Normalization by orientation anisotropy in human V1-V3

Zeming Fang[1,2], Catherine Olsson[3], Wei Ji Ma[1], Jonathan Winawer[1]

1. Department of Psychology and Center for Neural Science, New York University
2. Current affiliation: Cognitive Science, Rensselaer Polytechnic Institute
3. Anthropic, San Francisco, CA

Corresponding author: Zeming Fang, fangz5@rpi.edu

Abstract

An influential account of neuronal responses in primary visual cortex is the normalized energy model. This model is often implemented as a two-stage computation. The first stage is the extraction of contrast energy, whereby a complex cell computes the squared and summed outputs of a pair of linear filters in quadrature phase. The second stage is normalization, in which a local population of complex cells mutually inhibit one another. Because the population includes cells tuned to a range of orientations and spatial frequencies, the result is that the responses are effectively normalized by the local stimulus contrast. Here, using evidence from human functional MRI, we show that the classical model fails to account for the relative responses to two classes of stimuli: straight, parallel, band-passed contours (*gratings*), and curved, band-passed contours (*snakes*). The snakes elicit fMRI responses that are about twice as large as the gratings, yet traditional energy models, including normalized energy models, predict responses that are about the same. Here, we propose a computational model, in which responses are normalized not by the sum of the contrast energy, but by the orientation anisotropy, computed as the variance in contrast energy across orientation channels. We first show that this model accounts for differential responses to these two classes of stimuli. We then show that the model successfully generalizes to other band-pass textures, both in V1 and in extrastriate cortex (V2 and V3). We speculate that high anisotropy in the orientation responses leads to larger outputs in downstream areas, which in turn normalizes responses in these later visual areas, as well as in V1 via feedback.

1. Introduction

Primary visual cortex ("V1") has served as a testing ground for studying physiology, anatomy, brain development, neuroimaging, and computational modeling. There has been considerable success in developing general model forms that capture many of the encoding properties of V1 neurons reasonably well over a range of stimulus conditions, including the normalized energy model of V1 complex cells (Heeger, 1993). This type of model, like many others (Hubel and Wiesel, 1962; reviewed in chapter 6 of Wandell, 1995), includes a linear filter as the first stage, i.e., a weighted sum of the stimulus intensity over space and time. In a second stage, the outputs of the filter are squared and summed across nearby spatial locations or across phase (Adelson and Bergen, 1985; Heeger, 1992a; Pollen and Ronner, 1983). If the outputs are summed across a pair of linear filters tuned to the same frequency, orientation, and location, but differing in phase by 90 deg, it is called an energy model. In the third stage, the response of each neuron is normalized (divisively suppressed) by the (un-normalized) outputs of the nearby neural population (Heeger, 1992b, 1993). This effectively adjusts the gain based on the contrast energy in the image or image patch. There is substantial evidence that each of these three operations–linear filtering, energy, and normalization–contribute to the responses of V1 neurons (reviewed by Carandini et al., 2005).

The normalized contrast energy model, though initially developed to explain the outputs of single neurons, has also been successfully applied to functional MRI data in human visual cortex. First, a contrast energy model without normalization, applied to voxels in V1, V2, and V3, was used for successful encoding and decoding of natural images (Kay et al., 2008). Subsequent work showed that incorporating a normalization-like non-linearity improved model accuracy when testing stimuli that varied substantially in size (Kay et al., 2013b) or pattern (Kay et al., 2013c), and that normalization could account for the BOLD contrast response function for gratings with and without masking stimuli at other orientations (Brouwer and Heeger, 2011). These models have also shown good prediction accuracy for similar stimulus sets used in human intracranial electrode recordings of visual cortex (Hermes et al., 2019; Winawer et al., 2013).

The normalized contrast energy model, although successful at accounting for responses to a range of stimuli, nonetheless fails to explain some phenomena. There is some evidence that the standard model declines in accuracy for natural images (David et al., 2004; Olshausen and Field, 2005; Vinje and Gallant, 2000) (but see also (Rust et al., 2005)). Recent V1 two-photon calcium recordings included a large stimulus set and found that many cells in the superficial layer were sensitive to curvature, corners, and junctions (Tang et al., 2018), which the standard energy model is unlikely to predict, although a model was not fit to the data. Other studies with single unit electrophysiology found that a normalization model could be successful but only if the normalization was flexible, such that its strength depended on statistical dependencies in the image (Coen-Cagli et al., 2015). In human fMRI studies, the responses to natural/complex images in V1 also appear to be influenced by statistical dependencies and image context (Mannion et al., 2015; Qiu et al., 2016), factors unlikely to influence the predictions of the normalized energy model.

In visual areas beyond V1, the normalized energy model is expected to be incomplete, as circuits in these areas contribute new computations. There are no widely adopted standard encoding models for these areas analogous to the normalized energy model for V1, but there has been some success in modeling patterns in the responses extrastriate responses by incorporating higher-order statistical dependencies of the modeled V1 outputs (Freeman et al., 2013; Movshon and Simoncelli, 2015; Okazawa et al., 2017) or sensitivity to second-order contrast (Kay et al., 2013c).

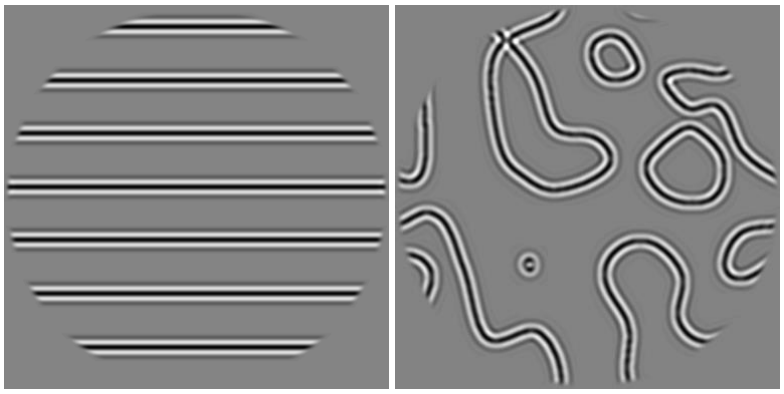


Figure 1. Example straight-line and curved-line stimuli from our experiments. We observe that in human V1, V2, and V3, stimuli with long straight lines (gratings) reliably evoke a smaller fMRI response than similar stimuli with curved lines.

Here, using evidence from human functional MRI, we show that the classical model fails to account for the relative responses to two classes of stimuli: straight, parallel, band-passed contours (*gratings*), and curved, band-passed contours (*snakes*) (Figure 1). The snakes elicit fMRI responses that are about twice as large as the gratings, yet traditional energy models, including normalized energy models, predict responses that are about the same. This is a large model failure which motivated us to develop a new model. We propose a computational model, in which responses are normalized not by the sum of the contrast energy, but by the variance in contrast energy computed across orientation channels.

2. Methods

We analyzed and modeled four fMRI data sets for this paper. Data sets 1 and 2 were collected at NYU for this project. Data sets 3 and 4 are re-analyzed from a previous paper **(**Kay et al., 2013c**),** for which the fMRI data and stimuli are freely available online (http://kendrickkay.net/socmodel/).

2.1 Participants

The two NYU participants were both experienced MRI subjects (female; 24, 29 yo). Data were collected at NYU's Center for Brain Imaging. The experimental protocol was approved by the University Committee on Activities Involving Human Subjects, and informed written consent was obtained from the participants before the study. Both participants had corrected-to-normal vision. The subjects participated in two separate scanning sessions, one for retinotopic mapping and one for the main study on encoding of textures.

2.2 Stimuli

Publicly accessible links to the stimuli, the names of the stimulus classes, and the correspondence between our naming convention and those in the Kay *et al.* (2013c) paper are described in Supplementary Tables 1 and 2.

2.2.1 Stimuli for data sets 3 and 4

Data sets 3 and 4 correspond to subject 1 and subject 2, respectively, in (Kay et al., 2013c). The stimuli for the two subjects were the same and are referred to as "Stimulus set 2" on the website (http://kendrickkay.net/socmodel/). The publicly available stimulus set includes 156 stimuli, 39 of which

were used for this paper (Supplementary Tables 2 and 3). The reason we use only a subset of the stimuli is that we modeled how visual areas respond to textures ignoring retinotopic preference, and many of the stimuli used in the original paper varied systematically over space in order to map spatial receptive fields. We used only the large-field textures, i.e., the subset of stimuli whose patterns were similar across the whole circular aperture.

The 39 stimuli are organized into 7 groups, each of which we describe with two terms, one term for the type of texture and one term for the way in which the stimuli within the group vary. For example, GRATINGS (contrast) are stimuli which come from the grating family and vary from low to high contrast. GRATINGS (density) stimuli come from the same family but have uniform contrast and vary in the spacing between the contours. The correspondence between how we refer to the stimuli and how Kay et al referred to them is in Supplementary Table 2. Below we describe the general stimulus characteristics and the 7 specific classes used for this paper. Most of the text is duplicated from (Kay et al., 2013c)(p. 11), indicated by italics.

General stimulus characteristics. Stimuli were constructed at a resolution of 256 pixels x 256 pixels and were upsampled to 800 pixels x 800 pixels for display purposes. All stimuli were presented within a circular aperture filling the height of the display; the rest of the display was filled with neutral gray. The outer 0.5 deg of the circular aperture was smoothly blended into the background using a half-cosine function.

Stimuli consisted of grayscale images restricted to a band-pass range of spatial frequencies centered at 3 cycles per degree. To enforce this restriction, a custom band-pass filter was used in the generation of some of the stimuli. The filter was a zero-mean isotropic 2D Difference-of-Gaussians filter whose amplitude spectrum peaks at 3 cycles per degree and drops to half-maximum at 1.4 and 4.7 cycles per degree. Restricting the spatial frequency content of the stimuli avoids the complications of building multiscale models and helps constrain the scope of the modeling endeavor. Even with the spatial frequency restriction, it is possible to construct a rich diversity of stimuli including objects and other naturalistic stimuli…. Each stimulus consisted of nine distinct images that were presented in quick succession. The purpose of this design was to take advantage of the slow dynamics of the BOLD response and average over stimulus dimensions of no interest (e.g., using sinusoidal gratings differing in phase to average over phase).

A key motivating observation for this paper is that the response to gratings was lower than to curved stimuli. Four groups of stimuli, two groups of gratings and two groups of snakes, were studied first and are referred to throughout the paper as *target stimuli.* These are described first.

Target stimuli

SNAKES (contrast, 10 stimuli). Kay et al. refers to these stimuli as noise patterns: *Noise patterns were created by low-pass filtering white noise at a cutoff frequency of 0.5 cycles per degree, thresholding the result, performing edge detection using derivative filters, inverting image polarity such that edges are black, and applying the custom band-pass filter (described previously). We generated nine distinct noise patterns and scaled the contrast of the patterns to fill the full luminance range.* [The contrast stimuli were then] *constructed by varying the contrast of the noise patterns.... Ten different contrast levels were used: 1%, 2%, 3%, 4%, 6%, 9%, 14%, 21%, 32%, and 50%. These contrast levels are relative to the contrast of the patterns used in SPACE* [not used in this study]*, which is taken to be 100%.*

SNAKES (density, 5 stimuli). *These stimuli used the same type of noise patterns as SPACE* [not used here] *but varied the amount of separation between contours. We generated noise patterns using cutoff frequencies of 2.8, 1.6, 0.9, 0.5, and 0.3 cycles per degree, and numbered these from 1 (smallest*

separation) to 5 (largest separation). The noise patterns used in SPACE correspond to separation 4; thus, we only constructed stimuli for the remaining separations 1, 2, 3, and 5. The noise patterns occupied the full stimulus extent (no aperture masking)

GRATINGS (contrast, 4 stimuli). *These stimuli consisted of horizontal sinusoidal gratings at 2%, 4%, 9%, and 20% Michelson contrast. The spatial frequency of the gratings was fixed at 3 cycles per degree.*

GRATINGS (density, 5 stimuli). The highest density stimulus in this group is similar to the horizontally oriented stimulus in GRATINGS (orientation), i.e., similar to a horizontal high-contrast grating, but it is not precisely a sinusoidal grating. It is made by convolving equally spaced horizontal lines with the custom band-pass filter (described previously). When the gratings are spaced appropriately (⅓ deg spacing) and filtered by a band-pass filter centered at 3 cycles per deg, the result is close to a sinusoidal grating at 3 cycles per deg. When the spacing is larger, there are several parallel band-pass contours with uniform gray between them. The spacing between parallel lines for the 5 stimuli varied in powers of 2, as 1/3 deg x 1,2,4,8, or 16, from densest to sparsest. Because the grating and snakes stimuli were both constructed by convolving lines with the same band-pass filter, they have some similar properties. They differ in that the lines here were straight whereas the lines used for constructing the snakes stimuli were curved.

