outline stimuli as

$$\mathrm{AAD} = \frac{1}{n}\sum \left|\overline{x_i} - \bar{x}\right| \quad (1)$$

where $\bar{x}$ is the mean response of the neuron to a stimulus type (i.e., filled or outline), $\overline{x_i}$ is the mean response to a unique stimulus across all presentations, and $n = 362$ stimuli. We then computed the relative response range metric as follows:

$$\rho = \frac{\mathrm{AAD}_{\mathrm{outline}}}{\mathrm{AAD}_{\mathrm{filled}}} \quad (2)$$

A ρ value near 1 indicates similar responsiveness to filled and outline stimuli; $\rho > 1$ indicates a larger range of responses to outline than to filled stimuli, and the converse for $\rho < 1$. Results were very similar when we used standard deviation instead of AAD to measure the range of responses within a stimulus category.

To ask whether preferences for filled versus outline stimuli relate to differences in overall stimulus intensity, we computed a normalized stimulus area by dividing the number of pixels used to depict each shape by the number needed to depict the largest filled shape (the large filled circle in Fig. 2*A*). These values depend on the RF eccentricity for stimulus presentation and, for outline stimuli, the width of the outline. For example, at an eccentricity of 2.4° (median eccentricity in our recordings), normalized stimulus area ranged from 0.065 to 1.0 for filled shapes; from 0.014 to 0.058 for outline shapes with a width of 2 pixels; and from 0.023 to 0.098 for outline shapes with a width of 4 pixels.

To characterize the shape selectivity of each neuron in terms of tuning for boundary curvature, we fit neuronal responses to the angular-position curvature (APC) model of Pasupathy and Connor (2001). First, for each stimulus in our set of 362 shapes, we considered the 8 contour segments along the shape boundary that were used to construct the shape and thus were most characteristic of the shape. For each segment, we calculated the curvature and angular position (see Pasupathy and Connor 2001 for details). The curvature values of segments ranged from −0.3 (moderate concavity) to 1.0 (sharp convexity); their angular position values ranged from 0° (right of center) to 360°, where 0° was to the right of object center and values increased in a counterclockwise direction. Next, we related neuronal responses to these stimulus descriptions: using nonlinear regression, we identified the best-fitting two-dimensional (2D) Gaussian in a space with dimensions of angular position and curvature that best predicted the observed responses. The 2D Gaussian function was defined by five parameters that best predicted the observed responses: two means and two standard deviations for the curvature and angular position dimensions (μ_{curv}, σ_{curv} and μ_θ, σ_θ, respectively) and an amplitude.

To assess neuronal response reliability, we used the Spearman-Brown-corrected correlation of half-split responses. To obtain the half-splits, we computed mean responses for each stimulus based on the odd and even repetitions separately. In the case of an odd number of stimulus repetitions, we omitted the last repetition to ensure that the same number of repeats contributed to each half-split. We then computed a Pearson's correlation coefficient between the half-splits and corrected it with the Spearman-Brown formula [for halves of equal length: $r_{\mathrm{corrected}} = 2 \times r/(1 + r)$].

Model Architecture and Function

Overview. Our instantiation of the feedforward HMax model follows that of Cadieu et al. (2007) and has four main stages. The first two stages are named S1 and C1, in an analogy to simple and complex cells in V1. The S1 units have oriented Gabor function linear filters (Fig. 3*A*) that operate on the input stimulus. C1 units then compute a Max-function (i.e., they simply report the maximum) over their inputs, which consist of many shifted S1 units (Fig. 3*A*) that have the same orientation. This creates phase- and position-invariant tuning for

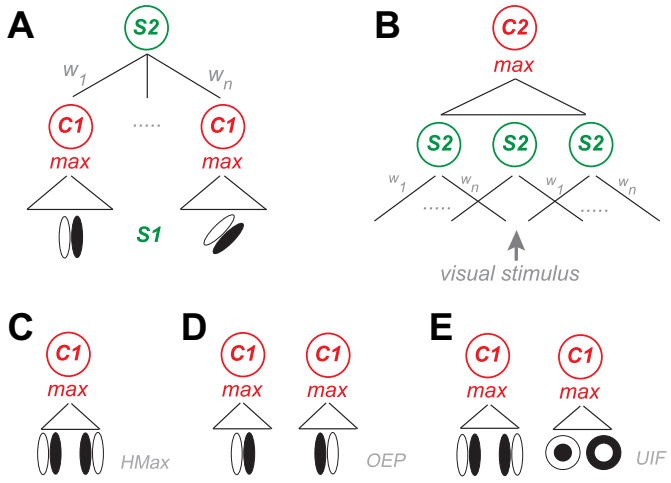


Fig. 3. Hierarchical-max (HMax) model structure and proposed modifications. *A* and *B*: the HMax model described by Cadieu et al. (2007) creates V4-like units selective for specific boundary orientation through alternating operations imparting selectivity (S layers) and invariance (C layers). *A*: S1 units (oval icons) are Gabor filters varying in position, orientation (4 orientations), phase (2 phases), and size (6 sizes). C1 units are arranged in 3 spatial grids (2 × 2, 3 × 3, 4 × 4), each receiving inputs from oriented S1 units and taking a maximum (triangle) response over S1 unit phase, position, and size (2 sizes per grid, inversely proportional to grid tiling). Each C1 grid location has 4 C1 units, 1 per orientation. Each S2 unit receives input from 2–25 C1 units. The selection and strength of these weights (w_1 to w_n) are determined by the fitting procedure to produce a shape-selective template matching the response profile of a given V4 neuron. See METHODS for implementation details. *B*: the C2 unit produces a V4-like response by taking the maximum response over a 3 × 3 grid of S2 units. *C–E*: comparison of original HMax model and the proposed modifications. *C*: in the HMax model, each C1 unit takes a maximum response over the phases, spatial position, and size of S1 units. *D*: to preserve information about interior fill, in the oriented edge polarity (OEP) model each C1 unit takes a maximum response over position and size, but not phase, of S1 units. This results in a doubling of the number of C1 units available for fitting. *E*: as an alternative strategy, in the unoriented interior fill (UIF) model interior fill information is made available to the S2 stage by adding unoriented difference-of-Gaussian units in the S1 stage. Each C1 unit takes a maximum response over position, size, and phase for oriented and unoriented units. This adds 29 additional units to the C1 stage.

orientation, roughly analogous to that of V1 complex cells. The third stage, S2, contains units that compute a weighted average (Fig. 3*A*) of the outputs of a particular set of C1 units. An S2 unit can be thought of as defining a flexible shape template based on the positions and orientations of its chosen C1 inputs. The template is flexible because of the position invariance built into the C1 units. The final stage, C2 (Fig. 3*B*), consists of a Max-function over a grid of spatially offset, but otherwise identical, S2 units. This second Max stage is designed to achieve greater translation invariance (TI). This model was designed to respond selectively to particular shapes regardless of their position within the C2 RF, thereby matching the boundary curvature selectivity and TI of V4 neurons (Cadieu et al. 2007). Each stage in the model is described in detail below. Visual stimuli presented to the model were represented as grayscale images on a 256 × 256-pixel grid and were scaled such that the diameter of the large circle stimulus was 100 pixels.

S1 stage. This stage consists of 2D spatial Gabor (cosine × Gaussian) linear filters at four orientations (0°, 45°, 90°, and 135°), two spatial phases (either 0° and 180° or 90° and 270°), and six spatial scales. Each filter is represented on a square "tile" of $N \times N$ where the tile size, N, defines the scale in pixels and is an element of {32, 36, 40, 45, 54, 60}. The spatial frequency (SF) of the Gabor filter is inversely proportional to N (SF = 3/N, in cycles/pixel), and the SDs of the spatial Gaussian (orthogonal and parallel to the sinusoid) are $SD_{orth} = N/9$ and $SD_{par} = N/6$ (in pixels).

To establish measurement units comparable to electrophysiology, we define 1 pixel to be 0.05° of visual angle. Thus the smallest filters (tile size 32) have SF = 1.875 cyc/° and $SD_{orth} = 0.18°$, and the largest filters have SF = 1 cyc/°and $SD_{orth} = 0.33°$.

The set of S1 responses for each image is computed as the dot product between the filter and the image at each location in the image. The response is then normalized as

$$r_{norm} = \frac{r_{raw}}{\sqrt{\sum_{i,j} S_{i,j}^2 + \varepsilon_{s1}}} \quad (3)$$

where $S_{i,j}$ is the stimulus luminance at pixel (i,j), and the Σ is over a local region (covered by the filter at location i,j) that scales with the tile size. In particular, it consists of all pixels in the stimulus that are within the radius $SD_{par} \times 2.5$. The constant, ε_{s1}, is set to 0.0001 to avoid division by zero.

C1 stage. The C1 units are arranged in three grids: a 4 × 4 grid with small RFs, a 3 × 3 grid with medium RFs, and a 2 × 2 grid with large RFs. Each C1 unit takes the Max over S1 units at two spatial scales and at a specific orientation preference. For example, the C1 units in the 4 × 4 grid take the Max over all S1 units at the first and second tile sizes, i.e., $N = 32$ and 36 (see *S1 stage*), that have a given orientation preference and that have their RFs contained within the 48-pixel square centered on the C1 unit grid position (for full details of C1 RF size and spacing, see Table 1 of Cadieu et al. 2007). The Max is also taken over filters having two spatial phases that are 180° apart (Fig. 3*C*). Thus at each grid position, there are four C1 units corresponding to the four orientations of the S1 Gabor filters. Overall, the three grids, with four orientations at each point, provide 116 C1 units (varying in position, orientation, and scale) for the next stage, S2.

S2 stage. Each S2 unit computes a weighted sum of a set of C1 inputs (Fig. 3*A*) as follows:

$$u = \frac{\sum w_i x_i}{\sqrt{\sum x_i^2 + \varepsilon_{s2}}} \quad (4)$$

where x_i is the response of the ith C1 unit, w_i are the weights, and $\varepsilon_{s2} = 0.0001$. The denominator normalizes the response in a way that makes u relatively insensitive to a scaling over all input signals. The normalized, weighted sum is then passed through a sigmoid function:

$$g(u) = \frac{s}{1 + e^{-a(u-b)}} \quad (5)$$

where s is the maximum value, a is the slope, and b is the horizontal offset. These three sigmoid parameters and the weights in the previous equation are the only parameters that are fit to the V4 data (see *Model Fitting and Testing*).

C2 stage. There is one C2 unit, and it computes Max over a 3 × 3 grid of S2 units, each with identical weights on their respective C1 inputs. This C2 unit thus has a larger RF than the S2 units, and its output value is that which we compare to V4 responses during training and testing.

Model Fitting and Testing

The HMax model was fit to the V4 neuron responses in a manner similar to that described by Cadieu et al. The neuronal data consisted of the responses to the 362 shape stimuli, averaged across repeated trials for each stimulus. These 362 responses were divided randomly into a training set (5/6 of the data) and a testing set (1/6 of the data). During the training phase, only the three parameters of the S2 sigmoid (s, a, and b, *Eq. 5*) and the weights from the C1 to the S2 layer were allowed to vary. The training phase consisted of two parts; first, the best fit was found for a model having only two nonzero weights from the C1 to the S2 layer. In particular, all unique pairs of C1 weights

were tried from the middle C1 grid (the 3 × 3 grid with 4 orientations at each point). For each pair of weights, gradient descent was performed on the five parameters (s, a, b, w_1, w_2) from many starts (typically 625). Once the best model was found that had two C1 units from the middle grid, the second phase of fitting began. Here, we iteratively tested and included the next best C1 unit by trying all remaining C1 units and fitting the three sigmoid parameters and existing C1 weights along with the weight, w_i, of the additional C1 unit, using gradient descent from an array of starting points for the new weight. We continued this procedure until 25 C1 units (following Cadieu et al.) had been chosen.

For each of the 43 neurons studied here, we repeated this fitting procedure 10 times, using average responses to 5/6 of the stimuli (randomly selected) as the training set and the remaining 1/6 as the testing set. We computed an average mean squared error (MSE) across the 10 test sets as a function of the number of C1 units (from 2 to 25). We then determined the smallest number of C1 units necessary for the model to achieve an MSE value within 1% of the minimum average cross-validated MSE. We obtained final results for each neuron by fitting the model to all data (combining the training and testing sets) but constraining the number of C1 units as determined above.

We compared model responses to neural data with Pearson's correlation coefficient (r) between predicted and observed responses, the MSE from the fitting procedure, and the relative response range metric (ρ) described above. We also compared the variance explained by each model (Yamins et al. 2014), calculated as r^2 values normalized by the reliability of the measured responses (Spearman-Brown-corrected correlation of half-split responses).

Model Modifications

To better describe V4 responses that were dictated by both interior fill and boundary form, we modified the original HMax model in two ways, as described below, to allow information about interior fill to be maintained.

Oriented edge polarity. The original HMax model discards information about spatial phase, and thus the polarity of edges, by taking the Max over two complementary phases (Fig. 3*C*). We created a variant of the HMax model that maintains this phase information, passing it from the S1 to the C1 level (Fig. 3*D*), and we refer to this as the oriented edge polarity (OEP) model. This change doubled the number of C1 units (116 for each phase), for a total of 232 units available during the selection of the best 2–25 C1 inputs to each S2 unit, but the rest of the model structure and fitting procedure were preserved. To ensure that any change in performance was not caused simply by the addition of units, we also created a version of the OEP model with the same number of C1 units as the HMax model. Specifically, this model (the "restricted OEP" model) had access only to one phase for each pair of C1 units; the other phase, chosen randomly, was discarded. This maintained a total of 116 units available for the fitting stage.