Additional stimuli

GRATINGS (orientation, 8 stimuli). *These stimuli consisted of full-contrast sinusoidal gratings at eight different orientations. The spatial frequency of the gratings was fixed at 3 cycles per degree. Each [of the 9 exemplars per stimulus] consisted of gratings with the same orientation but nine different phases (equally spaced from 0 to 2π).*

PLAID (contrast, 4 stimuli). *These stimuli consisted of plaids at 2%, 4%, 9%, and 20% contrast (defined below). Each condition comprised nine plaids, and each plaid was constructed as the sum of a horizontal and a vertical sinusoidal grating (spatial frequency 3 cycles per degree, random phase). The plaids were scaled in contrast to match the root-mean-square (RMS) contrast of the GRATING stimuli. For example, the plaids in the 9% condition were scaled such that the average RMS contrast of the plaids is identical to the average RMS contrast of the gratings in the 9% GRATING stimulus.*

CIRCULAR (contrast, 4 stimuli). *These stimuli were identical to the PLAID stimuli except that sixteen different orientations were used instead of two.*

2.2.2 Stimuli for data set 1

Data set 1 was collected at NYU. The data set was designed to replicate some of the effects observed from data sets 3 and 4 (the greater response to snakes than gratings), but also to extend the measurements to new stimulus classes. The general stimulus characteristics were the same as those used in data sets 3 and 4. However, because the display size differed, the image resolution in pixels also differed (400 x 400 here, vs 800 x 800 above), and there were slight differences in the bandpass filter. The stimulus size in degrees of visual angle was the same (12.5 deg diameter). A total of 50 stimuli were tested. (The numbers below total more than 50 because some stimuli belong to more than one group, as indicated in Supplementary Table 2).

Target stimuli

GRATINGS (contrast, 5 stimuli). These stimuli are horizontal gratings (but not quite sinusoids), with a similar spatial pattern to the middle stimulus in the GRATINGS (density) stimuli from data sets 3 and 4.

They were made by convolving horizontal lines spaced every 1.75 deg with a custom band-pass filter. The images were scaled to yield 5 different contrasts of 3%, 10%, 25%, 50% and 100%.

GRATINGS (density; 5 stimuli). These stimuli are similar to horizontal gratings, made by convolving equally spaced horizontal lines with the custom band-pass filter. The 5 stimuli differed in the spacing of the horizontal lines, spaced every 3, 2.5, 1.75, 1, 0.33 deg. The contrast of all stimuli was 25%. The middle stimulus in this sequence was the same as the middle stimulus in the contrast sequence (spacing 1.75 deg, 25% contrast). The highest density (0.33 deg spacing) is close to a sinusoidal grating, as the spacing is the inverse of the peak spatial frequency of the band-pass filter (3 cycles per degree).

snakes (contrast, 5 stimuli). The spatial pattern is similar to the snakes stimuli in DS3 and DS4. Contrasts matched the grating contrasts (3%, 10%, 25%, 50% and 100%).

snakes (density, 5 stimuli). The spatial pattern is similar to the snakes stimuli in DS3 and DS4, but with 5 different densities of the contours. The contrast for all stimuli was 25% (lower than the contrast of the corresponding stimuli in DS3 and DS4). The range of densities used here was also lower than the range used in DS3 and DS4, with the densest pattern here similar to the middle stimuli in DS3 and DS4.

Additional stimuli

GRATINGS (orientation, 4 stimuli). These stimuli are the same as the third stimulus in the GRATINGS (density) group (25% contrast, 1.75 deg spacing between contours), except that they are rotated by 0, 45, 90, or 135 deg. Because the 0 deg rotation does not change the image, a new stimulus was not created; for visualization of results, the BOLD measurements and model predictions are plotted for both groups.

GRATINGS (cross, 4 stimuli). These stimuli contain horizontal contours similar to two of the stimuli in the GRATINGS (density) sequence, except that they have periodic vertical blank regions which interrupt the contours. For two of the stimuli, the spacing of the horizontal contours matches the densest stimuli in the density sequence (spacing of 0.33 deg) and for two of the stimuli, the spacing matched the middle stimulus in the density sequence (1.75 deg spacing). In all 4 images, the horizontal contours are interrupted by vertical blanks spaced every 1.75 deg. The vertical blanks are either thick (50% duty cycle; 1st and 3rd stimulus) or thin (25% duty cycle; 2nd and 4th stimuli).

NOISE BARS (density, 5 stimuli). These stimuli have the same contrast apertures as the GRATINGS (density) stimuli. Specifically, there are horizontal bands containing contrast patterns, spaced the same as the grating stimuli (bands every 3, 2.5, 1.75, 1, or 0.33 deg). These stimuli differ from the gratings in that each band contains band-pass filtered noise, equal in power across orientations, rather than horizontal contours.

NOISE BARS (contrast, 5 stimuli). These stimuli are matched in spatial pattern to the middle density of the NOISE BARS (density) stimuli (horizontal lines, spacing 1.75 deg), but scaled in contrast similar to the grating stimuli (3%, 10%, 25%, 50%, 100%).

NOISE BARS (orientation, 4 stimuli). The orientation sequence rotated the middle stimulus of the NOISE BARS (density) group (spacing 1.75 deg, contrast 25%) by 0, 45, 90, or 135 deg.

WAVES (density, 6 stimuli). These are identical to the snakes (density) stimuli, except that they have been filtered by orientation, such that they only contain power at or near the horizontal.

WAVES (contrast, 5 stimuli). These are identical to the snakes (contrast) stimuli, except that they have been filtered by orientation, such that they only contain power at or near the horizontal.

WAVES (orientation, 4 stimuli). These are identical to the densest stimulus in the snakes (density) group, except that they have been filtered by orientation, with filter centered at either 0, 45, 90, or 135 deg.

2.2.3 Stimuli for data set 2

Data set 2 was collected at NYU. The stimuli were nearly identical to those in data set 1, differing only in the following ways. First, the stimuli were 50% larger (18.75 x 18.75 deg and 600 x 600 pixels, rather than 12.5 x 12.5 deg and 400 x 400 pixels). The difference in size did not entail a difference in spatial frequency: The spatial frequency was matched between the two data sets (meaning that the stimuli were re-made with a larger aperture rather than by re-scaling). Second, those stimuli which were oriented were oriented vertically rather than horizontally. This applies to all GRATING stimuli, as well as NOISE BARS and WAVES. Third, the WAVES (density) stimuli had only 4 densities rather than 6. We reduced the number of stimuli to slightly shorten the MRI scans.

2.3 MRI

The methods for MRI acquisition and preprocessing for data sets 3 and 4 are described in Kay et al (2013c). In brief, each data set comes from one subject, who viewed a variety of stimuli in an event-related fMRI design. Data set 3 was collected over two scan sessions and each stimulus was presented 6 times. Data set 4 was collected over one scan session and each stimulus was presented 3 times. (Note that both in this paper and the Kay website (http://kendrickkay.net/socmodel/), these two datasets are referred to as data sets 3 and 4. However, the Kay website refers to the stimuli for these two datasets as stimulus set 2 and the subjects themselves as “subject B” and “subject C”. We do not adopt these latter two conventions.)

After preprocessing the data (slice-time correction, co-registration, spatial unwarping), a general linear model was applied using the GLMdenoise toolbox (Kay et al., 2013a). The output of this algorithm includes a coefficient (beta weight) for each stimulus for each voxel solved from the whole fMRI session, as well as 30 bootstrapped estimates of each beta weight (bootstrapping across fMRI runs). The publicly available data (http://kendrickkay.net/socmodel/) are already pre-processed, denoised, and organized by ROI. Specifically, the data we used are in the files called “data set03.mat” and “data set04.mat” (http://kendrickkay.net/socmodel/data set03.mat, http://kendrickkay.net/socmodel/data set04.mat). The datasets are described on the website as “Dataset 3 (subject B)” and “Dataset 4 (subject C),” respectively. Within the MATLAB files, we used the stored 3D array called “betas” (voxels x stimuli x bootstraps), limited to V1, V2, V3 as indicated in the grouping variables “roi” and “roilabels”, and limited to the 39 stimuli indicated in Supplementary Table 2. Visual areas were identified by retinotopic mapping in a separate session.

2.3.1 Acquisition of data sets 1 and 2

Datasets 1 and 2 were acquired in one scanning session each. Each scanning session had 12 fMRI runs of 249 s each (data set 1) or 241.5 s each (data set 2). For each data set, half of the stimuli were assigned to odd fMRI runs and half to even runs, so that each stimulus was shown 6 times in the session. The stimulus events were 3 s long, consisting of 9 alternations between stimulus exemplar and blank, ⅙ s each. Trial onsets were every 7.5 s (so 4.5 s blank between trials). To help estimate the hemodynamic response function, there were 12 s of blank at the beginning and end of each run, as well as 5 additional trials randomly interspersed with no stimulus (meaning that 5 times during the scan, trials were separated by 15 s instead of 7.5 s). Thus, each complete run consisted of either (25 stimuli + 5 blanks) * 7.5 s + 24 s = 249 s (data set 1) or (24 stimuli + 5 blanks) * 7.5 s + 24 s = 241.5 s (data set 2).

All MRI data were acquired at New York University Center for Brain Imaging using a Siemens Allegra 3T head-only scanner with a Nova Medical phased array, 8-channel receive surface coil (NMSC072). For each participant, we collected functional images (single shot echo planar images, 1500 ms TR, 30 ms TE, and 72° flip angle). Voxels were 2.0 mm^3 isotropic, with 24 slices, with an inplane sampling of 104 x 80 voxels (208 mm A/P x 160 mm L/R). The slice prescription covered most of the occipital lobe, and the posterior part of both the temporal and parietal lobes. Images were corrected for B0 field inhomogeneity using a calibration scan and Center for Brain Imaging algorithms during offline image reconstruction.

We also acquired 1 or 2 T1-weighted whole-brain anatomical scans (MPRAGE sequence; $1mm^3$), as well as a T1-weighted "inplane" image with the same slice prescription as the functional scans. This scan had an inplane resolution of 1.25 X 1.25 mm and a slice thickness of 2.5 mm, and was collected to aid alignment of the functional images to the high-resolution T1 weighted anatomical images.

In a separate session, retinotopy scans were collected and analyzed using a pRF model as implemented in the Vistasoft software tool (*https://github.com/vistalab/vistasoft*). The methods for acquisition and analysis of the retinotopy data are identical to that described by Zhou et al. (2018).

2.3.2 Data Preprocessing and Analysis

Data preprocessing. Processing of the fMRI data was identical to that described by Zhou et al. (2018):

> *We coregistered and segmented the T1 weighted whole-brain anatomical images into gray and white matter voxels using FreeSurfer's autosegmentation algorithm (http://surfer.nmr.mgh.harvard.edu). Using custom software Vistasoft (https://github.com/vistalab/vistasoft), the functional data were slice-time corrected by resampling the time series in each slice to the center of each 1.5 s volume. Data were then motion-corrected by coregistering all volumes of all scans to the first volume of the first scan. The first 8 volumes (12 s) of each scan were discarded for analysis to allow longitudinal magnetization and stabilized hemodynamic response.*

GLM. The preprocessed fMRI data were then fit by a general linear model, GLMDenoise (Kay et al., 2013a). This algorithm denoises the data by projecting out nuisance regressors derived in a data-driven manner, and estimates coefficients for each of the 48 or 50 stimuli for each voxel in the functional images. The algorithm bootstraps the data over fMRI runs. For data sets 1 and 2, we generated 50 bootstraps for data set 1 and 100 bootstraps for data set 2. The publicly available data from Kay et al. (2013c), included 30 bootstraps per subject. The algorithm also estimated a hemodynamic impulse response function as a finite impulse response function, with 35 time points (52.5 s) per subject.

ROIs. Regions of interest for V1, V2, V3 were delineated manually using the Vistasoft (https://github.com/vistalab/vistasoft) graphical user interface to visualize the results of the pRF models. These methods for identifying these boundaries are well established, as described in many publications (Engel et al., 1997; Sereno et al., 1995; summarized by Wandell and Winawer, 2011). The ROIs for V1, V2 and V3 were identified on the cortical surface and then projected to the functional images. For purposes of data summary and model fitting, we took the average signal from each ROI. We did this by averaging the beta weight across voxels within an ROI separately for each stimulus, after voxel selection (**Supplementary Table 1**). Because noise can be correlated across voxels, but should not be correlated across scans, when we bootstrapped the data, we average across voxels within an ROI for each bootstrap. For the purposes of model fitting, each of the 4 datasets comprised two matrices, one for the means and one for the standard deviation across bootstraps, each of which had a size equal to the number of stimuli by number of ROIs.

2.4 Model Equations

In the Results, we compared the accuracy of 4 models fit to the data, three of which are based on existing models or empirical findings–a contrast energy model, a second order contrast model, and an orientation tuned surround model–and one new model, which computes normalization ny orientation anisotropy. In this section we describe the computation that comprises each model.

All four models consist of three primary steps: (1) computation of oriented contrast energy, (2) pooling across orientation and space, and (3) a power-law nonlinearity. Steps 1 and 3 are identical for all models. Step 2, spatial pooling, varies between models. To reduce the computational demand, all stimuli were downsampled to 150 x 150 (datasets 1, 3, 4) or 225 x 225 (dataset 2) and remapped to -0.5 to 0.5 intensity values (Supplementary table S2). The images were then padded to 166 x 166 (datasets 1, 3, 4) or 241 x 241 (dataset 2).