Unoriented interior fill. To evaluate whether unoriented units, separate from those representing the oriented boundary, were important to signal object fill, we added difference-of-Gaussian (DoG) units to the pool of oriented Gabor filters in the S1 layer (Fig. 3*E*). It is important to note that this remains a phase-independent model, in which C1 units take the Max over phase. We refer to this model variant as the unoriented interior fill (UIF) model. The Gaussian RFs for the unoriented units were generated as follows:

$$\mathrm{DoG} = \frac{1}{2\pi\sigma_{\mathrm{ctr}}^2} e^{\frac{-[(x-x_0)^2+(y-y_0)^2]}{2\sigma_{\mathrm{ctr}}^2}} - \frac{1}{2\pi\sigma_{\mathrm{surr}}^2} e^{\frac{-[(x-x_0)^2+(y-y_0)^2]}{2\sigma_{\mathrm{surr}}^2}} \quad (6)$$

where x and y are coordinates along the horizontal and vertical axes of space, x_0 and y_0 define the spatial center of the DoG, and σ_{ctr} and σ_{surr} are the standard deviations of the center and surround Gaussian components, respectively (N/5 and N/3, where N is the tile size defined in *S1 stage*). The final amplitude of the DoG, as with the Gabor filters, was determined by the weights chosen during the fitting process. As in the HMax model, we took the Max over complementary phases of the DoGs at the C1 level. This model modification added 29 unoriented C1 units to the original 116 oriented C1 units, for a total of 145 units available during the selection of C1 to S2 inputs. Other than the addition of DoG units in the S1 layer, the rest of the model architecture and fitting procedure remained identical. To compare models that had the same number of distinct C1 units, we also created a "restricted UIF" model that had a total of 116 units available. To accomplish this, we reduced the number of orientations for the Gabor C1 units to three (0°, 60°, and 120°) instead of four to produce 87 oriented units and added 29 unoriented units for a total of 116.

In addition to comparing the performance of our restricted models, which match the original HMax model in total free parameters (total C1 units available for fitting), we also analyzed the performance of our unrestricted modified models and accounted for the difference in the number of free parameters with the Bayesian information criterion (BIC). We computed BIC as $n\cdot\ln(\mathrm{RSS}/n) + k\cdot\ln(n)$, where RSS is the residual sum of squared differences between model and neuronal responses, n is the number of observed responses (724, 1 per stimulus), and k is the number of model parameters (3 sigmoid parameters plus the number of C1 units used in each model). We found equivalent results when using the Akaike information criterion (AIC) as an alternative to BIC.

To estimate translation invariance (TI), we measured the responses of each model to our stimulus set presented on a 5 × 5 grid of 25 locations (following Cadieu et al. 2007, their Fig. 6), where the center grid position was the one used for model fitting. We computed TI as the normalized covariance (Pospisil et al. 2017) between responses across all unique pairs of grid locations:

$$\mathrm{TI} = \frac{\sum_{i\neq j}\mathrm{Cov}(\vec{p}_i,\ \vec{p}_j)}{\sum_{i\neq j}\mathrm{SD}(\vec{p}_i)\mathrm{SD}(\vec{p}_j)} \quad (7)$$

where i and j are positions in the grid, p_i is the mean-subtracted column of responses to shapes at the ith position, and the sums are taken over all unique pairs of locations. This metric can be thought of as a generalization of the correlation coefficient (r value) from 2 to n variables, and like the r value it is bounded between −1 and 1. It has an advantage over simply computing the average r value across all unique pairs of locations because the latter would weight the r value from RF locations with very weak responses just the same as those with very strong responses.

For TI, as for other comparisons between models, we report values averaged across models with odd- and even-symmetric filters because individual (odd- or even-symmetric) predictions were typically similar and did not affect our conclusions. However, for one outlier, the even-symmetric HMax model produced a degenerate fit to the neuronal data ($r = 0.05$, compared with $r = 0.69$ for the model fit to same data with odd-symmetric filters). In this instance, we report TI values for only the odd-symmetric model.

Statistical Analysis

For all statistical tests presented here, we rejected the null hypothesis at the 5% significance level ($\alpha = 0.05$). For all correlation values reported, we computed the Pearson correlation coefficient and associated P values testing the hypothesis that variables are not correlated. To determine whether neurons were sensitive to stimulus luminance contrast, we used data from preliminary characterization of each neuron's color and luminance preferences. We performed a two-way ANOVA to test for effects of stimulus luminance contrast, color, or interaction between luminance contrast and color on neuronal re-

sponses. We considered neurons to be luminance sensitive if there was a significant main effect of luminance contrast and/or the interaction term ($\alpha = 0.05$). To compare performance of models across our population, we performed Wilcoxon signed-rank tests and report P values testing the null hypothesis that the difference between model performances for individual neurons is zero. Additional details are reported in RESULTS and corresponding figures.

RESULTS

We first describe our experimental observations related to the neuronal encoding of filled versus outline shapes in V4. We then demonstrate how an existing biologically inspired model of V4 shape tuning, an HMax model, can be improved to account for our new data.

Responses of V4 Neurons to Filled and Outline Shapes

To determine whether form selectivity in area V4 is invariant across filled and outline shapes, we recorded responses from well-isolated V4 neurons ($n = 43$) in two awake, fixating macaques for the two stimulus sets shown in Fig. 2. The filled shapes (Fig. 2*A*) consisted of unique combinations of convex and concave features and have been used in a series of studies to characterize V4 tuning for boundary curvature (Bushnell et al. 2011a; Bushnell and Pasupathy 2012; Kosai et al. 2014; Oleskiw et al. 2014; Pasupathy and Connor 2001). The outline shapes had the same boundary curvature and were rendered in the same color and luminance as the filled shapes.

Example neurons. Neurons in V4 showed a variety of response patterns to filled and outline shapes, which we demonstrate with four example neurons. The average responses of our first example neuron, *neuron 1*, are shown in Fig. 4 for filled and outline shapes (Fig. 4, *A* and *B*, respectively), where darker background shading indicates stronger responses. This neuron exhibited similar shape preferences for filled and outline stimuli. Specifically, it responded strongly to many of the filled shapes that included a rightward-facing concavity (Fig. 4*A*). In comparison, stimuli with a convexity to the right of the shape evoked virtually no spikes (Fig. 4*A*). This neuron's responses to outlines (Fig. 4*B*) were weaker than those to filled shapes, but the shape tuning was similar: responses were stronger for outline shapes with a concavity to the right and weaker for outline shapes with a convexity to the right of object center. Responses to filled and outline stimuli, plotted against each other in Fig. 5*A*, were highly correlated ($r = 0.91$, $P < 0.001$), reflecting the similarity in shape tuning for both sets of stimuli. However, the scatter was compressed along the vertical axis, indicating weaker responses for outlines. We quantified this latter trend with a relative response range metric, ρ (see METHODS), which was 0.59 for *neuron 1*. Overall, despite a greater modulation of neuronal responses by filled stimuli, shape preferences were similar for filled and outline stimuli.

A strong correlation in responses between filled and outline stimuli, as observed in *neuron 1*, is consistent with the structure of the HMax model (see METHODS) for V4 shape selectivity proposed by Cadieu and colleagues (Cadieu et al. 2007). In this model, shape selectivity and TI are built by pooling weighted signals from V1-like oriented filters followed by Max-pooling (Fig. 3, *A* and *B*). Because the HMax model relies on phase-invariant band-pass orientation-tuned filters to build shape selectivity, its response should largely be determined by the shape, not fill, of the stimulus. In other words, the shape selectivity predicted by the model should be invariant of interior fill. To verify this, we fit the HMax model to the filled responses of *neuron 1* (following Cadieu et al.; see METHODS) and then used the model to predict outline responses. The best fit model, which had a high correlation between predicted and observed responses to filled stimuli ($r = 0.66$; $P < 0.001$), did indeed produce responses to outline stimuli that were strongly correlated with those to filled shapes (Fig. 5*E*; $r = 0.87$, $P < 0.001$). The model produced a ρ value close to 1 ($\rho = 1.16$), indicating that this model unit was driven equally well by filled

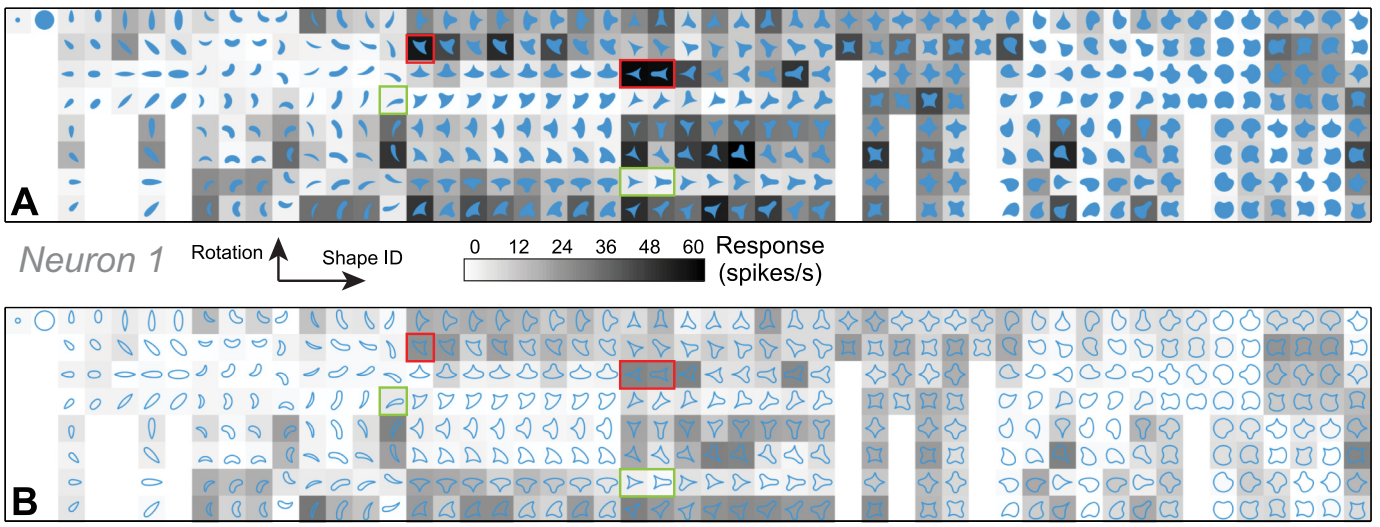


Fig. 4. Responses of an example V4 neuron to filled and outline stimuli. Responses of *neuron 1* (*o140915*) to filled (*A*) and outline (*B*) versions of 51 shapes (*x*-axis) at up to 8 rotations (*y*-axis) are shown. The background gray level of each icon represents the mean response to the stimulus drawn within. Darker backgrounds correspond to higher firing rates (see scale bar). Responses were weaker for outline stimuli, but shape preferences were similar. Stimuli that elicited stronger spiking responses from this neuron (red boxes) included a concave feature to the right side of the shape, whereas shapes that evoked a weaker response (green boxes) did not. Responses to both filled and outline stimuli were well fit by an angular position and curvature model with a shallow concavity to the right. Fitted means for the curvature and angular position dimensions based on filled stimuli: $\mu_\theta = 350°$, $\mu_{curv} = -0.12$ and outline stimuli: $\mu_\theta = 345°$, $\mu_{curv} = -0.14$. In both cases, correlation between predicted and observed responses was 0.87.

and outline stimuli. We obtained similar results when we fit the HMax model to the outline responses of *neuron 1* and used the model to predict filled responses (Fig. 5*F*). The best fit model had a high correlation between predicted and observed responses ($r = 0.69$, $P < 0.001$). Responses to filled stimuli were strongly correlated with those to outline shapes ($r = 0.84$, $P < 0.001$), and the model unit was driven equally well by filled and outline stimuli ($\rho = 1.12$). Fitting the model to filled or outline responses of other neurons produced similar predictions. These results highlight the HMax model's ability to generalize form selectivity from filled stimuli to outline stimuli and vice versa.

Many neurons in our data set, however, were unlike *neuron 1*: their responses to filled and outline stimuli were not well correlated. This is illustrated by *neurons 2–4* in Fig. 5, *B–D*. *Neuron 2* responded strongly and selectively to filled shapes (Fig. 6*A*), but the responses to outline shapes (Fig. 6*B*) were much weaker and did not reflect the selectivity observed for filled shapes. The scatterplot of outline versus filled responses for *neuron 2* (Fig. 5*B*) was distinctly different from that for *neuron 1* (Fig. 5*A*): the responses of filled and outline shapes were poorly correlated ($r = 0.25$, $P < 0.01$), and the relative response range, $\rho = 0.38$, was substantially lower. *Neuron 3* responded strongly and selectively to outline stimuli (Fig. 7*B*), but unlike the previous examples it responded weakly to filled shapes (Fig. 7*A*). There was little correlation between filled and outline responses ($r = 0.12$, $P = 0.026$; Fig. 5*C*), and ρ was 2.67, reflecting a stronger response modulation for outlines. *Neuron 4* (Fig. 5*D*) was driven above baseline (20.7 spikes/s) and over a similar range by both stimulus types ($\rho = 0.99$), but the patterns of selectivity for filled and outline shapes were dissimilar ($r = 0.15$, $P = 0.003$). The diversity in these three examples, and their dissimilarity from the behavior of *neuron 1*, can be visualized by comparing the four scatterplots for outline versus filled responses depicted in Fig. 5, *A–D*. Only the behavior of *neuron 1* is consistent with the predictions of the HMax model, that responses for filled and outline stimuli should be similar in magnitude and strongly correlated.

Population data. Figure 8*A* shows the distribution of our two metrics, correlation in responses between filled and outline stimuli (*y*-axis) and the relative range of responses ρ (*x*-axis), across our data set of 43 V4 neurons. Although a few of the neurons we recorded showed a strong correlation between responses to filled and outline stimuli, a majority of the neurons did not: only four neurons (1 in *M1*, 3 in *M2*) had $r > 0.5$, and the median was 0.16. V4 neurons showed a broad range of ρ values (0.2–2.67, median 0.69), with many neurons exhibiting a compressed range of responses to one or the other category of stimuli. Across our data set, a majority of neurons exhibited a larger range of responses for filled stimuli (33/43 neurons had $\rho < 1$; *M1*: 6/7, *M2*: 27/36), and the mean (0.86)

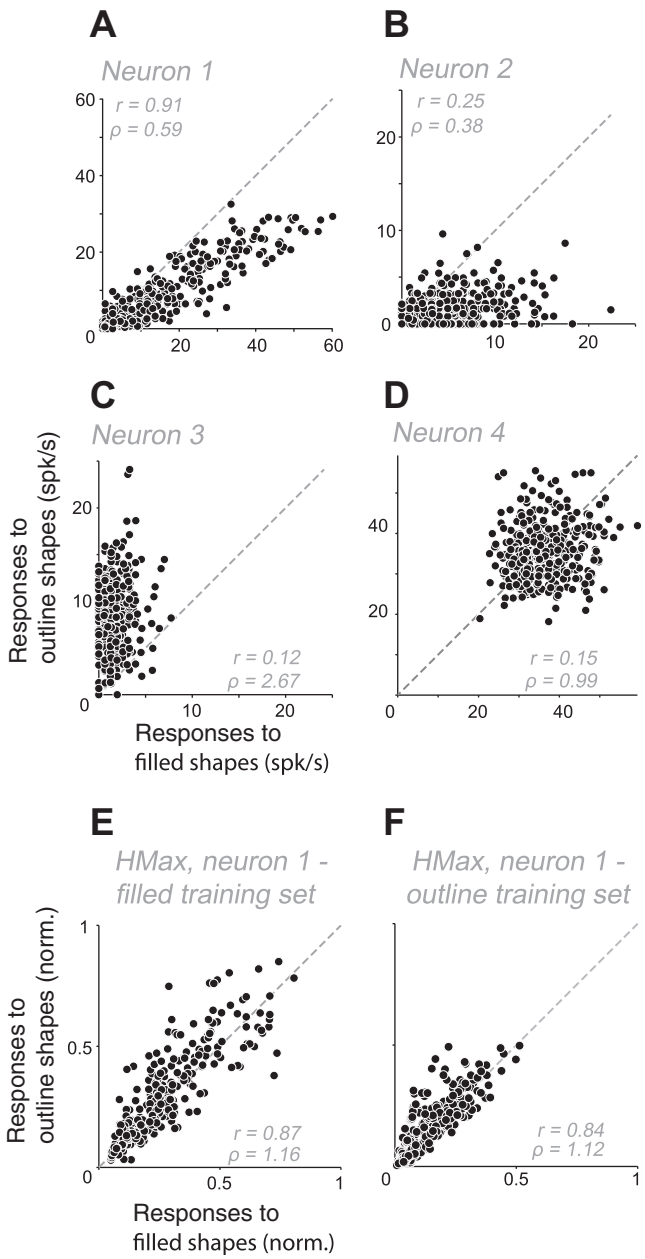


Fig. 5. Responses of example V4 neurons and hierarchical-max (HMax) model predictions. *A–D*: responses of *neurons 1–4* to filled and outline stimuli. Each point represents the mean response [spikes (spk)/s] to the presentation of an outline (*y*-axis) and a filled (*x*-axis) stimulus with the same boundary shape. *A*: for *neuron 1* (same as in Fig. 4), responses to filled and outline stimuli were strongly correlated ($r = 0.91$, $P < 0.001$). Filled stimuli evoked a larger range of responses than outline stimuli, and the relative range of responses (ρ) was 0.59. *B*: for *neuron 2*, filled shapes evoked a larger range of responses than outline shapes ($\rho = 0.38$). Correlation between responses to filled and outline stimuli was weak ($r = 0.25$, $P < 0.001$). *C*: for *neuron 3*, outline stimuli evoked a large range of responses and ρ was 2.67. Here again, correlation between responses to filled and outline stimuli was weak ($r = 0.12$, $P = 0.026$). *D*: filled and outline stimuli evoked a similar range of responses from *neuron 4* ($\rho = 0.99$), but correlation between filled and outline responses was weak ($r = 0.15$, $P = 0.003$). *E*: responses to filled and outline stimuli predicted by an instantiation of the HMax model that was trained on responses of *neuron 1* to filled stimuli (*x*-axis in *A*). Most points lie along the line of equality, i.e., the model predicts similar responses to outline and filled shapes that share identical boundary configuration. Correlation between predicted responses to filled and outline stimuli was high ($r = 0.87$, $P < 0.001$), and ρ was 1.16. The HMax model template included 25 C1 units, and the predictions shown here are based on odd-symmetric S1 filters, which achieved better performance than even-symmetric filters in this case. *F*: responses predicted by an instantiation of the HMax model that was trained on responses of *neuron 1* to outline stimuli (*y*-axis in *A*). As in *E*, the model predicts similar responses to outline and filled shapes that share identical boundary configuration. Correlation between predicted responses to filled and outline stimuli was high ($r = 0.84$; $P < 0.001$), and ρ was 1.12. The HMax model template included 22 C1 units, and the predictions shown here are based on odd-symmetric S1 filters, which achieved better performance than even-symmetric filters in this case.

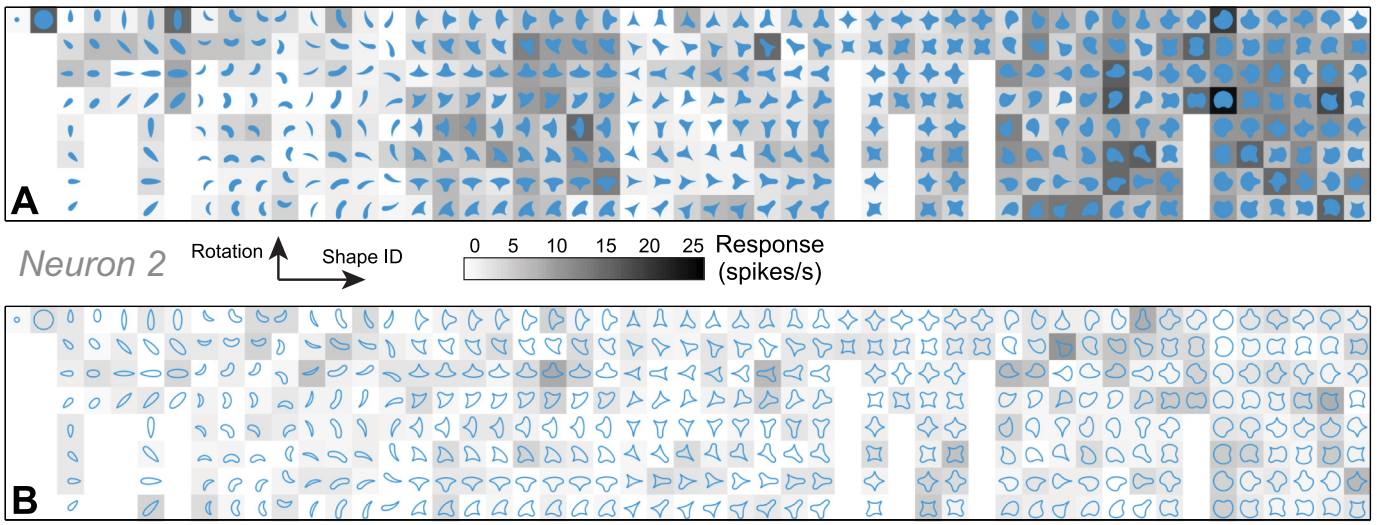


Fig. 6. Responses of *neuron 2* to filled and outline stimuli. Same format as in Fig. 4. *Neuron 2* (*p150428*; also shown in Fig. 5*B*) responded strongly and selectively to filled stimuli but not to outlines.

was significantly less than 1 (Wilcoxon signed-rank test using log-transformed ρ, $P < 0.001$). Despite the small data set from *M1*, the two primary results reported here are consistent between the two animals: few neurons showed a strong correlation between filled and outline stimuli, and a majority of neurons exhibited a larger range of responses to filled than to outline stimuli.

For each neuron in our data set, we fit the HMax model to the responses to filled shapes and used the best fit model to predict outline responses (as shown in Fig. 5*E* for *neuron 1*). Across the 43 HMax models that best fit our data (1 model per neuron), the mean correlation and ρ were 0.85 (SD 0.11) and 1.35 (SD 0.48), respectively. For comparison to the neuronal data, this point is denoted in Fig. 8*A*. These values deviated substantially from those observed across our data set; compare with the equivalent point for neuronal data in Fig. 8*A*. Thus our data suggest that V4 neurons show more diversity in their responses to filled and outline stimuli than what is predicted by HMax models of the form previously used to account for boundary curvature tuning in V4 (Cadieu et al. 2007).

Next, we explored whether the diversity in our metrics across neurons was associated with variations in other characteristics, including properties such as their RF eccentricity, response reliability, preference for stimulus contrast or overall stimulus intensity, and curvature tuning.

Factors Contributing to Different Responses to Filled vs. Outline Stimuli

RF eccentricity. It is possible that a larger range of responses, i.e., greater shape selectivity, with filled versus outline stimuli is correlated with RF eccentricity. Filled and outline stimuli vary in terms of their spatial frequency (SF) content, with filled stimuli having more power at low SF. Given that SF tuning varies with eccentricity in V1 (Tootell et al. 1988) and that SF tuning properties in V4 scale similarly to those in V1 (Desimone and Schein 1987), it is conceivable that neurons with more eccentric RFs exhibit a larger range of responses with filled stimuli. Across our population of neurons, RF eccentricity ranged from 0.4° to 6.1° (Fig. 8*B*, *y*-axis) and there was a negative correlation between ρ and RF eccen-

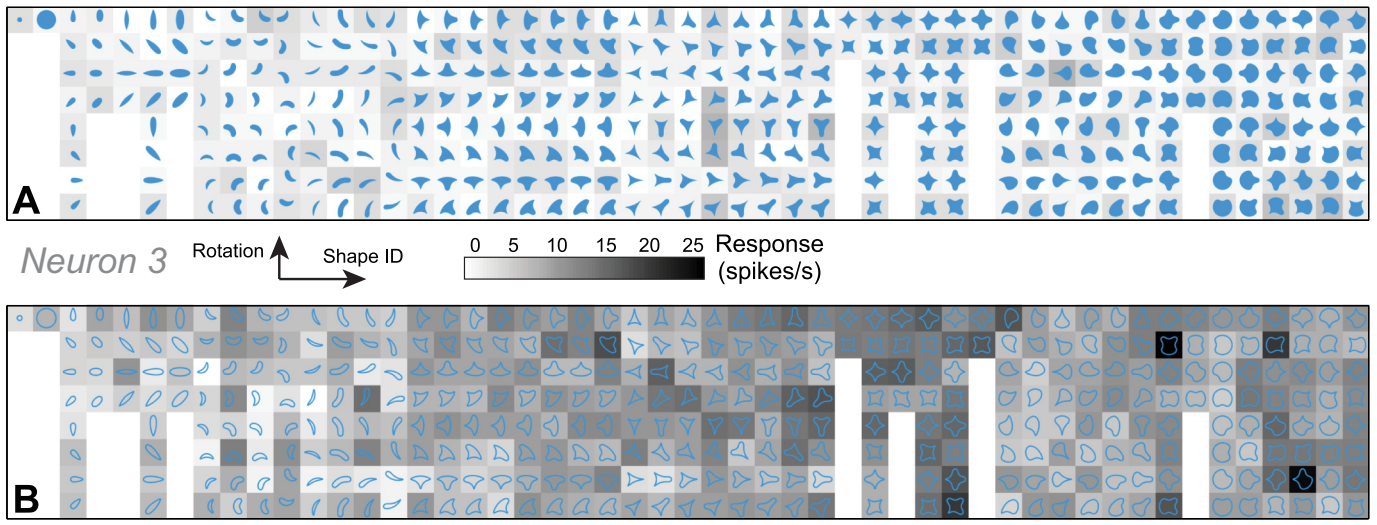


Fig. 7. Responses of *neuron 3* to filled and outline stimuli. Same format as in Fig. 4. *Neuron 3* (*p150415*; also shown in Fig. 5*C*) responded strongly and selectively to outlines but not to filled stimuli.