1. Contrast energy: We denote by $I(x, y)$ the value of the pre-processed input image at coordinates (*x*,*y*). The pre-processing causes the image values to have mean 0 and range -0.5 to 0.5. The image is projected onto a set of 16 Gabor filters, which comprise 8 orientations(θ) and 2 phases (ϕ), separated by 90º (i.e., "quadrature"). $F(x, y, \theta, \phi)$ indicates the Gabor filter at a spatial location (*x*,*y*), orientation θ, and phase ϕ. Because the stimuli are all band-passed, we only filter at one spatial scale, matched to the stimulus power (3 cpd). The filters had two cycles each. The outputs over the two phases are squared, summed, and square-rooted. Finally, the contrast energy as a function of spatial position (x, y) and orientation (θ) becomes

$$E(x, y, \theta) \;=\; \sqrt{\sum_{\phi}\left(\sum_{x',y'} I(x - x', y - y')F(x', y', \theta, \phi)\right)^2}$$

We convolve the image I and the filter F. The computation of contrast energy has no free parameters. Typically, the calculation of contrast energy does not have a square root, but we included it for consistency with some prior work (Kay et al., 2013c).

2. Spatial pooling. Each model differs in how contrast energy is pooled to yield a scalar value, s:

$$s \;=\; \Phi[E],$$

where we use square brackets to indicate a function of a function (also called a *functional*). We describe the pooling functional, Φ, for each model below.

3. Power-law nonlinearity. Finally, the scalar is passed through a power-law nonlinearity to predict the BOLD amplitude *r* in units of percent signal change:

$$r = g\, s^{\alpha}$$

Where g is the gain and α is the exponent parameter. These are free parameters fit to the human data. The power-law nonlinearity is similar to divisive normalization in the case where each unit in a population is normalized by the same pool (Kay et al., 2013b).

2.4.1 Contrast energy model

In the contrast energy model, the contrast energy is summed over orientations and space to yield a scalar output, s.

$$s = \frac{1}{N_{\text{ori}} \cdot N_{\text{pixels}}} \sum_{x,y,\theta} E(x,y,\theta)$$

N_{ori} is the number of orientation channels (always 8) and N_{pixels} is the number of pixels per stimulus in the downsampled and padded images (166^2 or 241^2). There are no free parameters in this step, so the contrast energy model has only two free parameters, g and α, both from the power-law nonlinearity step.

2.4.2 Second-order contrast model

Rather than summing the energy over all channels (space and orientation), the SOC model first sums over orientations at each location to generate a single contrast image, E_{spatial}:

$$E_{\text{spatial}}(x,y) = \frac{1}{N_{ori}} \sum_{\theta} E(x,y,\theta)$$

The model then computes the second-order contrast (contrast energy within the contrast image), to yield a scalar, *s*:

$$s = \frac{1}{N_{\text{pixels}}} \sum_{x,y} \left(E_{\text{spatial}}(x,y) - w\bar{E}_{\text{spatial}}\right)^2$$

Where $\bar{E}_{\text{spatial}}$ is the mean of the contrast image, $\bar{E}_{\text{spatial}}(x,y)$, computed over space. w is a free parameter bounded by 0 and 1. When w is 1, *s* is the sum of squared deviations from the mean, analogous to the variance in the contrast image. When w is 0, *s* is the sum of squares (the total power in the contrast image), and hence is similar to the contrast energy model.

This step adds one free parameter, w. Thus, the SOC model has 3 free parameters, w, g, and α .

2.4.3 Orientation-tuned surround model

The contrast energy in the OTS model is normalized before it is summed: Each (x,y,θ) element in the energy image is divided by a weighted sum of the other elements. Those weights are relatively high for elements at the same (x,y) location but other orientations, and for elements within an ellipse in the (x,y) slice at the same theta value. The ellipse is elongated in the orientation defined by theta. This is akin to summing two forms of normalization, cross-orientation suppression (same location, other orientations), and an orientation-tuned surround (same orientation, other locations), with the shape of the surround elongated.

$$Z(x,y,\theta) = \sum_{x',y',\theta'} E(x-x',y-y',\theta) F(x',y',\theta';\theta)$$

$Z(x,y,\theta)$ is the normalizer of the contrast energy image $E(x,y,\theta)$. The filter $F(x,y,\theta';\theta)$ is *n* pixels x *n* pixels x 8 orientations, where *n* is 25% of the image width. The filter assigns relative weights for normalization. When the orientational channel of the image and filter matches $\theta' = \theta$, the 2-D filter of the channel is an elongated Gaussian with standard deviation of 0.85 x *n* pixels at the matched orientation and $0.1 \times n$ pixels at the orthogonal orientation. At all other orientations (i.e., $\theta' \neq \theta$), the filter is a symmetric 2D Gaussian distribution with a standard deviation with 0.01 x *n* pixels.

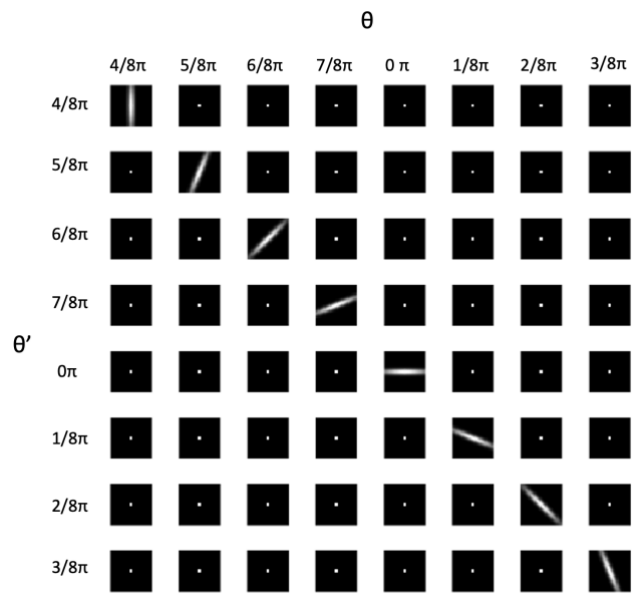


Figure 2. The orientation-tuned surround normalization filter. When $\theta' = \theta$(diagonal), the filter is elongated in the direction of θ. For other orientations (off diagonals) the filter is near zero except at the center pixel (cross-orientation suppression).

The normalized contrast energy is

$$d(x, y, \theta) = \frac{E(x,y,\theta)}{1 + w\, Z(x,y,\theta)}$$

where w is a parameter to control the strength of normalization. When w approaches 0, the normalization is low, and the overall expression approximates the contrast energy model. When w is large, there is strong normalization. We then sum d across space and orientation to result in the scalar, s.

$$s = \frac{1}{N_{\text{ori}} \cdot N_{\text{pixels}}} \sum_{x,y,\theta} d(x, y, \theta)$$

Other than the overall weight of normalization, w, the parameters of the surround filter were chosen to maximize the suppression in grating-like stimuli, which contain elongated contours in a single direction. Prior analyses showed that a model with an orientation-tuned surround that was spatially symmetric (circular) predicted little difference between the responses to gratings and to noise patterns (Olsson, Kay, & Winawer, 2015).

We note that this model has two normalization-like calculations, one that is orientation tuned and one on the final output. This is consistent with prior work showing that two stage of normalization (a cascade model) improved model accuracy (Kay et al., 2013c).

2.4.4 Normalization by orientation anisotropy

In a new model developed here, the normalization by orientation anisotropy (NOA) model, the pooling step first sums the contrast energy across space within an orientation band, resulting on one value per orientation band, $E_{\text{ori}}(\theta)$:

$$E_{\text{ori}}(\theta) = \frac{1}{N_{\text{pixels}}} \sum_{x,y} E(x, y, \theta)$$

E_{ori} indicates the oriented energy. This energy at each orientation is then normalized by the variance across the 8 orientations, and then summed to produce a scalar.

$$s = \frac{1}{N_{\text{ori}}} \sum_{\theta} \frac{E_{\text{ori}}(\theta)^2}{1 + w^2 Var(E_{\text{ori}})}$$

Where $Var(E_{\text{ori}}) = \sum_{\theta}(E(\theta) - \bar{E}_{\text{ori}})^2$ calculates the variance of the oriented energy and a non-negative parameter w controls the strength of the normalization. When $w$approaches 0, the normalization is low, and the overall expression approximates the contrast energy model. When w is large, there is strong normalization by the variance across orientation channel outputs. Calculating the variance of oriented energy involves a squaring operation. To keep the parameters comparable across different models, we also square the numerator and the parameter w.

The pooling step adds one free parameter, w. Thus, the NOA model has 3 free parameters, w, $g$and α.

2.5 Optimization

In each model, we fitted the model free parameters using the MATLAB optimization tool *fmincon* by minimizing the squared error between the model prediction and the corresponding BOLD amplitude. Because each stimulus consisted of 9 exemplars shown to the subject in rapid succession, the model prediction for each stimulus was obtained by averaging the model predictions across the exemplars. To avoid getting stuck in the local minima of the nonconvex landscape, we ran the optimization algorithm with 40 different parameter initializations. Each initialized value was picked randomly. Except for the parameter w of the SOC model, which was constrained to lie between 0 and 1, all parameters were unbounded in the search, minimizing human interference in the fitting.

2.6 Cross-validation scheme

All models were fit using an *n*-fold leave-one-out cross-validation scheme, where *n* is the number of stimuli. Thus, the BOLD signal prediction for each stimulus was generated by a model fit to all stimuli except that one. Under this scheme, the models are less likely to overfit data sets.

2.7 Accuracy metric

The model accuracy was quantified as the percentage of the explained variance (R^2) in the human BOLD data by the cross-validated model predictions,

$$R^2 = 1 - \frac{\sum_{i=1}^{M}(r_i - \hat{r}_i)^2}{\sum_{i=1}^{M}(r_i - \bar{r})^2}$$

where r_i represents the BOLD amplitude to the ith stimulus, $\hat{r}_i$ represents the corresponding model prediction, and $\bar{r}$ is the mean response across stimuli. We can understand this metric as the extra uncertainty reduction brought by the model beyond describing the human data by its mean.

3. Results

3.1. The fMRI BOLD responses to curved patterns are larger than to straight patterns

We first consider an observation about fMRI responses to two classes of simple, gray-scale, band-passed, static images. For one class, the stimuli contain several curved contours, which we refer to throughout as snakes. For the other class, the stimuli contain several straight, parallel contours, which we refer to as gratings. The surprising observation is that for V1, V2 and V3, the fMRI responses are substantially larger for the snakes than for gratings (**Figure 3**). The responses to the gratings, irrespective of density, are only about as high as the response to the sparsest snakes. We confirm this pattern with three additional fMRI data sets, which also show larger responses to snakes than gratings in V1, V2 and V3 (**Supplementary Figure S2**).

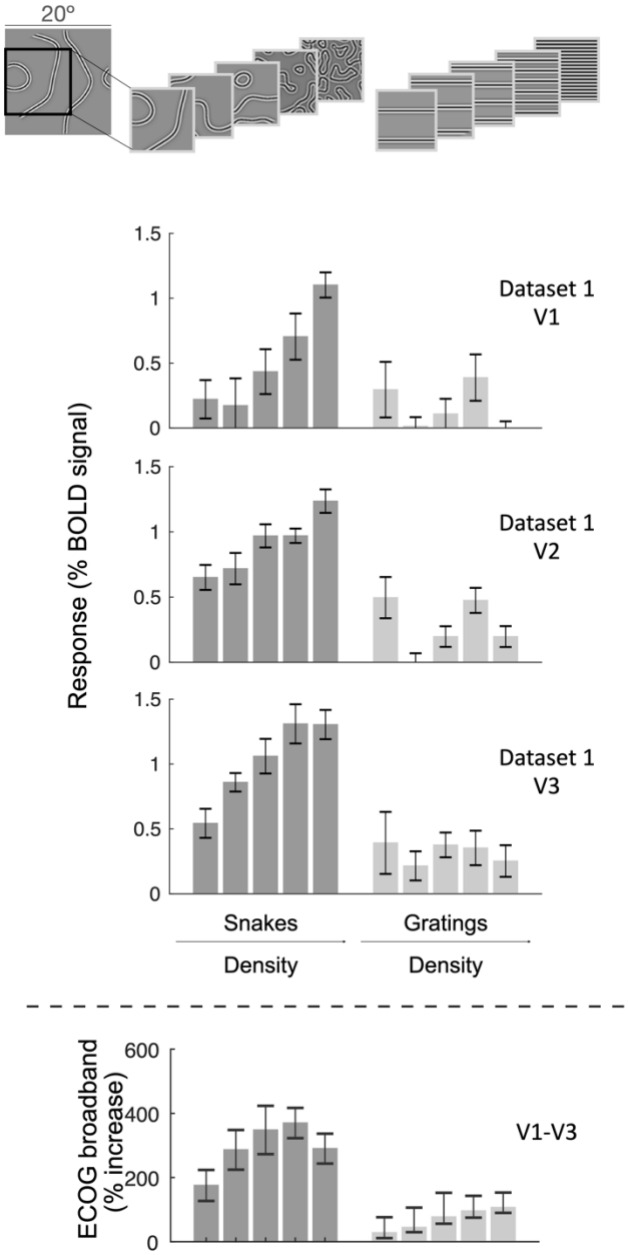


Figure 3. FMRI responses are larger for curved patterns than for straight patterns. The mean fMRI responses within a visual area are plotted for V1, V2, and V3 in subject S1. For both curved stimuli (*snakes*, dark bars) and straight stimuli (*gratings*, light bars), data are plotted in order of increasing texture density from left to right. Examples of the 5 densities are shown above. Error bars are the standard deviation of the mean, bootstrapped across fMRI runs. Responses are larger for the *snakes* than *gratings*. The same effect is also observed for subjects S2-S4 (Supplementary Figure 2). Bottom panel: For comparison, we replot data from intracranial recordings (ECoG) from Hermes et al 2019, which also show larger responses for *snakes* than for *gratings*. The full set of stimuli is shown in Supplementary Figure 1. Data plotted from the function *s4_visualize('figure3')* in the GitHub code repository.

To check whether this pattern of results is specific to fMRI, we replot published intracranial data from Hermes et al (2019), in which human subjects with ECoG recordings viewed a similar set of stimuli (**Figure 3**, bottom panel). The ECoG data, plotted as the percent increase in broadband power over baseline, show the same general effect as the fMRI data: The responses to snakes are much larger than to gratings, irrespective of texture density, indicating that the effect is not limited to the fMRI BOLD response. See Hermes et al. (2019) for methodological details.

In addition to the difference in the mean responses to snakes and gratings, there are also differences in the slope, meaning the increase in response amplitude as a function of texture density. From the sparsest to densest patterns, the responses rise from about 0.25% (V1) or 0.5% (V2 and V3) BOLD signal change to over 1%. The responses to the grating stimuli vary much less with density. For the ECoG data, the responses also rise much more as a function of density for the snakes than for the gratings.

In the next four sections, we describe the four models that are fit to the data. For simplicity, for each model we consider, the parameters are fit to the aggregate (average) data within a visual area, rather than to each voxel individually. All of the stimuli are texture-like, meaning that they have similar properties across the whole image aperture, and for each stimulus class, nine different exemplars were shown per 3-s trial. The different exemplars have the same higher-order statistics but vary in the precise spatial distributions. Hence model computations, such as contrast energy, would result in similar values for spatially localized portions of the image (as one would compute for an individual voxel) and for the whole image (as we compute to model the aggregate response of a visual area). For this reason, we did not include model parameters for the spatial location or spatial extent of the receptive fields for individual voxels. We first describe models fit to the two classes of stimuli described above (gratings and snakes) that vary in either density or contrast, which we refer to as *target stimuli*. We then summarize fits to the larger set of stimuli viewed by the subjects.

3.2. The larger response to snake stimuli is not captured in a simple energy model

The larger response to snakes than to gratings is not predicted by a standard contrast energy model. The model computes the oriented contrast energy at the stimulus spatial frequency. The contrast energy is then summed across space and orientations to give a total contrast energy. Finally, this value is passed through a power-law nonlinearity to predict the percent BOLD responses. The power-law nonlinearity acts similarly to a divisive normalization computation in which the normalization pool is shared across units (Kay et al., 2013b, section “Relationship to Divisive Normalization”).

The contrast energy model predicts that responses should increase with both stimulus contrast and with density of the pattern. It does not predict systematic differences between the two classes of stimuli (**Figure 4**). Specifically, it does not predict larger responses to snakes than to gratings, contrary to the data, nor does it predict a higher slope as a function of density or contrast. In fact, the cross-validated variance explained is negative for all 3 visual areas in data set 1, meaning that the model prediction is less accurate than it would have been if it simply predicted the mean response for all stimuli. In short, the contrast energy model provides a poor fit to the fMRI data in V1-V3 for these classes of stimuli. It is also a poor fit to the target stimuli in the other three data sets (**Table 1**, first row). This does not mean contrast energy models are always poor fits to fMRI responses in V1-V3. For example, when stimuli vary in how contrast energy is distributed across space, a contrast energy model can capture a lot of the variance in the responses across images, as shown by Kay et al (2008).

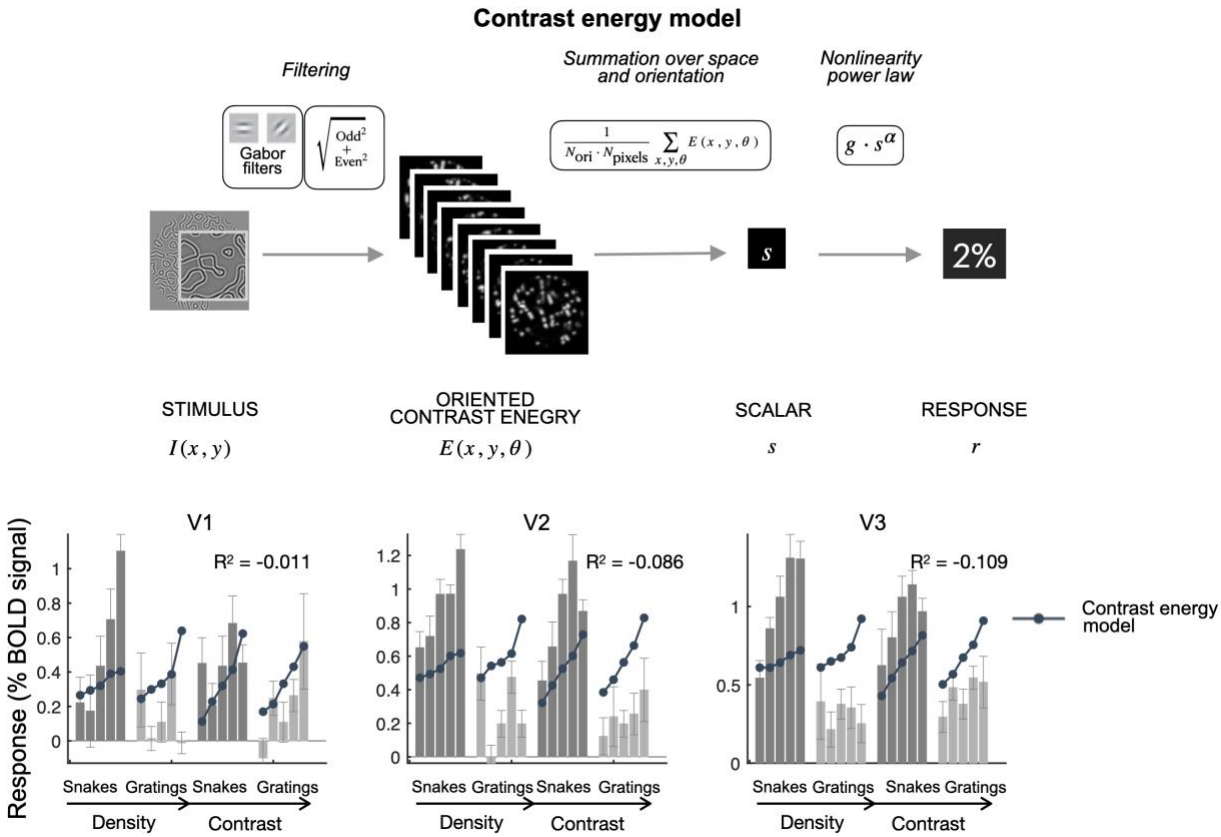


Figure 4: Contrast energy model does not account for V1-V3 responses to these stimuli. The mean fMRI responses from V1, V2, and V3 are shown for snakes and grating stimuli that vary in density and contrast for subject S1. The bar plots show the mean and standard error of the responses. The dots are the cross-validated predictions from the contrast energy model. Dark bars represent snake stimuli and light bars represent grating stimuli. Each group of stimuli is arranged in increasing order of either density or contrast. Data and model fits plotted from the function *s4_visualize('figure4')* in the code repository. The upper panel is a schematic representation of the contrast energy model. See *Figure S2b* for fits to all 4 datasets.

Because the contrast energy model failed to capture the effect of interest, we considered two alternative models, a Second-order contrast model (SOC) that has been previously fit to both fMRI data (Kay et al., 2013c) and ECoG data (Hermes et al., 2019), and a contrast energy model with orientation tuned surround suppression. A preliminary implementation of the Orientation-tuned surround model was presented as a conference abstract (Olsson et al., 2015).

3.3. The larger response to snake stimuli is not captured by a second-order contrast model

The SOC model is a two-stage cascade model developed by Kay et al (2013c), which computes the contrast energy in the image (first stage), and then the contrast within the contrast image (second stage). The contrast in the contrast image is called second-order contrast, which gives the model its name. The motivation for incorporating a second-order contrast term is that the observed BOLD response is sometimes greater for a stimulus where the contrast energy is distributed heterogeneously compared to a stimulus with a similar or greater total contrast energy distributed more homogeneously across the image. Second-order contrast computations are also a well-established component of many texture models (Landy and Graham, 2004; Larsson et al., 2006; Shooner et al., 2015). As with all the

models in this paper, we did not fit spatially local receptive fields to each voxel, so we compute both first-order contrast energy and second-order contrast over the whole stimulus. This required a modification from the published SOC model.

The SOC model predicts a similar BOLD amplitude for snakes and gratings, thereby failing to account for the observation about the two stimulus classes (**Figure 5**). While the SOC model is sensitive to heterogeneity in contrast energy across space, it is not sensitive to heterogeneity across orientations. This is because the non-linearities computed in the second stage come after energy is pooled across orientation channels in the first stage. Therefore, the output does not depend on whether the contrast energy is concentrated in one orientation channel as in the gratings, or spread across many channels as in the snakes.

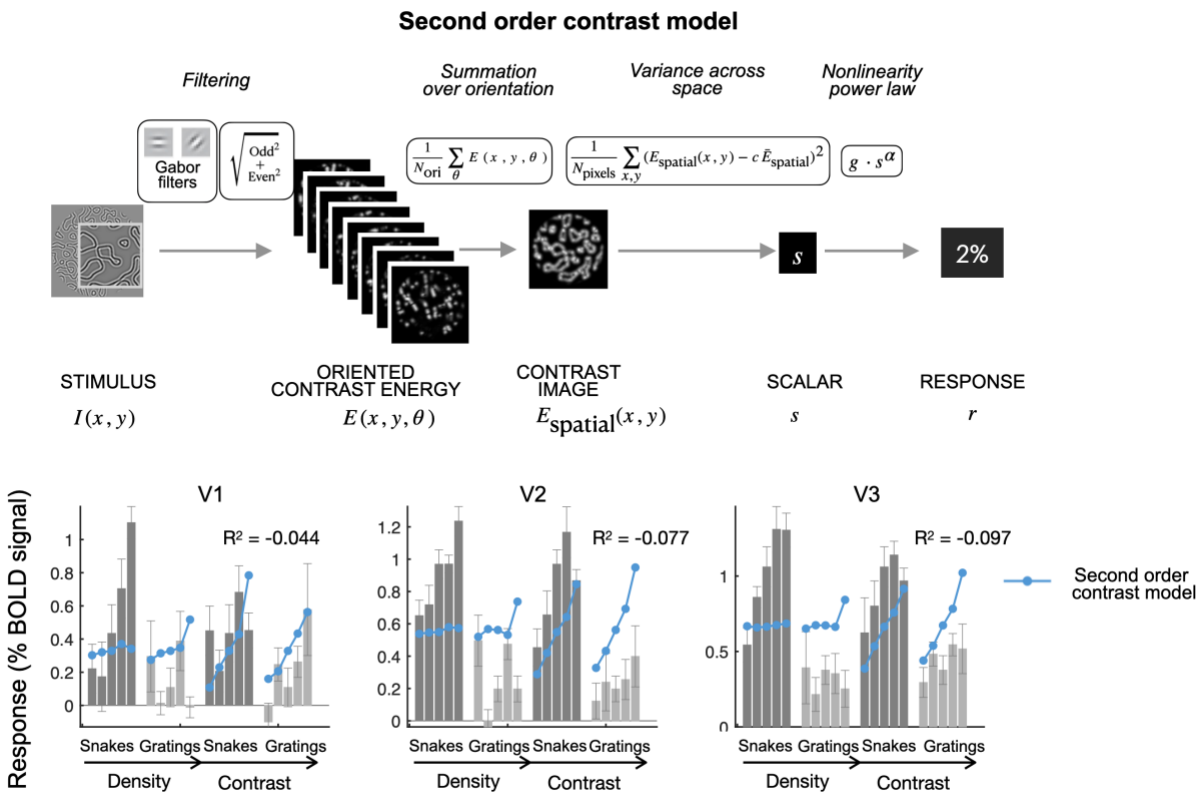


Figure 5. The Second-order contrast model does not account for V1-V3 responses to these stimuli. The mean fMRI responses from V1, V2, and V3 are replotted from Figure 4. The blue dots are the cross-validated predictions from the Second-order contrast model. Data and model fits plotted from the function *s4_visualize('figure5')* in the code repository. See *Figure S2c* for fits to all 4 datasets.

The SOC model responds differently to density variations than the contrast energy model. For the contrast energy model, greater density of gratings or snakes results in more contrast energy and hence a larger predicted output. For the SOC model, there is a tradeoff: as density increases, the first-order contrast increases while the second-order contrast can decrease. In the limit, a pure harmonic has no second-order contrast. Because of the tradeoff, for some stimuli and model parameters, the two effects approximately cancel, which is why the predictions are close to flat in V1-V3 for the SOC model predictions in figure 5. When density is held fixed, the SOC model predictions increase as the stimulus first-order contrast increases, similar to the contrast energy model.

The SOC model's failure to account for the greater response to snakes is confirmed by low variance explained in each of the four data sets with these target stimuli (**Table 1,** second row).