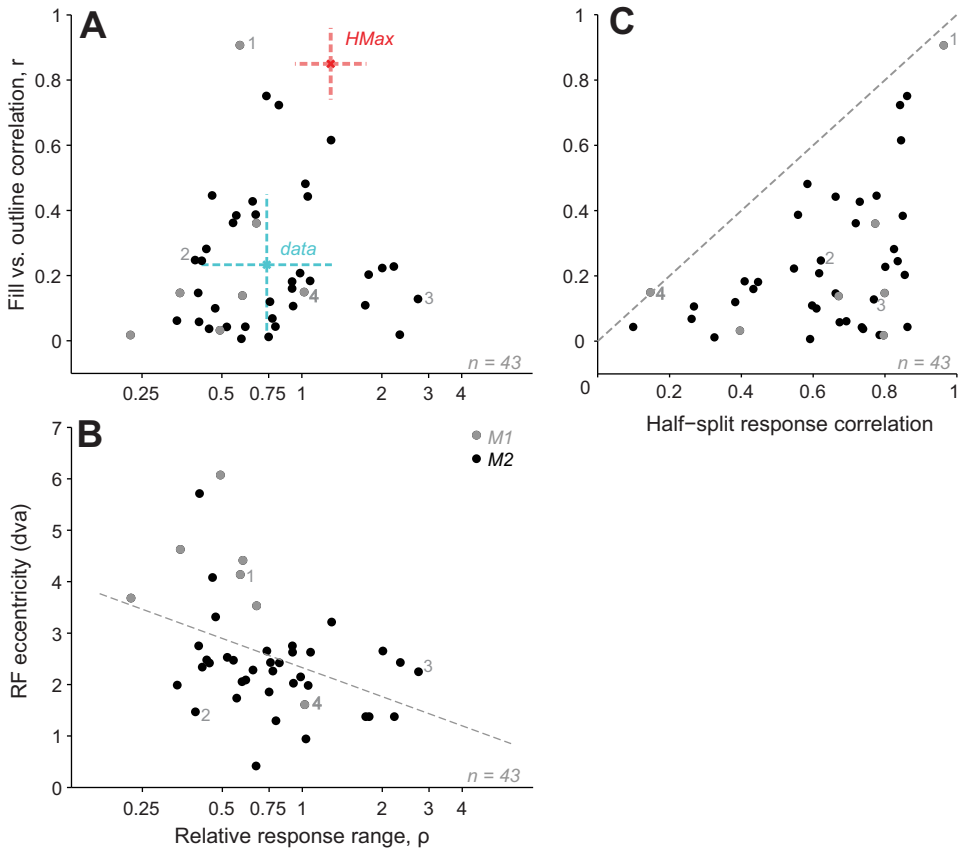


Fig. 8. Population results. *A*: correlations between the responses to filled and outline stimuli (*y*-axis) are plotted against the relative range of responses to filled vs. outline stimuli (ρ, *x*-axis; see METHODS) for each neuron in our data set. Neurons with $\rho > 1$ have a larger range in their responses to outline compared with filled stimuli. Mean values across the recorded neurons for correlation and ρ were 0.23 (SD 0.22) and 0.74 (SD 0.56), respectively. For comparison, symbols representing the mean (denoted by ×) and standard deviation (dashed lines) of fill/outline response correlation and ρ for the data (cyan) are shown alongside the same values across a set of 86 hierarchical-max (HMax) models trained on data from the 43 V4 neurons (2 models per neuron, corresponding to even- and odd-symmetric S1 filters; red). Example *neurons 1–4* are identified, neurons from *animal M1* in gray and from *animal M2* in black. *B*: receptive field (RF) eccentricity (*y*-axis) is plotted against the relative range of responses to filled vs. outline stimuli (ρ, *x*-axis) for each neuron. There was a statistically significant negative correlation ($r = -0.4$, $P = 0.008$), suggesting that filled stimuli evoked a larger range of responses at greater eccentricities. The regression slope is indicated (dashed line). Example *neurons 1–4* are identified, neurons from *M1* in gray and from *M2* in black. dva, Degrees of visual angle. *C*: correlations between the responses to filled and outline stimuli (*y*-axis) are plotted against response reliability (Spearman-Brown-corrected half-split correlation, *x*-axis) for each neuron. Example *neurons 1–4* are identified, neurons from *M1* in gray and from *M2* in black.

tricity for both animals (*M1*: $r = -0.45$, $P = 0.3$; *M2*: $r = -0.27$, $P = 0.1$) but neither reached significance, presumably because of the small data set. When the data from the two animals are taken together, the negative correlation between ρ and RF eccentricity was statistically significant ($r = -0.4$, $P = 0.008$; Fig. 8*B*). For neurons with RFs within 4° of the fovea, ρ ranged from 0.2 to 2.67, reflecting the presence of both types of neurons: those with a response range that was independent of fill ($\rho \approx 1$) and others with a response range that was dependent on interior fill. The small sample from *M1* included only one neuron in this RF eccentricity range that was driven by both filled and outline stimuli ($\rho \approx 1$). For neurons at more peripheral RF locations (eccentricity > 4°), ρ was <0.59 for all neurons, i.e., neurons exhibited a greater diversity of responses to filled stimuli than to outlines. This bias toward greater shape selectivity with filled stimuli at larger eccentricities is consistent with a preference for lower spatial frequencies away from the fovea. The correlation between responses to filled and outline stimuli was not significantly associated with eccentricity (using *z*-transformed values; $r = 0.05$, $P = 0.74$).

Response reliability. Because the mean response for each stimulus was based on five or fewer repeats, noise accounted for a substantial fraction of the total variance: 37.6% on average across neurons (noise variance = average squared standard error across stimuli; see Pasupathy and Connor 2001). This observation is consistent with previous data from Pasupathy and Connor (2001), where stimuli were also presented for five repetitions. To examine whether the inherent variability in neuronal responses contributed to the observed differences between responses to filled and outline stimuli, we compared the correlation between filled and outline responses against half-split response correlation for each neuron (Spearman-Brown corrected; Fig. 8*C*). These values were significantly correlated (Fisher *r*-to-*z* transformed, $r = 0.6$, $P < 0.001$) but only because of the inevitable lack of points in the upper left: there cannot be highly unreliable neurons (i.e., neurons with low half-split correlation) that reliably signal the same shape tuning for filled and outline stimuli. More importantly, many neurons with high split-half reliability values (e.g., >0.5), exhibited a low correlation between filled and outline responses, suggesting that response reliability does not

explain the observation that many neurons have low correlation between outline and fill responses.

To further consider the possibility that poor correlation between fill and outline responses was simply due to a dramatic gain change along one axis that pushes the responses down to noise levels (without a change in selectivity), we conducted a simulation to ask how much the fill-outline correlation would be underestimated for *neurons 2–4* because of noise. Briefly, for each neuron, we took the mean responses for fill or outline (whichever exhibited the larger range of responses) to be the true responses to that class of stimuli and subtracted the baseline firing rate (i.e., response to a blank stimulus). We then predicted the mean responses to the other class by applying a gain scaling factor and then adding back the baseline firing rate. Using this prediction, we generated individual trial responses assuming Poisson noise. Finally, we computed mean responses and the observed correlation. This process was repeated 1,000 times to obtain the following 95% confidence intervals: for *neuron 2* correlation values fell in the range 0.86–0.91, for *neuron 3* in the range 0.59–0.70, and for *neuron 4* in the range 0.92–0.94. Assuming even worse noise levels (Fano factor = 2), responses simulated via gain scaling maintained correlation values > 0.46. These simulations demonstrate beyond a doubt that the low correlation between fill and outline responses cannot be driven simply by noise. Thus our data exhibit changes in both gain and selectivity for responses to fill and outline stimuli.

Response reliability was high for example *neurons 1–3* (Spearman-Brown-corrected $r = 0.95, 0.62, 0.77$, respectively; numbered in Fig. 8*C*) and low for *neuron 4* (Spearman-Brown-corrected $r = 0.15$). Neurons with properties similar to *neuron 4* (low fill-outline correlations and similar fill-outline response ranges) exhibited lower response reliability (mean = 0.34) than other neurons we recorded (mean = 0.73).

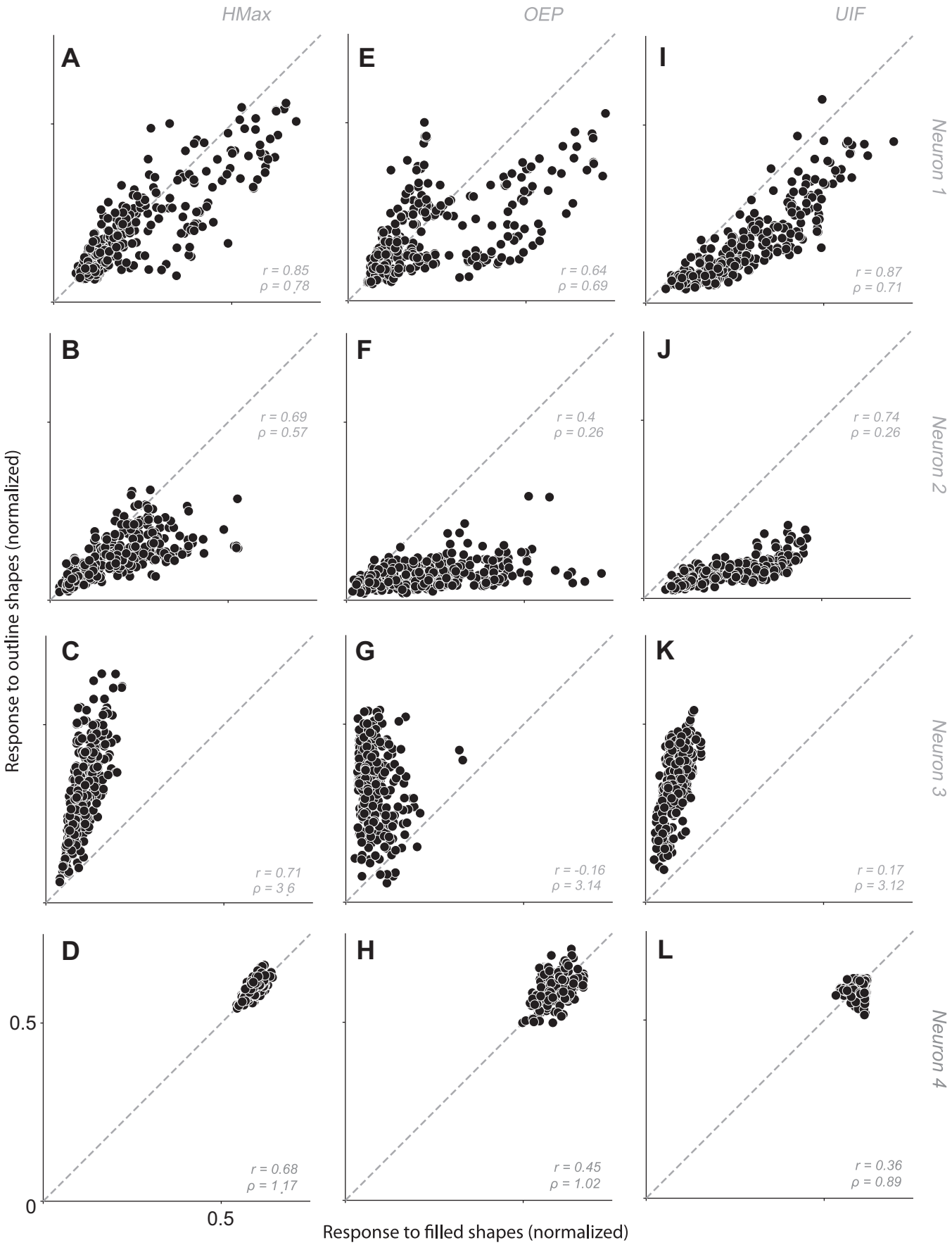
Overall stimulus intensity. Because outlines contain fewer stimulus pixels than corresponding filled stimuli, they impart less change in integrated luminance to the neuronal RF. To determine whether this factor could explain the difference in the range of responses to filled and outline stimuli, we compared the responses of 20 neurons for two outline widths, one twice the other (e.g., 2 and 4 pixels), for a subset of five or six shapes. For the range of eccentricities in our data set, increasing the line width from 2 to 4 pixels produced a sizable change in stimulus intensity; e.g., at an RF eccentricity of 2.4° (the median across our data set), the large circle outline stimulus at 2- and 4-pixel widths had a normalized stimulus area that was 5% and 9% of the large filled circle, respectively. Nevertheless, there was no significant change (paired *t*-test, $P > 0.05$) in response strength with doubling of the outline width for 11/20 neurons; 4/20 neurons showed a significant increase in response strength and 5/20 showed a significant decrease in response strength with doubling of outline width. Thus we found no systematic evidence across our set of 20 neurons for a monotonic increase in neuronal responses with increasing intensity.

Luminance contrast tuning. Although a majority of V4 neurons are sensitive to stimulus luminance contrast, we previously found that a small fraction (~11%) are not (Bushnell et al. 2011b). For this small fraction of neurons, responses to different chromatic contrast are not modulated by the addition of a luminance contrast. We wondered whether there was a relationship between insensitivity to luminance contrast and shape selectivity that is invariant of object fill. Specifically, are neurons with similar tuning for filled and outline shapes also insensitive to luminance contrasts? For six of eight neurons with a high correlation between responses to filled and outline shapes ($r > 0.4$), we had characterized color and luminance preferences in detail (see METHODS). Responses of five of six of these neurons were significantly modulated by luminance contrast (2-way ANOVA for luminance and color, significant main effect for luminance and/or interaction term, $P < 0.05$), suggesting that insensitivity to luminance contrast is not a hallmark feature of neurons that exhibit similar shape selectivity with filled and outline stimuli.