3.4. The larger response to snake stimuli is not captured by an Orientation-tuned surround model

The Second-order contrast model can be thought of as implementing surround suppression: Second-order contrast is low when both the center and surround of a receptive field are stimulated, and high when only the center is stimulated. The SOC model does not take into account the specific orientations stimulating the center or surround. Work from macaque electrophysiology, however, indicates that surround suppression is orientation-tuned. Specifically, Cavanaugh, Bair, and Movshon (2002) reported that the response of a neuron to a stimulus at its preferred orientation in its receptive field is suppressed more when the surrounding region contains contrast at the same orientation compared to contrast at different orientations. Because our grating stimuli have contrast energy concentrated at a single orientation, and the snake stimuli do not, one might surmise that an Orientation-tuned surround model would show greater suppression for the gratings, where the RF centers and surrounds will have matched orientations, than for the snakes, where the orientations are more likely to differ between center and surround. If so, this could then account for our observed effect.

To implement an Orientation-tuned surround model, we use the same first-stage contrast energy calculation as used in both the contrast energy model and the SOC model. We then normalize the contrast energy at each location and orientation by the outputs across orientations at the same location (also called cross-orientation suppression), and across nearby locations at the same orientation (orientation-tuned surround). The normalized energy images are then combined (summed over space and orientations) and passed through a nonlinearity power law, as in the contrast energy model.

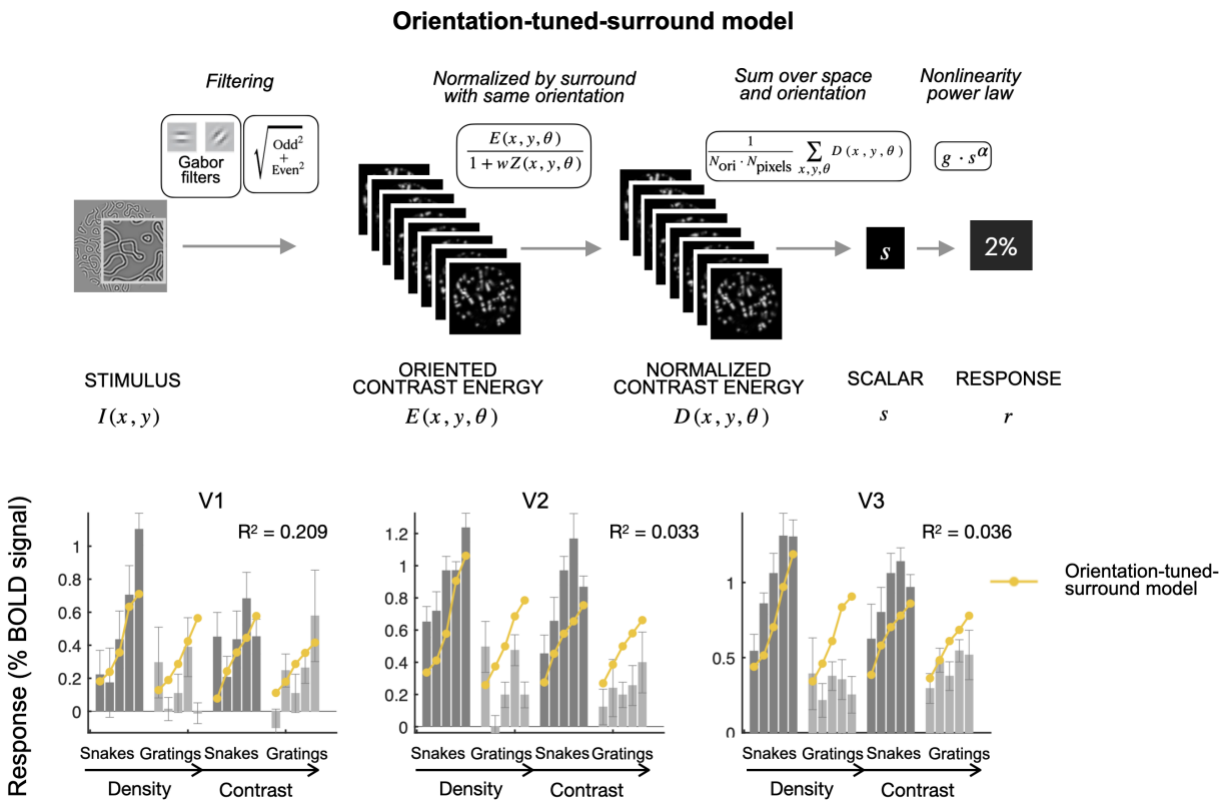


Figure 6. An Orientation-tuned surround model does not account for V1-V3 responses to these stimuli. The mean fMRI responses from V1, V2, and V3 are replotted from Figure 4. Z = FE. The yellow dots are the cross-validated predictions from the Orientation-tuned surround model. Data and model fits plotted from the function *s4_visualize('figure6')* in the code repository. See *Figure S2d* for fits to all 4 datasets.

The OTS model is slightly improved compared to the SOC model for these data, but nonetheless it fails to capture the large and systematic differences in response amplitude between gratings and snakes. In V2 and V3, the model underpredicts the responses to snakes and overpredicts the responses to gratings, similar to the contrast energy model and the SOC model. Across the 4 stimulus sets, its performance on these target stimuli is better than the first two models tested, but it still has low accuracy, especially for V2 and V3 (R^2 ranging from 0.033 to 0.414; **Table 1**, third row).

3.5. Normalization By Orientation Anisotropy Model accounts for the larger response to snake stimuli

The failures of the three models described above to account for the responses to the target stimuli motivated us to develop a new model. The reduced response to the gratings suggests the possibility that when the contrast energy is concentrated in a single orientation channel, there is greater normalization. We implemented this possibility in a model, in which the contrast energy is normalized by the variance across the outputs of the orientation channels, rather than the sum of the outputs. We refer to this as the NOA model (**Figure 7**).

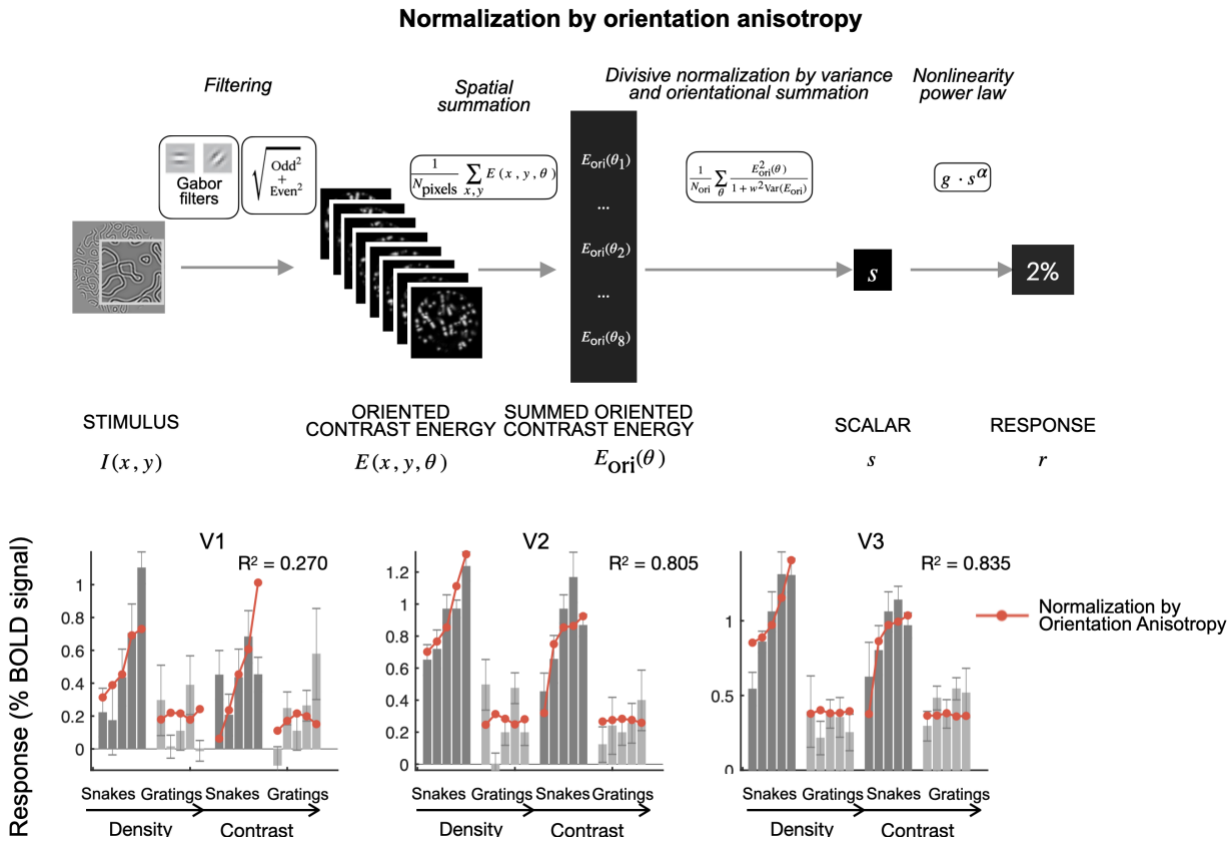


Figure 7. The NOA model better accounts for the responses in V1-V3. The mean fMRI responses from V1, V2, and V3 are replotted from Figure 4. The red dots are the cross-validated predictions from the normalization by orientation anisotropy (NOA) model. Data and model fits plotted from the function *s4_visualize('figure7')* in the code repository. See *Figure S2a* for fits to all 4 datasets.

The NOA model does an excellent job accounting for the differences in the responses to gratings and snakes. In data set 1, the BOLD amplitude and the NOA model predictions are about one-third for the gratings compared to the snakes, both for the density and contrast stimulus manipulations, and for all three visual areas. In addition to capturing this difference in the means between the two stimulus classes, the model also captures the difference in slope. As the density or contrast increases, the model predicts steeper slopes for snakes than gratings. These patterns, both in the data and in the model fits, were found across the four data sets (**Supplementary Figure 2**). The NOA model is more accurate than the other three models in all 12 cases (four datasets and three ROIs). The improvement over other models is modest for V1 and large for V2 and V3 (**Table 1**).

Model (free params)	**V1**				**V2**				**V3**			
	DS1	**DS2**	**DS3**	**DS4**	**DS1**	**DS2**	**DS3**	**DS4**	**DS1**	**DS2**	**DS3**	**DS4**
Contrast (2)	-0.011	0.599	-0.100	0.119	-0.086	0.248	-0.201	-0.134	-0.109	-0.013	-0.190	-0.181
SOC (3)	-0.044	0.583	-0.216	-0.038	-0.077	0.234	-0.056	-0.008	-0.097	0.148	-0.019	0.004
OTS (3)	0.209	0.621	0.431	0.474	0.033	0.414	0.104	0.245	0.036	0.032	0.012	0.088
NOA (3)	**0.270**	**0.805**	**0.652**	**0.568**	**0.805**	**0.767**	**0.744**	**0.669**	**0.835**	**0.788**	**0.687**	**0.694**

Table 1. Cross-validated variance explained for 4 models on *target stimuli*. The table shows the leave-one-out cross validated coefficient of determination (R^2) for four data sets in three visual areas. The number of fitted model parameters (degrees of freedom) is indicated in parentheses for each model type (column 1). The number of stimuli comprising the target set are 18 for data set 1 (DS1) and data set 2 (DS2) and 17 for data set 3 (DS3), data set (DS4). The bold-faced font indicates the model with the highest variance explained in that column.

To summarize the greater response to snakes than data, we computed an index of (snakes – gratings) / mean (**Figure 8**). For the contrast energy and SOC model predictions, this value is close to 0. For the data and the NOA model, it is about 0.15. For the OTS model, it is about half of the values in the data and in the NOA model.

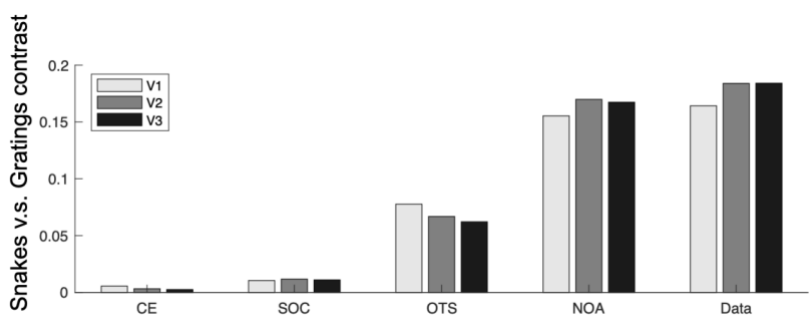


Figure 8. The NOA model accounts for the higher responses to snakes than gratings. For each visual area, V1-V3, we averaged the response to snakes and to gratings across stimuli in the target set and across the 4 data sets to yield one summary value for each class. We then computed a contrast of (snakes – gratings) / mean. In the data, there is a large difference between snakes and gratings (contrast > 0.6). This is matched in the NOA model predictions but not the predictions from other models. Data and model fits plotted from the function *s4_visualize*('figure8') in the code repository.

3.6. The NOA Model accurately predicts responses to a wide range of stimuli

The greater accuracy for the NOA model was not due to having more free parameters than the other models. The SOC, OTS, and NOA all have three free parameters, and the contrast energy model has two. Moreover, the prediction accuracy was summarized by *n*-fold cross validation, so that extra parameters do not necessarily lead to more accurate predictions. Nonetheless, because the NOA model was motivated to explain these data, it is important to validate it on other stimuli. We refit to model the full data sets shown to our four subjects, which consisted of 50 (data set 1), 48 (data set 2) and 39 (data set 3, data set 4) stimuli, spanning a variety of texture types. In addition to the snakes and gratings, there are textures we refer to as noise bars, waves, dust, and chess (**Supplementary Figure 1**).

Across the larger stimulus sets, the NOA model generalizes well. For data set 1, it explains between 48% and 51% of the variance in V1-V3 (**Figure 9**). The fits to the larger stimulus set, like the fits to the target stimuli alone, capture the observation about the two stimulus classes, meaning a larger predicted response for snakes than gratings. It also accurately predicts lower responses for waves than noise bars and snakes, which is a large effect in V1. One failure of the NOA model is that it predicts increasing response amplitude with texture density for the noise bars, whereas the observed responses do not increase with density. Across ROIs and data sets, the NOA model performs well, with R^2 ranging from 0.402 to 0.691 (**Table 2**, bottom row). Across the 4 models, it accounts for the highest variance in 4 of 4 V2 data sets and in 3 of 4 V3 data sets. It does not perform quite as well as the OTS model in V1, though it outperforms the contrast energy model and SOC.

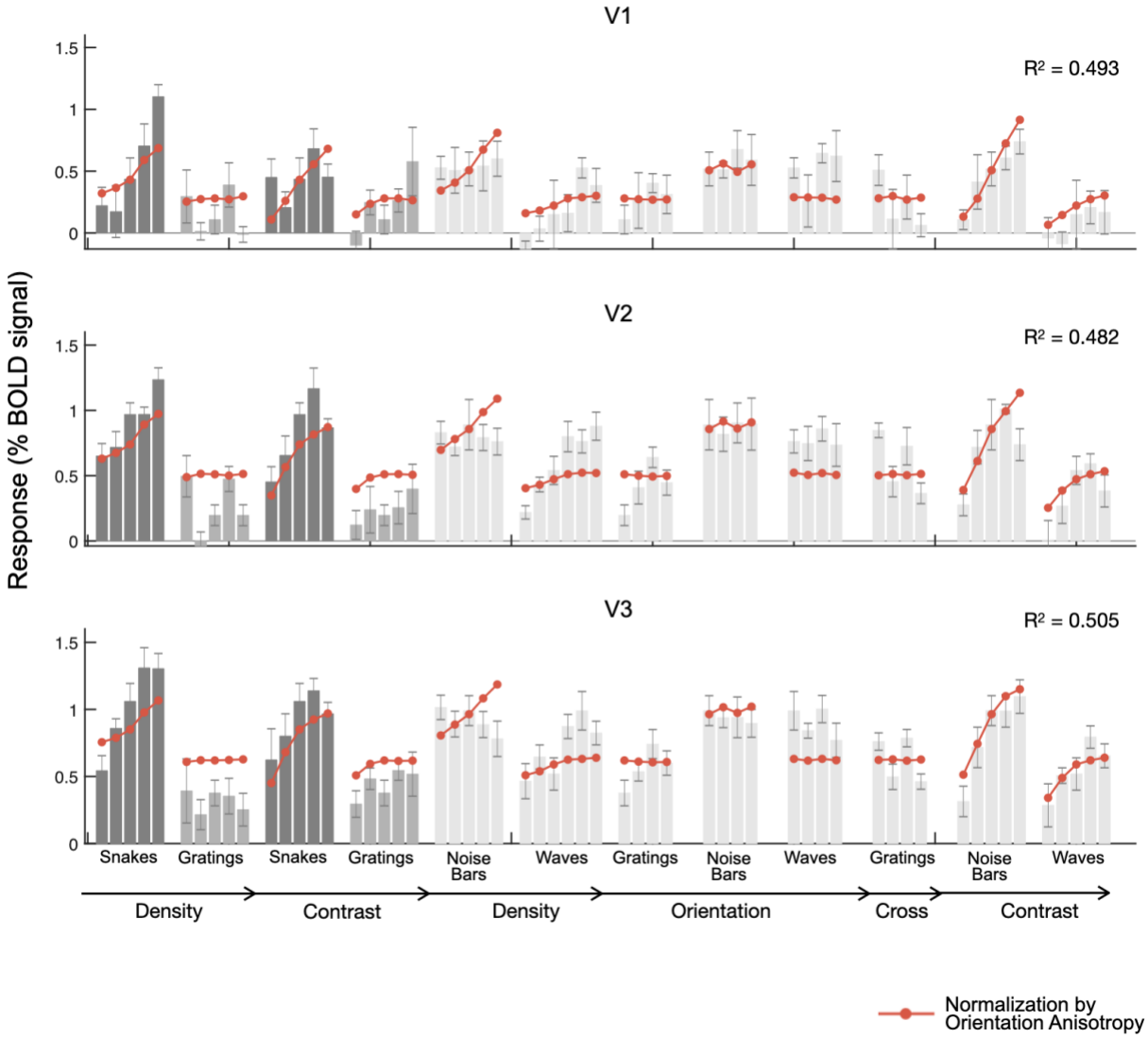


Figure 9. The NOA model fits a wide variety of stimuli. The mean fMRI responses from V1, V2, and V3 are shown for the full set of stimuli from data set 1 (See Supplementary Figure 1 for images). The data for snakes and gratings are replotted from figures 3-6, and again shown in dark and light gray. An even lighter gray is used for all other stimulus classes. The red dots are the cross-validated predictions from the NOA model. See Supplementary Figures

S3-S6 for similar plots for data sets 2-4. Data and model fits plotted from the function *s4_visualize*('figure9') in the code repository.

Model (free params)	**V1**				**V2**				**V3**			
	DS1	**DS2**	**DS3**	**DS4**	**DS1**	**DS2**	**DS3**	**DS4**	**DS1**	**DS2**	**DS3**	**DS4**
Contrast (2)	0.281	0.645	0.408	0.469	0.166	0.442	0.100	-0.004	0.134	0.260	0.093	0.013
SOC (3)	0.231	0.653	0.419	0.500	0.132	0.445	0.181	0.050	0.174	**0.433**	0.198	0.096
OTS (3)	0.460	**0.696**	**0.499**	**0.548**	0.280	0.527	0.206	0.090	0.275	0.296	0.159	0.078
NOA (3)	**0.493**	0.691	0.464	0.465	**0.482**	**0.553**	**0.519**	**0.536**	**0.505**	0.402	**0.439**	**0.464**

Table 2. Cross-validated variance explained for 4 models on the larger stimulus set. The table is organized in the same way as Table 1, differing only in the number of stimuli used to fit the models (50 for data set 1; 48 for data set 2; 39 for data set 3, data set 4).

4. Discussion

4.1 Why do existing models fail to account for the larger response to snake stimuli?

We began with the observation that the BOLD response is about twice as large for patterns with curved contours (snakes) than for similar stimuli with straight, parallel contours (gratings). The difference in the responses is not a peculiarity of the BOLD signal, as a similar pattern was also observed in human intracranial measures of the field potential (broadband power from ECoG electrodes). Three different models—a contrast energy model, a Second-order contrast model, and an Orientation-tuned surround model—all failed to predict the large differences between the responses to the two stimulus classes. Because the difference in response amplitude is large and systematic, it is unlikely to be explained by subtle confounds such as possible differences in contour density or contrast.

The reason why the contrast energy and SOC models do not predict systematic differences for snakes and gratings is relatively straightforward. Both models pool contrast energy over space and orientation without an orientation-specific normalization or other orientation specific non-linearity. If two images have the same total contrast energy, the contrast energy model output will be about the same. If they also have the same amount of second-order contrast (defined as spatial variation in contrast energy pooled over orientation), then the SOC output will also be about the same. Neither model is very sensitive to how the contrast energy is distributed across orientation channels.

We were more surprised by the failure of the OTS model to predict a large difference between the two stimulus types. The implementation of this model was motivated in part from electrophysiology results showing that in V1, surround suppressive fields tend to be elongated along the direction of the RF center's preferred orientation, and tend to be tuned to orientations close to the RF center's preferred orientation (Cavanaugh et al., 2002). Psychophysical experiments also show interactions between a target and surround that depend on matched orientation (e.g., Solomon et al., 1993). But the model was also motivated by the intuition that if the surround has orientation tuning matched to the center, then stimuli which are elongated in one direction (like gratings) will elicit strong surround suppression and hence small responses. In other words, we implemented the model specifically to account for the larger response to snakes than gratings. Nonetheless, the model generally does not do this: The results show that the OTS model predicts only a small difference in response amplitude between the snakes and

gratings. This appears to be due, in part, to the fact that in this model, the driven responses (numerator) and suppressive responses (denominator) tend to be correlated, such that normalization reduces the large responses at the preferred orientation while effectively enhancing response at other orientations. The result is that the total response, pooled across orientations and space, is not reduced much for gratings compared to snakes. And in fact, even this modest difference, evident when fit to the target stimuli (**Figure 6**), went away when fit to the larger stimulus sets (**Supplementary figure 6**). As part of our code repository, we include scripts that sweep out model parameters and visualize the intermediate outputs (responses for each orientation channel), showing that no combination of parameters leads to a very large difference between the stimulus classes (*s4_visualize('otsexplore'))*.

4.2 Why does the NOA model work? Model behavior.

The NOA model predicts larger outputs for snakes than gratings, even when the contrast energy is approximately matched. This is due to how the normalization is computed. The grating stimuli elicit large outputs in the orientation channels that are matched to the stimulus, moderate outputs in adjacent orientation channels, and little to no response in other channels. As a result, there is high anisotropy (variance across channel outputs). The outputs for the snake stimuli are similar across orientation channels, meaning less anisotropy. Because the NOA model normalizes (divides) the contrast energy by variance across orientation channels, the stimulus with higher variance receives more suppression. We illustrate this idea by showing the channel outputs, pooled across space, for several stimuli (**Figure 10)**. The contrast energy in the grating has relatively high orientation variance across channel outputs. Consequently, the normalized outputs are greatly diminished (first row, right vs left). A snake stimulus has similar contrast energy summed across different orientation channels (0.80 for the snake compared to 0.73 for the grating), but the variance is about 50 times lower, hence there is little normalization. This means that for the snake stimulus, the normalized and unnormalized outputs are nearly the same. After normalization, the summed output is about 3 times higher for the curve than the grating.

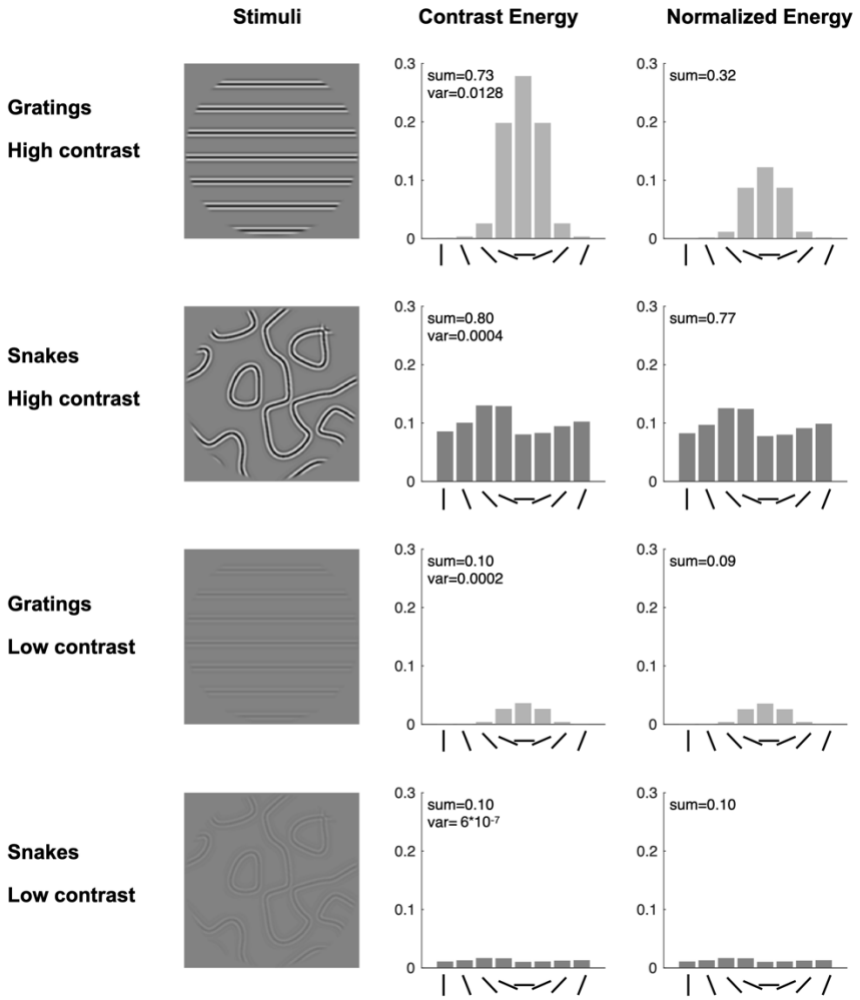


Figure 10. Model behavior. We illustrate the behavior of a simplified version of the NOA model for 4 stimuli. For a high contrast grating (top row), the contrast energy is high in channels where the channel orientation matches (or is close to) the stimulus orientation, and low for other channels. As a result, there is high anisotropy, and the variance in the channel outputs is high. This results in strong normalization, reducing the outputs substantially in all channels (right column). For all other stimuli, the anisotropy is lower, either because the stimulus has power at many orientations (snakes) or because the overall contrast is low. As a result, there is little effect of normalization in the other three rows (summed output across orientations is similar before and after normalization). For simplicity, we omitted the squaring on *E* in the simulations, as well as the exponent on the overall output. Data and model fits plotted from the function *s4_visualize*('figure10') in the code repository.