Curvature tuning. To determine whether a larger range of responses for filled versus outline stimuli related to selectivity for boundary curvature, we fit the responses of our V4 neurons to the angular position and curvature (APC) model that we previously used to characterize shape tuning in a series of studies (El-Shamayleh and Pasupathy 2016; Kosai et al. 2014; Pasupathy and Connor 2001). We assessed curvature tuning separately for each neuron's responses to filled or outline stimuli, since neurons with poor responses to one or the other stimulus type show low or no selectivity for those stimuli. For each neuron, we obtained two goodness-of-fit values (one for responses to filled stimuli and one for responses to outline stimuli) and took the higher value to be representative of that neuron's curvature tuning strength. The median goodness of fit across the population was 0.41 (ranging from 0.14 to 0.87; see DISCUSSION for comparison to previous studies). Goodness of fit was not correlated with ρ ($r = -0.16$, $P = 0.29$), suggesting that curvature tuning was independent of modulation by interior fill.

Extending the HMax Model

Results presented above suggest that the HMax model, previously shown to be a good fit to V4 boundary curvature tuning (Cadieu et al. 2007), predicts similar responses for filled and outline shapes and thus does not capture the diversity in our V4 data. Our evaluation of model performance above was based on fitting the HMax model with filled stimuli alone. To more comprehensively assess the ability of the HMax model to capture a range of fill-outline correlations and ρ values, we refit the model using neuronal responses to both filled and outline stimuli and assessed correlation and ρ values from the predicted responses. Training on both filled and outline stimuli still did not produce a model that accounted for the observed data, as exemplified by the best fit to the responses of *neuron 2* (compare neuronal data in Fig. 5*B* to model in Fig. 9*B*). The resulting model was able to produce a more compressed response range for outlines, but a strong positive correlation ($r = 0.69$, $P < 0.001$) between model responses to filled and outline shapes persisted. The same held for the other example neurons. For *neuron 3*, the model could account for the compressed range of responses to filled shapes (compare neuronal data in Fig. 5*C* to model in Fig. 9*C*), but a significant correlation ($r = 0.71$, $P < 0.001$) remained between model responses to filled and outline stimuli. For *neuron 4*, the model matched the relative range metric of responses (compare neuronal data in Fig. 5*D* to model in Fig. 9*D*) but not the actual spread, and it predicted a higher correlation between filled and



outline stimuli ($r = 0.68$, $P < 0.001$) than was observed for this neuron. Unlike *neurons 1–3*, responses of *neuron 4* were less reliable (as mentioned above) and poorly fit by the APC model: the goodness of fit was $r = 0.3$ for *neuron 4* (compared with $r = 0.87$, 0.47, and 0.56 for *neurons 1–3*). This neuron may be selective for a stimulus property not represented in our stimulus set (e.g., color, texture). In such a case, poor curvature tuning may explain why the HMax model could not capture these responses. Overall, even when trained to distinguish filled and outline stimuli, the HMax model was not able to capture the differential responses to these stimulus types observed in the V4 data.

To overcome this limitation of the original model, we considered two modifications aimed at making additional information about interior fill more explicitly available to the V4-like S2 (shape template) stage of the HMax model. In the original model, C1 units were phase insensitive because of Max-pooling across phase of the simple S1 Gabor units (Fig. 3*C*), resulting in a loss of information about the interior of the stimuli. To preserve phase information, we removed this operation (Fig. 3*D*) to create the oriented edge polarity (OEP) variant of the model, which contains oriented C1 units sensitive to contrast polarity of edges. This model has twice as many C1 units available as inputs to the shape-template stage (S2) during the fitting procedure, but the rest of the fitting procedure remained the same. Thus the model was free to assign weights to either or both phases of a filter with a particular orientation and position in space. The resulting fits of the OEP model for the example neurons are shown in Fig. 9, *E–H*, training on both filled and outline stimuli. For *neuron 2*, the OEP model was better able to capture the lack of correlation ($r = 0.4$, $P < 0.001$) and the compressed response range for outlines (Fig. 9*F*; cf. Fig. 5*B*). Similarly, this phase-preserving model provided a better account of the responses of *neuron 3* (Fig. 9*G*; cf. Fig. 5*C*). For *neuron 4*, the OEP model matched the range of responses but predicted a higher correlation between responses to filled and outline stimuli (Fig. 9*H*; cf. Fig. 5*D*). This observation was similar to the HMax model fit for this neuron (Fig. 9*D*). Both HMax and OEP models provided a poor fit to the observed data ($r = 0.3$ and $r = 0.32$, respectively). On average across the population, the OEP model was significantly better than the original HMax model at fitting the data (Wilcoxon signed-rank test, $P < 0.001$; Fig. 10*A*). The OEP model also used significantly fewer C1 units to fit the neuronal responses than HMax (Mann-Whitney *U*-test, $P = 0.019$).

As an alternative to preserving phase information in the orientation-tuned C1 channels, we modified the HMax model by introducing a new set of unoriented DoG S1 filters that drove a set of phase-pooling C1 channels (Fig. 3*E*). We reasoned that this would allow the shape template implemented by the S2 unit to detect the interior of the object separately from the oriented boundary, giving the model an ability to discriminate filled and outline stimuli. We continued to pool across phase for unoriented filters; thus, like the HMax model, this model was phase independent. The responses of this modified model, which we refer to as the unoriented interior fill (UIF) model, to filled and outline shapes are shown for *neurons 1–4* in Fig. 9, *I–L*, for the best-fitting models, training on both filled and outline stimuli. In comparison to the original HMax model, access to unoriented units allowed the model to better capture the bias of *neuron 2* for filled shapes (Fig. 9*J*; cf. Fig. 5*B*) and the bias of *neuron 3* for outlines (Fig. 9*K*; cf. Fig. 5*C*), while substantially reducing the correlation between filled and outline shapes. For the poorly curvature-tuned *neuron 4*, the UIF model was similar to the other models in producing a poor fit ($r = 0.27$) and failing to capture the low correlation between filled and outline responses (Fig. 9*L*; cf. Fig. 5*D*).

Across the population of 43 neurons, like the OEP model, the UIF model fit the data significantly better than the original HMax model (Wilcoxon signed-rank test, $P < 0.001$; Fig. 10*A*) and also used significantly fewer C1 units to fit the neuronal responses than HMax (Mann-Whitney *U*-test, $P = 0.046$). The UIF model significantly outperformed the OEP model in terms of having higher *r* values for fitting neuronal responses (paired *t*-test, $P = 0.001$; Fig. 10*A*). These metrics are for model fits to all data (train and test, see METHODS), but performance computed on the test set alone (Fig. 10*B*) was consistent with Fig. 10*A*: both modified models still fit the data significantly better than the HMax model ($P < 0.001$ for both).

In evaluating the models, we also considered the explainable variance in the data (Yamins et al. 2014). Specifically, we normalized r^2 values by the reliability of the measured responses, calculated as the Spearman-Brown-corrected correlation of half-split responses (see METHODS). Explained variance computed with this metric was still higher for the OEP and UIF models than for the original HMax model (HMax: 0.15–0.80, median = 0.47; OEP: 0.32–1.18, median = 0.69; UIF: 0.25–1.04, median = 0.83). For 40/43 and 39/43 neurons, respectively, the explained variance values for the OEP and UIF models were greater than those for the HMax mode. For 30/43

Fig. 9. Model comparison for example neurons. *A–D*: predicted responses to filled and outline stimuli for hierarchical-max (HMax) model trained on responses to both stimulus types (cf. Fig. 5, *A–D*, for corresponding scatterplots based on observed responses). For example *neurons 1–3*, the best HMax model (even-symmetric S1 filters for *A* and *C*, odd-symmetric for *B*) predicted responses that were reasonably well correlated with observed data: $r_{fit} = 0.76$ (*A*), 0.57 (*B*), 0.80 (*C*). The fit was poor for *neuron 4* (*D*): $r_{fit} = 0.3$ (even-symmetric S1 filter). The HMax model failed to capture the poor correlation between filled and outline responses for *neurons 2–4* (compare *r* values listed in the panels with those in Fig. 5, *B–D*) and the compressed range of outline responses for *neurons 1* and *2* (compare ρ values listed in the panels with Fig. 5, *B* and *C*). Model S2 templates included 23 C1 units (*A*), 19 C1 units (*B*), 21 C1 units (*C*), and 15 units (*D*). *E–H*: predicted responses based on the oriented edge polarity (OEP) model. Predicted responses were well correlated with observed responses, except for *neuron 4*: $r_{fit} = 0.76$ (*E*), 0.79 (*F*), 0.83 (*G*), 0.32 (*H*). Fill/outline response correlations and relative response ranges for this model (listed in the panels) were comparable to the values based on observed data, except fill-outline correlations for *neuron 4* (see Fig. 5, *A–D*). OEP models presented here were based on even-symmetric S1 filters and a template with 24 C1 units (*E*); odd-symmetric S1 filters and a template with 24 C1 units (*F*); odd-symmetric S1 filters and a template with 23 C1 units (*G*); and even-symmetric S1 filters and a template with 4 C1 units (*H*). *I–L*: predicted responses based on unoriented interior fill (UIF) model. Here again, predicted and observed responses were well correlated, except for *neuron 4*: $r_{fit} = 0.74$ (*I*), 0.76 (*J*), 0.84 (*K*), 0.27 (*L*). Fill/outline response correlations and relative response ranges for this model (listed in the panels) were comparable to the values based on observed data, except fill/outline correlations for *neuron 4*. UIF models presented here were based on odd-symmetric S1 filters and a template with 24 C1 units (*I*); even-symmetric S1 filters and a template with 25 C1 units (*J*); odd-symmetric S1 filters and a template with 12 C1 units (*K*); and even-symmetric S1 filters and a template with 4 C1 units (*L*). All *r* values reported here are statistically significant at $P < 0.01$.

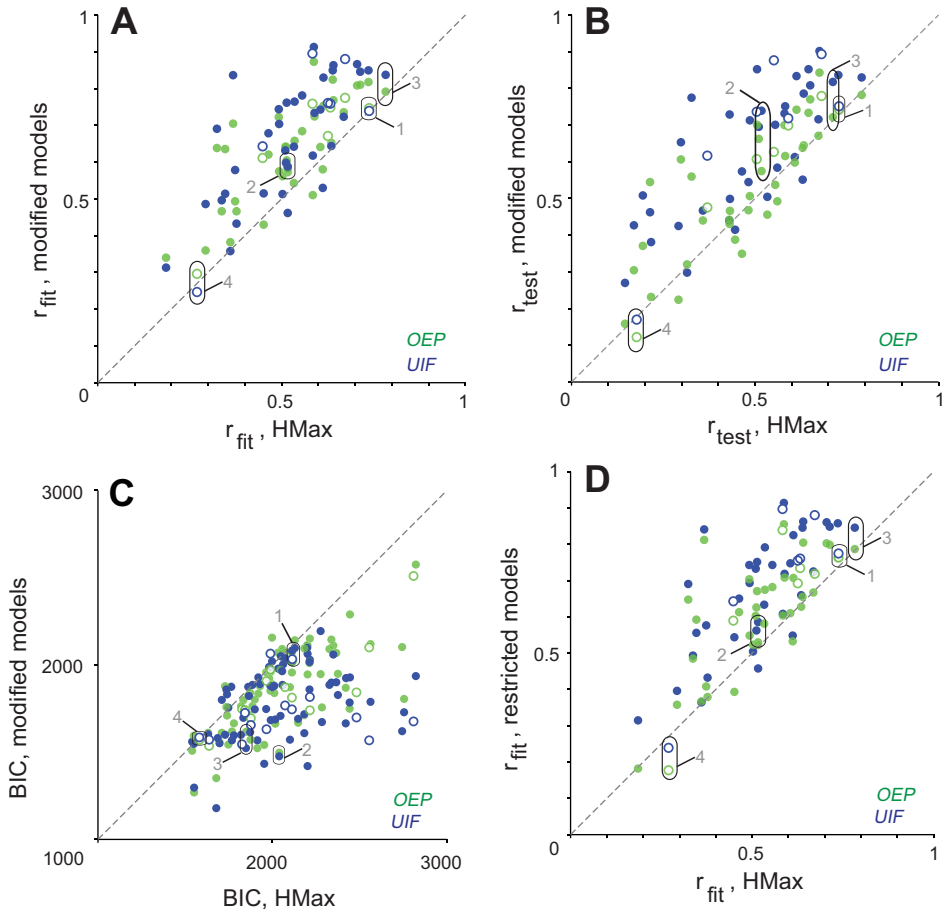


Fig. 10. Population model results. *A*: correlations between predicted and observed responses, r_{fit}, for hierarchical-max (HMax) model fits (*x*-axis; 1 model per neuron, values reported are averaged across even- and odd-symmetric S1 filters) are plotted against corresponding values for oriented edge polarity (OEP) and unoriented interior fill (UIF) models. Correlation values were significantly greater for OEP and UIF models compared with HMax ($P < 0.001$); values for UIF were also significantly greater than for OEP model ($P = 0.001$). Here, and in all other panels, open and filled symbols indicate fits to data from *animals M1* and *M2*, respectively, and values for example *neurons 1–4* are indicated (grouped by ellipses). *B*: as in *A*, but showing correlations between predicted and observed responses computed for the cross-validation test set only, r_{test}. Correlation values were significantly greater for OEP and UIF models compared with HMax ($P = 0.005$ and $P < 0.001$, respectively). *C*: Bayesian information criterion (BIC) values for HMax model fits (*x*-axis; 2 models per neuron; 1 model is indicated for example *neurons 1–4*) are plotted against the corresponding values for OEP and UIF models (*y*-axis). The modified models produced significantly lower BIC values compared with the HMax model ($P < 0.001$). *D*: r_{fit} for HMax model fits (*x*-axis) and the restricted OEP and UIF models (*y*-axis), which had the same number of C1 units as the HMax model. Correlation values were significantly greater for restricted OEP and UIF models compared with HMax ($P < 0.001$ for both restricted models vs. HMax).

neurons, these values were greater for the UIF model than for the OEP model.

We performed two additional analyses to verify that the increased performance of our modified models did not result from the increase in number of distinct C1 units available in these models. First, we compared BIC associated with each model (1 model per neuron; see METHODS for details). The modified models produced significantly lower BIC values compared with the HMax model (Fig. 10*C*; Wilcoxon signed-rank test, $P = 0.02$ and $P < 0.001$ for OEP and UIF, respectively). For 41/43 neurons and 36/43 neurons, the BIC values for the OEP and UIF models, respectively, were lower than the same values for the HMax model. We obtained the same results by using AIC instead of BIC to compare model performance. Next, we restricted the number of C1 units available to the modified models to match that in the HMax model (see METHODS). Across the population, even these restricted models produced significantly better fits to the data than the HMax model (Fig. 10*D*; cf. Fig. 10*A*). Together, these results indicate that the improved performance of our models cannot be explained by the number of C1 units available but basic functional changes are able to make the HMax model predictions qualitatively and quantitatively more consistent with observed responses in area V4.

In designing the OEP and UIF models, we attempted to explicitly represent information about interior fill within the HMax framework. The population-wide performance improvements of the OEP and UIF models suggest that this strategy was successful. Figure 11*A* shows the correlation between fill/outline shape responses for each of the models plotted against the same metric for neurons. The OEP model showed the best correspondence to neuronal data in this comparison in two respects. First, the correlation between fill and outline *r* values was highest between the model and the data ($r = 0.72$, $P < 0.001$, *z*-transformed values), whereas the HMax and UIF model values were less correlated with the neuronal data (HMax: $r = 0.54$, $P < 0.001$; UIF: $r = 0.57$, $P < 0.001$). Second, correlations between filled and outline stimuli for HMax and UIF models were typically much larger than those observed in the neuronal data (HMax and UIF points fall overwhelmingly above the line of identity), whereas those for the OEP model were better matched to the range of values observed for the neurons. This is quantified in Fig. 11*B*, which shows that the differences between the neuronal data and models in terms of fill-outline correlation were centered near 0 (mean $= -0.08$) for the OEP model but were more negative for the HMax and UIF models (mean $= -0.47$ and -0.32, respectively), i.e., the HMax and UIF models overestimate the fill-outline correlation. Thus OEP was better than HMax and UIF in matching the correlations between filled and outline stimuli observed in the data.

Another critical limitation of the original HMax model was its inability to capture the relative ranges of responses to filled and outline stimuli as quantified by ρ (Fig. 12*A*). In particular, for neurons that responded poorly to outline stimuli, i.e., when $\rho \ll 1$, the ρ of responses predicted by the HMax model consistently fell well above the line of equality in Fig. 12*A*. Although HMax overestimated the relative response range of

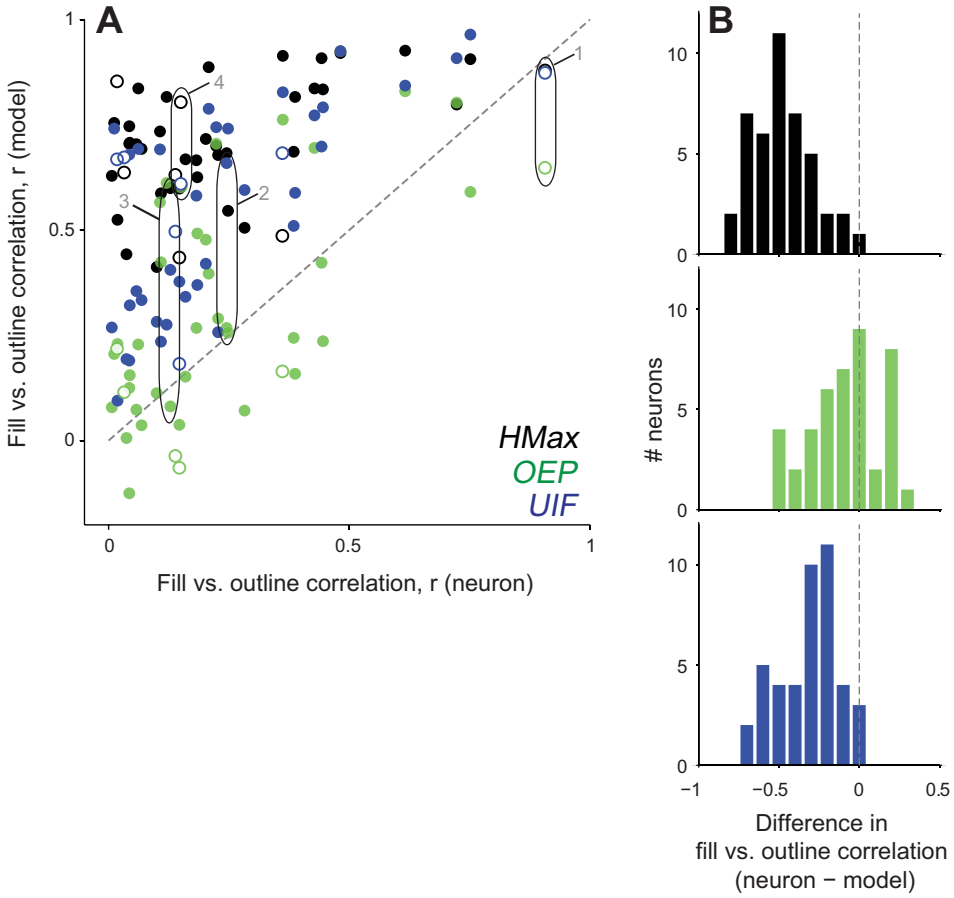


Fig. 11. Correlation between responses to filled and outline stimuli in models and data. *A*: correlations between the responses to filled and outline stimuli predicted by the models (*y*-axis) are plotted against the same correlations observed in the neuronal data (*x*-axis). Correlations for the hierarchical-max (HMax) and unoriented interior fill (UIF) models deviated more from the neuronal data than correlations for the oriented edge polarity (OEP) model. Open and filled symbols indicate fits to data from *animals M1* and *M2*, respectively. Values for example *neurons 1–4* are indicated (grouped by ellipses). *B*: distributions of differences between observed and predicted correlations are shown for each model; color coding as in *A*. Mean difference from observed data: −0.47 (HMax), −0.08 (OEP), −0.32 (UIF). Dashed vertical lines indicate no difference from observed data.

these neurons, the OEP model (Fig. 12*B*) tended to underestimate it. To understand whether the misestimation in the two models involved the same neurons, we compared the differences between model values of ρ and observed values of ρ. The magnitude of deviations from observed ρ was similar for OEP and HMax (Wilcoxon signed-rank test, $P = 0.94$), suggesting that the OEP model underestimates ρ for the same neurons whose response ranges are overestimated by HMax (compare Fig. 12, *A* and *B* vertically). In contrast, the UIF model was better able to match the relative ranges of responses for filled and outline stimuli (Fig. 12*B*).

Finally, we examined whether the property of TI for shape selectivity, which the HMax model was designed to capture, was maintained in our OEP and UIF modifications. Although we did not measure TI for neurons because of the large size of our stimulus set, it has been documented as an important feature of V4 responses in previous work (El-Shamayleh and Pasupathy 2016; Gallant et al. 1993; Pasupathy and Connor 2001; Rust and Dicarlo 2010). Thus our model modifications were designed to leave intact the properties of the HMax model that impart TI (Max-pooling operations over spatial location). We confirmed this by computing TI for each neuron's model fit (see details in METHODS) and comparing this value across models. The OEP and HMax model fits had similar TI, with HMax more invariant (this difference was small but significant; Fig. 13; Wilcoxon signed-rank test, $P = 0.02$), whereas the UIF model fits were significantly more translation invariant than both of the other models (Fig. 13; Wilcoxon signed-rank test, $P < 0.001$).

In summary, we found two straightforward modifications to the HMax model for V4 shape selectivity that provided a better account for two key observations across our population of V4 neurons: a low correlation between responses to filled and outline shapes and a difference in dynamic range in responses to these two stimulus sets.

DISCUSSION

To deepen our understanding and improve models of midlevel form processing, we investigated how the interior fill of simple shapes influenced responses of form-selective V4 neurons. We found that very few neurons responded similarly to filled and outline stimuli. Instead, a majority of neurons responded strongly and selectively to one category of stimuli (filled or outline) but poorly to the other. V4 neurons that exhibited differential responses to filled and outline stimuli were also form selective, suggesting that the midlevel representation of form involves a complex combination of boundary and surface features. Our results demonstrate that a prevalent assumption in form-encoding models, that form selectivity is built from boundary orientation signals regardless of phase, is not appropriate for a majority of neurons in area V4. Additionally, we propose modifications to an existing V4 HMax model that capture the observed diversity in V4 responses. More broadly, our results provide insights into how object-centered representations may be built and how V4 responses may play a role in image segmentation.

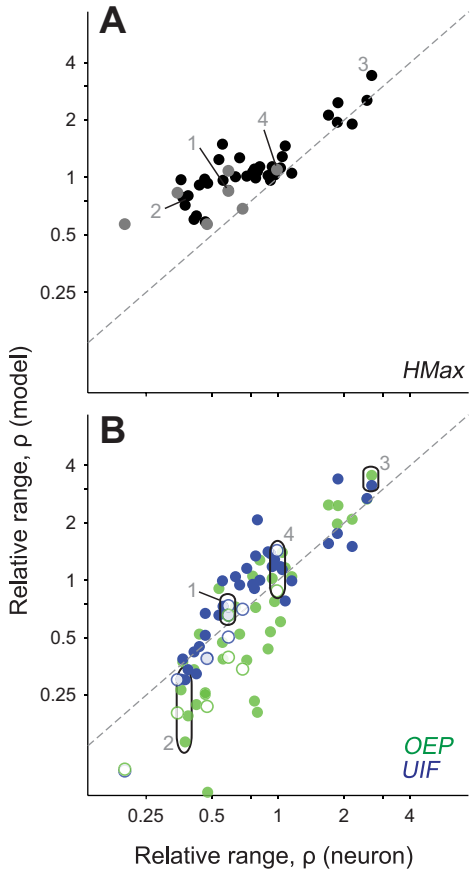


Fig. 12. Relative range of predicted vs. observed responses. Relative range of responses to filled vs. outline stimuli predicted by the hierarchical-max (HMax; *A*) and unoriented interior fill (UIF) and oriented edge polarity (OEP; *B*) models are plotted against relative response range (ρ) based on observed data. When observed ρ values are > 1, the HMax predictions (*A*) lie close to the diagonal. This was not the case when observed ρ was <1. ρ values based on OEP and UIF models closely track the line of equality, meaning responses predicted by these models match the lack of fill/outline invariance observed in the responses of many neurons. Values reported here are averaged across S1 phases (even- and odd-symmetric). ρ values for example *neurons 1–4* are indicated. Open (or gray) and filled (or black) symbols indicate fits to data from *animals M1* and *M2*, respectively.

Relation to Past Experimental Studies

Tuning for form and surface properties have typically been investigated independently in V4, with many past studies demonstrating sensitivity to visual form (Desimone and Schein 1987; Gallant et al. 1996; Kobatake and Tanaka 1994; Pasupathy and Connor 1999, 2001) and to surface characteristics, including luminance, chromaticity, and texture (Bushnell et al. 2011b; Conway et al. 2007; Goda et al. 2014; Namima et al. 2014; Okazawa et al. 2015; Schein and Desimone 1990). Because of the practical constraint of experimental time both attributes are seldom varied in the same experiment (but see Bushnell and Pasupathy 2012; Kobatake and Tanaka 1994), but such studies are required to determine how signals from earlier stages, which are sensitive to luminance, chromaticity (Gegenfurtner and Kiper 2003; Shapley and Hawken 2011), and more elaborate statistics (Freeman et al. 2013; Ziemba et al.. 2016), are pooled to give rise to form selectivity in V4. Our results document the strong influence of a surface characteristic, interior fill, on the responses of form-encoding V4 neurons.

We found that responses of only a minority of form-selective V4 neurons are invariant to interior fill. Responses of other neurons are better explained when surface information is also taken into account. Although this is consistent with some past findings in inferior temporal cortex (Ito et al. 1994; Kayaert et al. 2003), it is strikingly different from one previous study by Kovács et al. (2003) that reported weaker inferior temporal cortex responses but similar tuning for outlines compared with filled stimuli.