The normalization in the model also makes the surprising prediction that the response should increase more with contrast for snakes than for gratings. Without normalization or an output non-linearity, the contrast response function would be linear for any stimulus. But with normalization, the shape of the contrast response function is stimulus dependent. For the grating stimulus, at high contrast the variance is high and the normalization is high, whereas at low contrast, the variance is low and there is little suppression. Because there is strong normalization for only one of the two stimuli, the difference in the outputs is compressed. Specifically, before normalization, the response difference between a high and low contrast stimulus is about 10x (row 1 vs row 3). After normalization, the difference is only about 4x. For the snakes, the variance is low irrespective of contrast. Hence there is little normalization, and the contrast response function has a greater slope. This general pattern is borne out in the model fits and in

the data. For example, in figure S2a, for the contrast stimuli, the data and the predicted snakes are steeper for the curved stimuli and more compressive for the gratings. The same logic also applies to the variation in density: for the grating stimuli, the predicted responses are compressed, meaning there is relatively small difference in the predictions for dense and sparse gratings, but large differences in the predictions for dense and sparse snake stimuli. We see this in the data as well as the model predictions (**Figure S2a**). The other models do not make this prediction.

Interestingly, the denominator in the NOA model (the normalizer) is almost the same as the numerator of an orientation variance model used to predict the amplitude of gamma oscillations measured with ECoG electrodes in human subjects (Hermes et al., 2019). The fact that the same term is used in the numerator to predict gamma oscillations and the denominator to predict BOLD is consistent with the empirical observation that many of the stimuli that are most effective for driving large gamma oscillations (high contrast luminance gratings) are relatively ineffective for eliciting BOLD signals (Butler et al., 2017; Hermes et al., 2017b), and multiunit action potentials (Hermes et al., 2017a; Ray and Maunsell, 2011). The link may be that gamma oscillations, rather than being the fundamental mechanism for perception and long-range cortical communication (Engel and Singer, 2001; Fries, 2005), are rather a result of the normalization process (Hermes et al., 2019; Ray et al., 2013).

4.3 The NOA model has the largest advantage in extrastriate maps

The normalization by orientation anisotropy model showed the largest advantage over other models for V2 and V3. There seem to be two reasons for this. First, the difference in BOLD response between snakes and gratings is larger in V2 and V3 than in V1, and only the NOA model predicts this difference (**Figure 8**). However, this is only a small effect, since all three areas show larger responses to snakes than gratings. The second reason is due to an interaction between the contrast response function and stimulus class. In V2 and V3, the contrast response function is relatively flat for gratings, but steep for other stimulus classes (snakes as well as circular patterns and plaids). As described in sections 3.5 and 4.2, the NOA model predicts a flatter contrast response function for gratings than for other stimulus classes, matching the data in V2 and V3. The other three models, however, do not show a strong interaction between stimulus class and the steepness of the contrast response function and hence cannot capture the pattern of data in V2 and V3.

The fact that the NOA model shows a large advantage for extrastriate areas, whereas the contrast energy model and the other two models do not, is consistent with findings from electrophysiology and computational modeling of behavior. In particular, a series of studies suggests that extrastriate maps, especially V2, are more sensitive to higher-order image statistics than to contrast energy per se (Freeman and Simoncelli, 2011; Freeman et al., 2013; Movshon and Simoncelli, 2015; Okazawa et al., 2017). The NOA model is also sensitive to higher-order image statistics, namely orientation anisotropy (variance across orientation channel outputs). The NOA model differs from these physiology models in that it predicts population data (MRI voxels containing many thousands of neurons), and it is image-computable; the physiology models predict which image pairs will produce similar responses, but not what the response will be for a single image. More recent V2 models for single units have begun to make forward predictions for response of V2 neurons, and it will be interesting to see whether they bear any relation to our NOA model (Oleskiw et al., 2021).

We consider one way that our NOA model might be linked to the V2 models proposed by Simoncelli and colleagues. Suppose the V1 population computes contrast energy localized in orientation and space. (For simplicity, we ignore spatial frequency, as we did here experimentally by band-pass filtering our stimuli.) We then suppose that the V2 cells compute various weighted sums of the V1 outputs. Specifically, we

assume that, for each spatial location, the weights among V2 cells form a Fourier basis set on the V1 outputs (across orientation). If these weights are arranged in pairs (similar to the odd and even V1 filters), and the outputs are squared and summed across phase (similar to the V1 energy model), then then the summed V2 population output will be proportional to the variance in the V1 response across orientation (David Heeger, personal communication). If this population output is the normalization pool for V2, then we get a model like NOA, which normalizes by the variance across orientation channels. Feedback to V1, or a related calculation in V1 might result in similar behavior even for V1.

4.4 Are standard V1 models missing something important?

No model is complete. The standard normalized contrast energy model of V1, when fit with appropriate parameters, can capture substantial variance in V1 responses (Carandini et al., 1997), but not all. Our results indicate that the model fails for at least some relatively simple stimulus classes, and that the failure can be large. But whether a more complex model is needed, such as a gated normalization model (Coen-Cagli et al., 2015) or the NOA model here, will depend on the stimulus set tested and the purpose of modeling. It is reasonably likely that the computations in the NOA model are more connected to computations in extrastriate visual maps than in V1, but may also influence the response in V1 via feedback.

More generally, since the development of divisive normalization models in the early 1990s, there has not yet been a generally agreed up description of exactly what cell populations (and with what weights) contribute to normalization. Some attempts have been made from assuming efficient coding of natural images {Schwartz, 2001 #2320} or from fitting model parameters to neural data {Burg, 2021 #2980}. In this sense, a standard model of V1, with parameters set, that is downloadable and executable on arbitrary input images, does not yet really exist.

The NOA model we presented makes some advance but also has some important limits, most notably the lack of spatial receptive fields, as well as a lack of sensitivity to second-order contrast. Hence it does not supersede other models, but rather provides a compact computational summary of response patterns that are not well captured by other models. Integrating better computational tools for validation (Lerma-Usabiaga et al., 2020), model-based stimulus development (Golan et al., 2020; Wang and Simoncelli, 2008), high-quality standardized datasets (Allen et al., 2021), and theory (Simoncelli and Olshausen, 2001), may offer the best path toward more complete understanding of neural circuits in visual cortex.

References

Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2, 284-299.

Allen, E.J., St-Yves, G., Wu, Y., Breedlove, J.L., Dowdle, L.T., Caron, B., Pestilli, F., Charest, I., Hutchinson, J.B., and Naselaris, T. (2021). A massive 7T fMRI dataset to bridge cognitive and computational neuroscience. bioRxiv.

Brouwer, G.J., and Heeger, D.J. (2011). Cross-orientation suppression in human visual cortex. J Neurophysiol 106, 2108-2119.

Butler, R., Bernier, P.M., Lefebvre, J., Gilbert, G., and Whittingstall, K. (2017). Decorrelated Input Dissociates Narrow Band gamma Power and BOLD in Human Visual Cortex. J Neurosci 37, 5408-5418.

Carandini, M., Demb, J.B., Mante, V., Tolhurst, D.J., Dan, Y., Olshausen, B.A., Gallant, J.L., and Rust, N.C. (2005). Do we know what the early visual system does? J Neurosci 25, 10577-10597.

Carandini, M., Heeger, D.J., and Movshon, J.A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17, 8621-8644.

Cavanaugh, J.R., Bair, W., and Movshon, J.A. (2002). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. J Neurophysiol 88, 2547-2556.

Coen-Cagli, R., Kohn, A., and Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. Nat Neurosci 18, 1648-1655.

David, S.V., Vinje, W.E., and Gallant, J.L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. J Neurosci 24, 6991-7006.

Engel, A.K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. Trends Cogn Sci 5, 16-25.

Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral Cortex 7, 181-192.

Freeman, J., and Simoncelli, E.P. (2011). Metamers of the ventral stream. Nat Neurosci 14, 1195-1201.

Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., and Movshon, J.A. (2013). A functional and perceptual signature of the second visual area in primates. Nat Neurosci 16, 974-981.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci 9, 474-480.

Golan, T., Raju, P.C., and Kriegeskorte, N. (2020). Controversial stimuli: Pitting neural networks against each other as models of human cognition. Proc Natl Acad Sci U S A 117, 29330-29337.

Heeger, D.J. (1992a). Half-squaring in responses of cat striate cells. Vis Neurosci 9, 427-443.

Heeger, D.J. (1992b). Normalization of cell responses in cat striate cortex. Vis Neurosci 9, 181-197.

Heeger, D.J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. J Neurophysiol 70, 1885-1898.

Hermes, D., Kasteleijn-Nolst Trenite, D.G.A., and Winawer, J. (2017a). Gamma oscillations and photosensitive epilepsy. Curr Biol 27, R336-R338.

Hermes, D., Nguyen, M., and Winawer, J. (2017b). Neuronal synchrony and the relation between the blood-oxygen-level dependent response and the local field potential. PLoS Biol 15, e2001461.

Hermes, D., Petridou, N., Kay, K.N., and Winawer, J. (2019). An image-computable model for the stimulus selectivity of gamma oscillations. Elife 8.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160, 106-154.

Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature 452, 352-355.

Kay, K.N., Rokem, A., Winawer, J., Dougherty, R.F., and Wandell, B.A. (2013a). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Front Neurosci 7, 247.

Kay, K.N., Winawer, J., Mezer, A., and Wandell, B.A. (2013b). Compressive spatial summation in human visual cortex. J Neurophysiol 110, 481-494.

Kay, K.N., Winawer, J., Rokem, A., Mezer, A., and Wandell, B.A. (2013c). A two-stage cascade model of BOLD responses in human visual cortex. PLoS Comput Biol 9, e1003079.

Landy, M.S., and Graham, N. (2004). Visual Perception of Texture. The visual neurosciences, 1106.

Larsson, J., Landy, M.S., and Heeger, D.J. (2006). Orientation-selective adaptation to first- and second-order patterns in human visual cortex. J Neurophysiol 95, 862-881.

Lerma-Usabiaga, G., Benson, N., Winawer, J., and Wandell, B.A. (2020). A validation framework for neuroimaging software: The case of population receptive fields. PLoS Comput Biol 16, e1007924.

Mannion, D.J., Kersten, D.J., and Olman, C.A. (2015). Scene coherence can affect the local response to natural images in human V1. Eur J Neurosci 42, 2895-2903.

Movshon, J.A., and Simoncelli, E.P. (2015). Representation of Naturalistic Image Structure in the Primate Visual Cortex. Cold Spring Harb Symp Quant Biol.

Okazawa, G., Tajima, S., and Komatsu, H. (2017). Gradual Development of Visual Texture-Selective Properties Between Macaque Areas V2 and V4. Cereb Cortex 27, 4867-4880.

Oleskiw, T.D., Diaz-Pacheco, R.R., Movshon, J.A., and Simoncelli, E.P. (2021). A two-stage model of V2 demonstrates efficient higher-order feature representation. Journal of Vision 21, 2654-2654.

Olshausen, B.A., and Field, D.J. (2005). How close are we to understanding v1? Neural Comput 17, 1665-1699.

Olsson, C., Kay, K., and Winawer, J. (2015). Orientation-tuned surround suppression improves computational models of human visual cortex. Journal of Vision 15, 1001-1001.

Pollen, D.A., and Ronner, S.F. (1983). Visual cortical neurons as localized spatial frequency filters. IEEE Transactions on Systems, Man, and Cybernetics, 907-916.

Qiu, C., Burton, P.C., Kersten, D., and Olman, C.A. (2016). Responses in early visual areas to contour integration are context dependent. J Vis 16, 19.

Ray, S., and Maunsell, J.H. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol 9, e1000610.

Ray, S., Ni, A.M., and Maunsell, J.H. (2013). Strength of gamma rhythm depends on normalization. PLoS Biol 11, e1001477.

Rust, N.C., Schwartz, O., Movshon, J.A., and Simoncelli, E.P. (2005). Spatiotemporal elements of macaque v1 receptive fields. Neuron 46, 945-956.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889-893.

Shooner, C., Hallum, L.E., Kumbhani, R.D., Ziemba, C.M., Garcia-Marin, V., Kelly, J.G., Majaj, N.J., Movshon, J.A., and Kiorpes, L. (2015). Population representation of visual information in areas V1 and V2 of amblyopic macaques. Vision Res.

Simoncelli, E.P., and Olshausen, B.A. (2001). Natural image statistics and neural representation. Annu Rev Neurosci 24, 1193-1216.

Solomon, J.A., Sperling, G., and Chubb, C. (1993). The lateral inhibition of perceived contrast is indifferent to on-center/off-center segregation, but specific to orientation. Vision Res 33, 2671-2683.

Tang, S., Lee, T.S., Li, M., Zhang, Y., Xu, Y., Liu, F., Teo, B., and Jiang, H. (2018). Complex Pattern Selectivity in Macaque Primary Visual Cortex Revealed by Large-Scale Two-Photon Imaging. Curr Biol 28, 38-48 e33.

Vinje, W.E., and Gallant, J.L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287, 1273-1276.