We also found that a majority of V4 neurons showed a larger dynamic range in their responses to filled stimuli. This is unlikely to arise from a loss of perceptibility of outlines at more eccentric locations: we verified in early experimental sessions that animals could discriminate outline shapes presented at the line widths and eccentricities in our recordings, and the preference for filled stimuli persisted even when we doubled outline widths. The larger range of responses evoked by filled stimuli suggests that these stimuli sample a larger region of the overall V4 representational space compared with outlines, but it remains unclear what image features underlie this difference (surface properties, SF, etc.). Our study specifically focuses on the encoding of a filled shape and its outline, and it is clear that these are typically not nearly equivalent in terms of single V4 neuron responses. To further characterize the role of surface properties in form encoding, future experiments should examine how surface luminance gradients and texture modulate V4 responses to boundary form.

Consistent with past findings, responses of many neurons in our data set were well fit by the boundary curvature (APC) model proposed by Pasupathy and Connor (2001). However, the median goodness of fit reported here is smaller than other data sets (e.g., El-Shamayleh and Pasupathy 2016) for several possible reasons: fewer stimulus repeats, smaller population sample, less strict inclusion criteria, and use of a 2D APC model (responses in area V4 are often dictated by adjoining

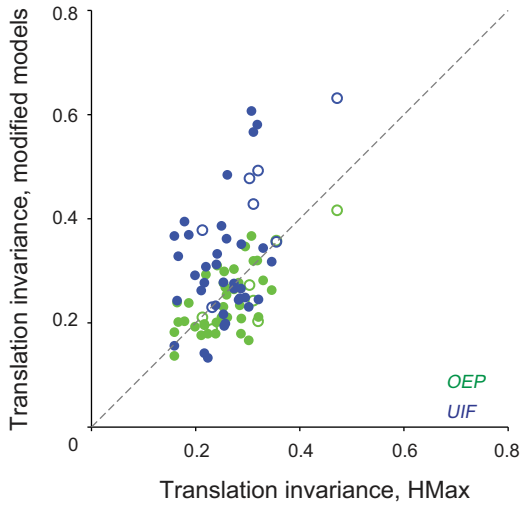


Fig. 13. Comparison of translation invariance (TI) between hierarchical-max (HMax) and modified models. TI values are shown for the HMax model fits (*x*-axis; values reported are averaged across even- and odd-symmetric filters for each neuron, except for one; see METHODS) plotted against TI values for fits of the modified models. TI was significantly better for the unoriented interior fill (UIF) model than for the other models ($P < 0.001$), whereas the oriented edge polarity (OEP) and HMax models exhibited a similar TI on average, HMax being slightly more invariant ($P = 0.02$). Open and filled symbols indicate fits to data from *animals M1* and *M2*, respectively.

curvatures, which would be better explained by a 4D model; see Pasupathy and Connor 2001). As in previous studies, our data set also included neurons whose responses were poorly explained by tuning for boundary curvature. Responses of these neurons can still capture important aspects of objects, e.g., surface properties. The existence of these neurons alongside curvature-tuned neurons underscores the diverse functions that could be supported by responses in area V4. Our results support the viewpoint that V4 neurons not only encode object boundaries but may be involved in other complex computations involving surface properties of objects, and thus contribute to object and scene recognition.

Building Models of the Ventral Stream

Research over the past several decades has yielded numerous models for how signals from the lateral geniculate nucleus are transformed into simple and complex cell response properties observed in V1 (for review see Carandini et al. 2005). However, it has been more difficult to capture the computational architecture for the transformation from V1 to V4. Past V4 models for form and texture encoding have relied on disparate strategies, focusing on boundary form (Cadieu et al. 2007; Rodriguez-Sánchez and Tsotsos 2012) and higher-order image statistics (Okazawa et al. 2015), respectively. The HMax model was developed and elaborated in a series of studies (Riesenhuber and Poggio 1999, 2000; Serre et al. 2005) and then later used by Cadieu et al. (2007) to successfully fit existing V4 boundary curvature data from Pasupathy and Connor (2001). Our findings suggest that fundamental parts of the model, such as building form selectivity from orientation-tuned inputs with low-bandwidth SFs and phase-destroying operations, may be inconsistent with our expanded V4 data. Importantly, by testing our models with a larger and more diverse data set, we have created a model that explains a richer set of V4 response properties. We examined modifications that affect the early layers (S1/C1), leaving the hierarchical structure intact. The guiding principle behind both of the modifications we made was to add relevant information that may support the emergence of shape that depends on interior fill and not just the boundary. The OEP and UIF models maintain the same level of form selectivity and TI as the original HMax model and at the same time successfully capture the diversity in V4 responses to filled and outline stimuli. Additionally, the modified models were able to achieve a better match to the data while using fewer C1 unit weights to fit the observed responses in area V4.

Our experimental results provide a key insight, that explicit surface contrast and form information is necessary to build accurate models of V4 responses. Because filled and outline stimuli differ in SF content, modulation of form responses could theoretically be implemented by accessing SF information. For example, the spectral receptive field model (David et al. 2006) could differentiate between filled and outline stimuli, even though it cannot capture V4 tuning for curvature (Oleskiw et al. 2014). The HMax model could create different responses to filled and outline stimuli by varying the use of inputs located in the center of the field and the relative use of inputs with different SF. For example *neuron 3*, the best-fitting HMax model contained strong negative (suppressive) weights on small C1 units in the center of the field while using moderate weights further away from the center to gain shape selectivity. For example *neuron 2*, the best-fitting HMax model had all positive weights (including in the center of the field), and the C1 units with the highest weights had lower SF. Despite this, the HMax model had limited success in fitting the data in comparison with our modified models. We believe that the HMax model was unable to leverage all the available SF information in the stimulus because of the limited SF range ($<$1 octave, from 1.0 to 1.9 cyc/°) of its S1 filters; adding more low-SF oriented units to the HMax model produced results that better matched neuronal data (not shown). This limitation of the HMax is consistent with the success of the UIF model, since unoriented units provide a source of low-SF information. However, the OEP model has the same SF range as the original HMax, yet it is better able to capture the differential responses to filled and outline stimuli observed in V4. Thus although a limited SF range can contribute to an inability to capture the differences in responses to filled and outline stimuli, it is not the sole factor distinguishing filled and outline stimuli. Lack of phase information could also explain why the HMax model was unable to produce different responses for these stimuli. However, the UIF model was better able to match the observed responses despite also lacking phase information. Overall, information about interior fill appears to be important in distinguishing filled and outline stimuli.

Improvements to the HMax Model

To provide model units with direct access to surface fill information, we tried two strategies: preserving phase information in the C1 units of the model, akin to V1 simple cells, and including units with unoriented spatially opponent (center surround) RFs in the S1 layer. Both strategies were more successful than HMax at capturing the diversity of responses to filled and outline stimuli in our V4 data, and they exhibit complementary strengths in their ability to explain other V4 response properties. Specifically, the addition of unoriented filters in the UIF model may facilitate the encoding of shape features in an object-centered reference frame. Previous neurophysiological studies on the position- and size-invariant encoding of contour segments in V4 (El-Shamayleh and Pasupathy 2016; Pasupathy and Connor 2001) suggest that V4 neurons encode parts of a shape boundary in an object-centered reference frame. However, the HMax model, which builds V4 selectivity by combining oriented boundary components into a contour template, is inconsistent with object-centered shape tuning. In the UIF model, if the unoriented filters reflect the interior fill of the stimulus, then the position of the contour template relative to the position of the unoriented filters could represent object-centered encoding. In this regard, the UIF model is consistent with the concept of "grouping cells" for figure/ground segregation (Craft et al. 2007). Although the OEP model is not consistent with object-centered encoding, the phase preservation at the C1 level could bestow the ability to reproduce V4 neuronal preference for different luminance contrasts (Bushnell et al. 2011b; Conway et al. 2007; Namima et al. 2014; Schein and Desimone 1990). Finally, the dissociation between the representation of contour and surface characteristics in the UIF model, but not the OEP model, makes it conceptually more consistent with models of scene perception where surface characteristics are filled into regions defined by

boundaries (Grossberg and Mingolla 1985; Pinna and Grossberg 2005). From our results, the OEP model better captures the correlations between filled and outline responses (and underestimates some relative ranges), whereas the UIF model better captures the relative ranges of responses to these stimuli (but overestimates correlations between responses to filled and outline stimuli). Given the above complementary strengths of the UIF and OEP models, it is conceivable that both strategies of preserving information about interior fill may be instantiated across the V4 population. Further experiments could confirm the role phase sensitivity or unoriented/low SF inputs might play in subserving object representations in area V4.

Both of these modifications also appear to be biologically plausible. The OEP model requires phase-sensitive signals to be available downstream from V1 complex cells. Our particular implementation uses units that are sensitive to edge polarity but generalize over spatial position, a concept consistent with the framework described by Hubel and Livingstone (1985) and consistent with observations that V2 neurons have signals related to border ownership (Zhou et al. 2000). There is abundant evidence that signals leaving V1, and signals present in V2, have spatial phase information (Ito and Komatsu 2004; Levitt et al. 1994; Mechler and Ringach 2002). Thus our modifications that include units with phase dependence are no less biologically plausible than the original HMax model that lacked phase-dependent signals at the C1 stage.

For the UIF model, unoriented information would need to be available at a stage within the model downstream from V1 complex cells (e.g., at the S2 stage of the HMax model). Such units are known to exist in V2: the "complex-unoriented" units of Hubel and Livingstone (1985) and the "spot cells" of Baizer et al. (1977) are important classes of neurons that have received less attention than their orientation-tuned counterparts. The major conceptual conclusion from Hubel and Livingstone (1985) is that the V1-to-V2 transformation could be one of spatial generalization, with "or-gates" acting across spatially offset but similar RFs. This concept underlies our modifications to the model, except that we no longer limit the features that are being generalized to the narrower class of orientation-tuned and perfectly phase-insensitive complex-cell units.

Implications for Object Recognition and Perception

Outline stimuli are an important, perceptually relevant stimulus class (Hochberg and Brooks 1962; Hodgson et al. 2010), but our results suggest that many neurons are poorly driven by these stimuli compared with their filled counterparts and appear not to be shape selective when tested with these outlines. Although this may appear to conflict with the human facility at recognizing outlines, it is important to note that most V4 neurons that responded well to outlines were within 4° of the fovea, the region of the visual field most relevant for reading and recognition (Latham and Whitaker 1996; Rosenholtz 2016). Because only a small proportion of V4 neurons responded similarly to filled and outline stimuli, our results support the hypothesis that behavioral learning underlies equivalency in recognition observed between line drawings and photographs (Close and Call 2015; Tanaka 2007). Moreover, although humans are good at recognizing isolated objects from their outlines, clutter makes outline-based recognition hard; in such cases, surface characteristics are critical for recognition (see Fig. 1). Thus the preponderance of neurons that respond differentially to filled shapes may reflect the importance of surface characteristics in interpreting visual scenes with clutter.

In natural vision, the information that reaches our eyes must be parsed into a meaningful arrangement of regions and objects. This process, called image segmentation, is one of the most challenging computations accomplished by the visual system. Computer vision and psychophysical studies suggest that segmentation is most effective when both contour- and region-based properties are considered (Hansen and Gegenfurtner 2007; Leung and Malik 1998; Mel 1997; Mumford et al. 1987; von der Heydt and Pierson 2006). Our results, demonstrating the joint encoding of surface and boundary characteristics in a majority of V4 neurons, may reflect this algorithmic preference. Thus these V4 neurons may play a role not just in form encoding but also in scene parsing.

ACKNOWLEDGMENTS

We thank Amber Fyall for expert animal care. We are also grateful to Gregory Horwitz and Yasmine El-Shamayleh for insightful comments and discussions.

GRANTS

This work was funded by National Institutes of Health (NIH) Grant R01 EY-018839 to A. Pasupathy; National Science Foundation (NSF) CRCNS Grant IIS-1309725 to A. Pasupathy and W. Bair.; NIH Center Core Grant for Vision Research P30 EY-01730 to the University of Washington; NIH/ORIP Grant P51 OD-010425 to the Washington National Primate Research Center; and University of Washington Vision Training Grant (NIH T32 EY-007031), NSF GRFP DGE-1256082, and University of Washington Computational Neuroscience Training Grant (NIH T90 DA-032436) to D. V. Popovkina.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

D.V.P. and A.P. conceived and designed research; D.V.P. performed experiments; D.V.P. and W.B. analyzed data; D.V.P., W.B., and A.P. interpreted results of experiments; D.V.P. prepared figures; D.V.P. and A.P. drafted manuscript; D.V.P., W.B., and A.P. edited and revised manuscript; D.V.P., W.B., and A.P. approved final version of manuscript.

REFERENCES

Baizer JS, Robinson DL, Dow BM. Visual responses of area 18 neurons in awake, behaving monkey. *J Neurophysiol* 40: 1024–1037, 1977. doi:10.1152/jn.1977.40.5.1024.

Biederman I, Ju G. Surface versus edge-based determinants of visual recognition. *Cogn Psychol* 20: 38–64, 1988. doi:10.1016/0010-0285(88)90024-2.

Bushnell BN, Harding PJ, Kosai Y, Bair W, Pasupathy A. Equiluminance cells in visual cortical area v4. *J Neurosci* 31: 12398–12412, 2011a. doi:10.1523/JNEUROSCI.1890-11.2011.

Bushnell BN, Harding PJ, Kosai Y, Pasupathy A. Partial occlusion modulates contour-based shape encoding in primate area V4. *J Neurosci* 31: 4012–4024, 2011b. doi:10.1523/JNEUROSCI.4766-10.2011.

Bushnell BN, Pasupathy A. Shape encoding consistency across colors in primate V4. *J Neurophysiol* 108: 1299–1308, 2012. doi:10.1152/jn.01063.2011.

Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T. A model of V4 shape selectivity and invariance. *J Neurophysiol* 98: 1733–1750, 2007. doi:10.1152/jn.01265.2006.

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC. Do we know what the early visual system does? *J Neurosci* 25: 10577–10597, 2005. doi:10.1523/JNEUROSCI.3726-05.2005.

Cavanagh P. What's up in top-down processing. In: *Representations of Vision*, edited by Gorei A. Cambridge, UK: Cambridge University Press, 1991, p. 295–304.

Close J, Call J. From colour photographs to black-and-white line drawings: an assessment of chimpanzees' (*Pan troglodytes*') transfer behaviour. *Anim Cogn* 18: 437–449, 2015. doi:10.1007/s10071-014-0813-5.

Conway BR, Moeller S, Tsao DY. Specialized color modules in macaque extrastriate cortex. *Neuron* 56: 560–573, 2007. doi:10.1016/j.neuron.2007.10.008.

Craft E, Schütze H, Niebur E, von der Heydt R. A neural model of figure-ground organization. *J Neurophysiol* 97: 4310–4326, 2007. doi:10.1152/jn.00203.2007.

David SV, Hayden BY, Gallant JL. Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96: 3492–3505, 2006. doi:10.1152/jn.00575.2006.

De Valois RL, Albrecht DG, Thorell LG. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res* 22: 545–559, 1982. doi:10.1016/0042-6989(82)90113-4.

De Weerd P, Desimone R, Ungerleider LG. Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques. *Vis Neurosci* 13: 529–538, 1996. doi:10.1017/S0952523800008208.

Desimone R, Schein SJ. Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol* 57: 835–868, 1987. doi:10.1152/jn.1987.57.3.835.

El-Shamayleh Y, Pasupathy A. Contour curvature as an invariant code for objects in visual area V4. *J Neurosci* 36: 5532–5543, 2016. doi:10.1523/JNEUROSCI.4139-15.2016.

Foster KH, Gaska JP, Nagler M, Pollen DA. Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol* 365: 331–363, 1985. doi:10.1113/jphysiol.1985.sp015776.

Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA. A functional and perceptual signature of the second visual area in primates. *Nat Neurosci* 16: 974–981, 2013. doi:10.1038/nn.3402.

Gallant JL, Braun J, Van Essen DC. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259: 100–103, 1993. doi:10.1126/science.8418487.

Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76: 2718–2739, 1996. doi:10.1152/jn.1996.76.4.2718.

Gallant JL, Shoup RE, Mazer JA. A human extrastriate area functionally homologous to macaque V4. *Neuron* 27: 227–235, 2000. doi:10.1016/S0896-6273(00)00032-5.

Gattass R, Sousa AP, Gross CG. Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8: 1831–1845, 1988. doi:10.1523/JNEUROSCI.08-06-01831.1988.

Gegenfurtner KR, Kiper DC. Color vision. *Annu Rev Neurosci* 26: 181–206, 2003. doi:10.1146/annurev.neuro.26.041002.131116.

Goda N, Tachibana A, Okazawa G, Komatsu H. Representation of the material properties of objects in the visual cortex of nonhuman primates. *J Neurosci* 34: 2660–2673, 2014. doi:10.1523/JNEUROSCI.2593-13.2014.

Grossberg S, Mingolla E. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol Rev* 92: 173–211, 1985. doi:10.1037/0033-295X.92.2.173.

Hansen T, Gegenfurtner KR. Higher order color mechanisms for image segmentation. In: *Advances in Brain, Vision, and Artificial Intelligence: Second International Symposium, BVAI 2007, Naples, Italy, October 10–12, 2007, Proceedings*, edited by Mele F, Ramella G, Santillo S, Ventriglia F. Berlin: Springer, 2007, p. 72–83.

Heywood CA, Cowey A. On the role of cortical area V4 in the discrimination of hue and pattern in macaque monkeys. *J Neurosci* 7: 2601–2617, 1987. doi:10.1523/JNEUROSCI.07-09-02601.1987.

Hochberg J, Brooks V. Pictorial recognition as an unlearned ability: a study of one child's performance. *Am J Psychol* 75: 624–628, 1962. doi:10.2307/1420286.

Hodgson E, Allen M, Bloom P. Do young children know what makes a picture useful to other people? *J Cogn Cult* 10: 27–37, 2010. doi:10.1163/156853710X497158.

Hubel DH, Livingstone MS. Complex-unoriented cells in a subregion of primate area 18. *Nature* 315: 325–327, 1985. doi:10.1038/315325a0.

Ito M, Fujita I, Tamura H, Tanaka K. Processing of contrast polarity of visual images in inferotemporal cortex of the macaque monkey. *Cereb Cortex* 4: 499–508, 1994. doi:10.1093/cercor/4.5.499.

Ito M, Komatsu H. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J Neurosci* 24: 3313–3324, 2004. doi:10.1523/JNEUROSCI.4364-03.2004.

Kayaert G, Biederman I, Vogels R. Shape tuning in macaque inferior temporal cortex. *J Neurosci* 23: 3016–3027, 2003. doi:10.1523/JNEUROSCI.23-07-03016.2003.

Kobatake E, Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994. doi:10.1152/jn.1994.71.3.856.

Kosai Y, El-Shamayleh Y, Fyall AM, Pasupathy A. The role of visual area V4 in the discrimination of partially occluded shapes. *J Neurosci* 34: 8570–8584, 2014. doi:10.1523/JNEUROSCI.1375-14.2014.

Kovács G, Sáry G, Köteles K, Chadaide Z, Tompa T, Vogels R, Benedek G. Effects of surface cues on macaque inferior temporal cortical responses. *Cereb Cortex* 13: 178–188, 2003. doi:10.1093/cercor/13.2.178.

Latham K, Whitaker D. A comparison of word recognition and reading performance in foveal and peripheral vision. *Vision Res* 36: 2665–2674, 1996. doi:10.1016/0042-6989(96)00022-3.

Leung T, Malik J. Contour continuity in region based image segmentation. In: *Computer Vision—ECCV'98: 5th European Conference on Computer Vision, Freiburg, Germany, June, 2–6, 1998 Proceedings*, edited by Burkhardt H, Neumann B. Berlin: Springer, 1998, vol. I, p. 544–559.

Levitt JB, Kiper DC, Movshon JA. Receptive fields and functional architecture of macaque V2. *J Neurophysiol* 71: 2517–2542, 1994. doi:10.1152/jn.1994.71.6.2517.

Mazer J. *pype3*. 2013. https://github.com/mazerj/pype3.

Mechler F, Ringach DL. On the classification of simple and complex cells. *Vision Res* 42: 1017–1033, 2002. doi:10.1016/S0042-6989(02)00025-1.

Mel BW. SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput* 9: 777–804, 1997. doi:10.1162/neco.1997.9.4.777.

Merigan WH. Basic visual capacities and shape discrimination after lesions of extrastriate area V4 in macaques. *Vis Neurosci* 13: 51–60, 1996. doi:10.1017/S0952523800007124.

Merigan WH, Pham HA. V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. *Vis Neurosci* 15: 359–367, 1998. doi:10.1017/S0952523898152112.

Mumford D, Kosslyn SM, Hillger LA, Herrnstein RJ. Discriminating figure from ground: the role of edge detection and region growing. *Proc Natl Acad Sci USA* 84: 7354–7358, 1987. doi:10.1073/pnas.84.20.7354.

Namima T, Yasuda M, Banno T, Okazawa G, Komatsu H. Effects of luminance contrast on the color selectivity of neurons in the macaque area v4 and inferior temporal cortex. *J Neurosci* 34: 14934–14947, 2014. doi:10.1523/JNEUROSCI.2289-14.2014.

Okazawa G, Tajima S, Komatsu H. Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc Natl Acad Sci USA* 112: E351–E360, 2015. doi:10.1073/pnas.1415146112.

Oleskiw TD, Pasupathy A, Bair W. Spectral receptive fields do not explain tuning for boundary curvature in V4. *J Neurophysiol* 112: 2114–2122, 2014. doi:10.1152/jn.00250.2014.

Pasupathy A, Connor CE. Responses to contour features in macaque area V4. *J Neurophysiol* 82: 2490–2502, 1999. doi:10.1152/jn.1999.82.5.2490.

Pasupathy A, Connor CE. Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86: 2505–2519, 2001. doi:10.1152/jn.2001.86.5.2505.

Pasupathy A, El-Shamayleh Y, Popovkina D. Visual shape and object perception. In: *Oxford Research Encyclopedia—Neuroscience*. New York: Oxford Univ. Press, 2018.

Pinna B, Grossberg S. The watercolor illusion and neon color spreading: a unified analysis of new cases and neural mechanisms. *J Opt Soc Am A Opt Image Sci Vis* 22: 2207–2221, 2005. doi:10.1364/JOSAA.22.002207.

Pospisil D, Pasupathy A, Bair W. *Evaluating and interpreting a convolutional neural net as a model of V4. MODVIS Workshop* (Online). West Lafayette, IN: Purdue e-Pubs, 2017. https://docs.lib.purdue.edu/modvis/2017/session01/4.

Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019–1025, 1999. doi:10.1038/14819.

Riesenhuber M, Poggio T. Models of object recognition. *Nat Neurosci* 3, *Suppl*: 1199–1204, 2000. doi:10.1038/81479.

Rodríguez-Sánchez AJ, Tsotsos JK. The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. *PLoS One* 7: e42058, 2012. doi:10.1371/journal.pone.0042058.

Rosenholtz R. Capabilities and limitations of peripheral vision. *Annu Rev Vis Sci* 2: 437–457, 2016. doi:10.1146/annurev-vision-082114-035733.

Rust NC, Dicarlo JJ. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30: 12978–12995, 2010. doi:10.1523/JNEUROSCI.0179-10.2010.

Sayim B, Cavanagh P. What line drawings reveal about the visual brain. *Front Hum Neurosci* 5: 118, 2011. doi:10.3389/fnhum.2011.00118.

Schein SJ, Desimone R. Spectral properties of V4 neurons in the macaque. *J Neurosci* 10: 3369–3389, 1990. doi:10.1523/JNEUROSCI.10-10-03369.1990.

Schiller PH. Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature* 376: 342–344, 1995. doi:10.1038/376342a0.

Schiller PH, Lee K. The role of the primate extrastriate area V4 in vision. *Science* 251: 1251–1253, 1991. doi:10.1126/science.2006413.

Serre T, Wolf L, Poggio T. Object recognition with features inspired by visual cortex. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, p. 994–1000.

Shapley R, Hawken MJ. Color in the cortex: single- and double-opponent cells. *Vision Res* 51: 701–717, 2011. doi:10.1016/j.visres.2011.02.012.

Tanaka M. Recognition of pictorial representations by chimpanzees (*Pan troglodytes*). *Anim Cogn* 10: 169–179, 2007. doi:10.1007/s10071-006-0056-1.

Tootell RB, Silverman MS, Hamilton SL, Switkes E, De Valois RL. Functional anatomy of macaque striate cortex. V. Spatial frequency. *J Neurosci* 8: 1610–1624, 1988. doi:10.1523/JNEUROSCI.08-05-01610.1988.

Truppa V, Spinozzi G, Stegagno T, Fagot J. Picture processing in tufted capuchin monkeys (*Cebus apella*). *Behav Processes* 82: 140–152, 2009. doi:10.1016/j.beproc.2009.05.004.

von der Heydt R, Pierson R. Dissociation of color and figure-ground effects in the watercolor illusion. *Spat Vis* 19: 323–340, 2006. doi:10.1163/156856806776923416.

Wagemans J, De Winter J, Op de Beeck H, Ploeger A, Beckers T, Vanroose P. Identification of everyday objects on the basis of silhouette and outline versions. *Perception* 37: 207–244, 2008. doi:10.1068/p5825.

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 111: 8619–8624, 2014. doi:10.1073/pnas.1403112111.

Yonas A, Arterberry ME. Infants perceive spatial structure specified by line junctions. *Perception* 23: 1427–1435, 1994. doi:10.1068/p231427.

Zhou H, Friedman HS, von der Heydt R. Coding of border ownership in monkey visual cortex. *J Neurosci* 20: 6594–6611, 2000. doi:10.1523/JNEUROSCI.20-17-06594.2000.

Ziemba CM, Freeman J, Movshon JA, Simoncelli EP. Selectivity and tolerance for visual texture in macaque V2. *Proc Natl Acad Sci USA* 113: E3140–E3149, 2016. doi:10.1073/pnas.1510847113.

Zimmerman RR, Hochberg J. Responses of infant monkeys to pictorial representations of a learned visual discrimination. *Psychon Sci* 18: 307–308, 1970. doi:10.3758/BF03331841.