Wandell, B.A. (1995). Foundations of vision (Sunderland, Mass.: Sinauer Associates).

Wandell, B.A., and Winawer, J. (2011). Imaging retinotopic maps in the human brain. Vision Res 51, 718-737.

Wang, Z., and Simoncelli, E.P. (2008). Maximum differentiation (MAD) competition: a methodology for comparing computational models of perceptual quantities. J Vis 8, 8 1-13.

Winawer, J., Kay, K.N., Foster, B.L., Rauschecker, A.M., Parvizi, J., and Wandell, B.A. (2013). Asynchronous broadband signals are the principal source of the BOLD response in human visual cortex. Curr Biol 23, 1145-1153.

Zhou, J., Benson, N.C., Kay, K.N., and Winawer, J. (2018). Compressive temporal summation in human visual cortex. J Neurosci 38, 691-709.

Supplementary figures

Table S1: Data set properties for all experiments

	url	Number of fMRI runs	V1 voxel count	V2 voxel count	V3 voxel count	voxel selection
DS1	https://osf.io/nm58r/	12	606	545	572	cross-validated $R^2 > 3\%$
DS2	https://osf.io/nxf65/	12	454	437	377	cross-validated $R^2 > 3\%$
DS3	https://osf.io/953yq/	24	202	283	215	cross-validated $R^2 > 0\%, mean(\beta) > 0$
DS4	https://osf.io/953yq/	12	189	163	227	cross-validated $R^2 > 0\%, mean(\beta) > 0$

Table S2: Stimulus set properties for all experiments

	url	Size (deg)	Size (pixels, shown to subjects)	Size (pixels, for analysis)	Spatial frequency (cpd)	Number of stimulus classes
DS1	https://osf.io/mbtf6/	12.5 x 12.5	400 x 400	150 x 150	3	50
DS2	https://osf.io/juya8/	18.75 x 18.75	600 x 600	225 x 225	3	48
DS3	https://osf.io/d2r8n/	12.5 x 12.5	800 x 800	150 x 150	3	39
DS4	https://osf.io/d2r8n/	12.5 x 12.5	800 x 800	150 x 150	3	39

Table S3: Stimulus set group for all experiments

	This paper		Kay et al 2013	
	OSF url		http://kendrickkay.net/socmodel/stimuli.mat	
	Number	**Description**	**Number**	**Description**
DS1	1 2 3 4 5	SNAKES (density)		
	47 48 3 49 50	SNAKES (contrast)		
	6 7 8 9 10	GRATINGS (density)		
	35 36 8 37 38	GRATINGS (contrast)		
	11 12 13 14 15	NOISE BARS (density)		
	16 17 18 19 20 21	WAVES (density)		
	8 22 23 24	GRATINGS (orientation)		
	13 25 26 27	NOISE BARS (orientation)		
	20 28 29 30	WAVES (orientation)		
	31 32 33 34	GRATINGS (spacing)		
	39 40 13 41 42	NOISE BARS (contrast)		
	43 44 18 45 46	WAVES (contrast)	Not part of Kay et al.	
DS2	1 2 3 4 5	SNAKES (density)		
	45 46 2 47 48	SNAKES (contrast)		
	6 7 8 9 10	GRATINGS (density)		
	33 34 7 35 36	GRATINGS (contrast)		
	11 12 13 14 15	NOISE BARS (density)		
	16 17 18 19	WAVES (density)		
	8 20 21 22	GRATINGS (orientation)		
	13 23 24 25	NOISE BARS (orientation)		
	18 26 27 28	WAVES (orientation)		
	29 30 31 32	GRATINGS (cross)		
	37 38 13 39 40	NOISE BARS (contrast)		
	41 42 18 43 44	WAVES (contrast)	Not part of Kay et al.	
DS3, DS4	36 37 38 39	SNAKES (density)	181-184	SEPARATION
	21 22 23 24 25 26 27 28 29 30	SNAKES (contrast)	159-168	CONTRAST

31 32 33 34 35	GRATINGS (density)	176-180	not used
9 10 11 12 1	GRATINGS (contrast)	147-150	GRATING
1 2 3 4 5 6 7 8	GRATINGS (orientation)	139-146	ORIENTATION
13 14 15 16	PLAID (contrast)	151-154	PLAID
17 18 19 20	CIRCULAR (contrast)	155-158	CIRCULAR
NA	not used	70-138	SPACE
NA	not used	169-175	not used (objects/scenes)
NA	not used	185-225	not used (mixtures)

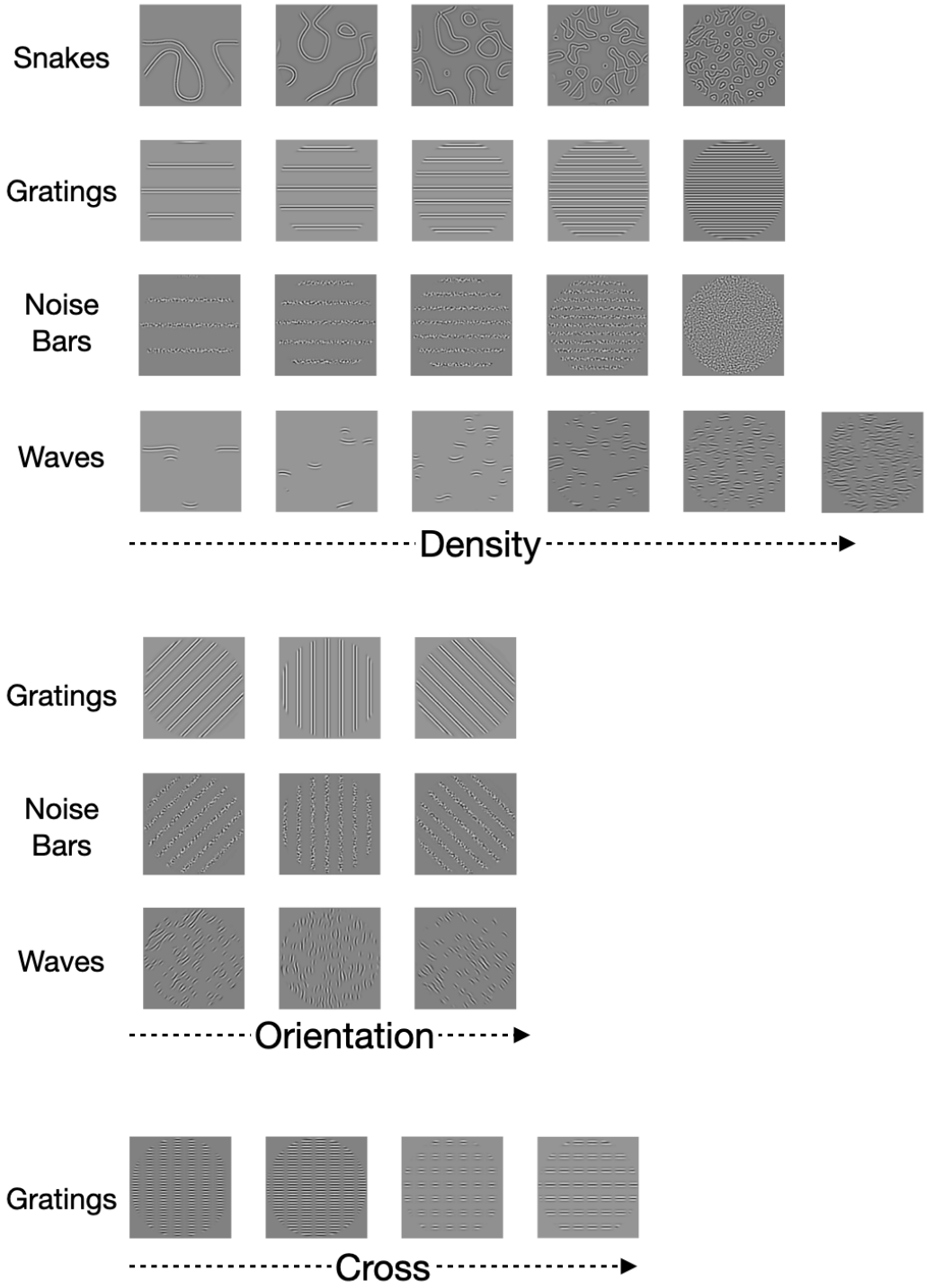
TableS4a: Parameters for cross-validated fits for all stimuli

		V1				V2				V3			
		DS1	DS2	DS3	DS4	DS1	DS2	DS3	DS4	DS1	DS2	DS3	DS4
Contrast energy	g	1.490	3.430	2.120	3.140	1.300	1.560	1.790	2.320	1.270	1.330	1.750	2.580
	n	0.406	0.289	0.221	0.285	0.230	0.202	0.100	0.197	0.174	0.144	0.088	0.158
Second-order contrast	c	0.111	0.123	0.235	0.353	0.944	0.944	1.000	1.000	0.944	0.985	1.000	1.000
	g	1.280	3.170	2.000	3.140	1.470	1.650	2.060	2.750	1.420	1.620	2.140	3.350
	n	0.216	0.159	0.116	0.157	0.144	0.118	0.072	0.125	0.111	0.107	0.076	0.120
Orientation-tuned surround	w	357.0	82.90	403.0	403.0	403.0	348.0	403.0	403.0	403.0	168.0	399.0	403.0
	g	233.0	9.680	67.00	117.0	34.80	10.50	10.30	50.40	18.40	3.290	6.820	30.20
	n	1.200	0.456	0.786	0.853	0.730	0.485	0.389	0.699	0.586	0.287	0.308	0.556
Normalization by orientation anisotropy	w	152.0	63.60	222.0	406.0	681.0	135.0	542.0	782.0	678.0	340.0	766.0	859.0
	g	23.20	6.390	12.70	23.30	77.90	4.130	11.80	36.00	40.40	6.820	11.60	38.60
	n	0.491	0.214	0.280	0.291	0.451	0.200	0.212	0.306	0.373	0.220	0.197	0.287

TableS4b: Parameters for cross-validated fits for target stimuli

		V1				V2				V3			
		DS1	DS2	DS3	DS4	DS1	DS2	DS3	DS4	DS1	DS2	DS3	DS4
Contrast energy	g	1.760	3.860	1.840	2.440	1.430	1.930	1.500	1.390	1.380	1.620	1.530	1.760
	n	0.438	0.301	0.209	0.231	0.224	0.239	0.092	0.088	0.171	0.186	0.086	0.093
Second-order contrast	c	0.020	0.004	1.000	1.000	0.960	0.979	1.000	1.000	0.986	1.000	1.000	1.000
	g	1.490	3.440	2.210	3.010	1.520	2.020	1.730	1.620	1.610	1.890	1.780	2.130
	n	0.240	0.162	0.125	0.139	0.136	0.137	0.063	0.064	0.119	0.125	0.062	0.071
Orientation-tuned surround	w	358.0	51.90	92.30	57.50	297.0	84.10	128.0	183.0	234.0	78.70	94.60	100.0
	g	100.0	8.540	8.070	8.420	10.600	5.190	3.430	5.030	5.010	4.320	2.770	4.060
	n	1.020	0.445	0.497	0.485	0.525	0.419	0.247	0.333	0.350	0.364	0.198	0.258
Normalization by orientation anisotropy	w	117.0	38.90	30.50	18.00	191.0	50.10	202.0	349.0	197.0	79.50	263.0	324.0
	g	6.200	5.540	4.020	4.540	4.460	2.910	5.630	14.70	3.810	2.570	5.350	14.20
	n	0.344	0.196	0.198	0.197	0.219	0.168	0.172	0.264	0.182	0.143	0.156	0.239

Figure S1: Stimuli for data set 1



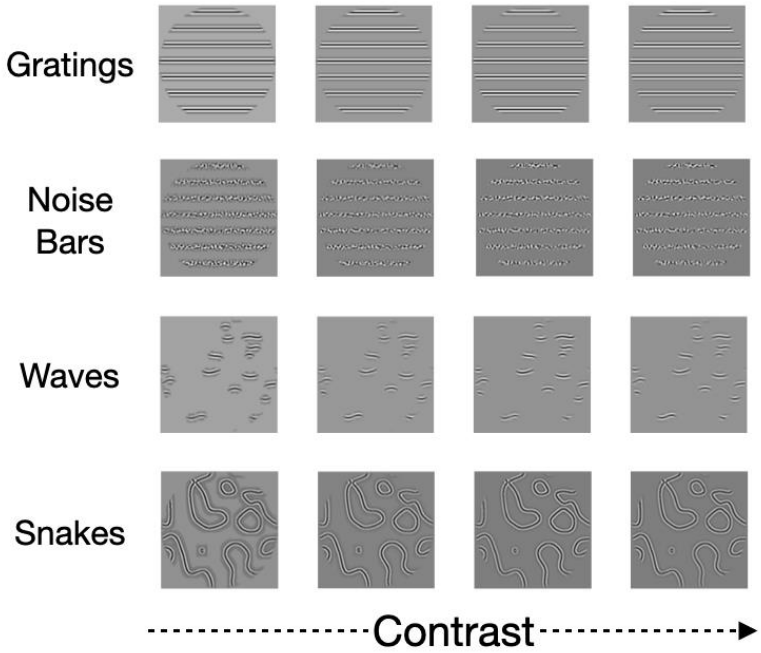

Gratings
Noise
Bars
Waves
Snakes
Contrast

Figure S2a: Responses and NOA model fits for target stimuli, all datasets

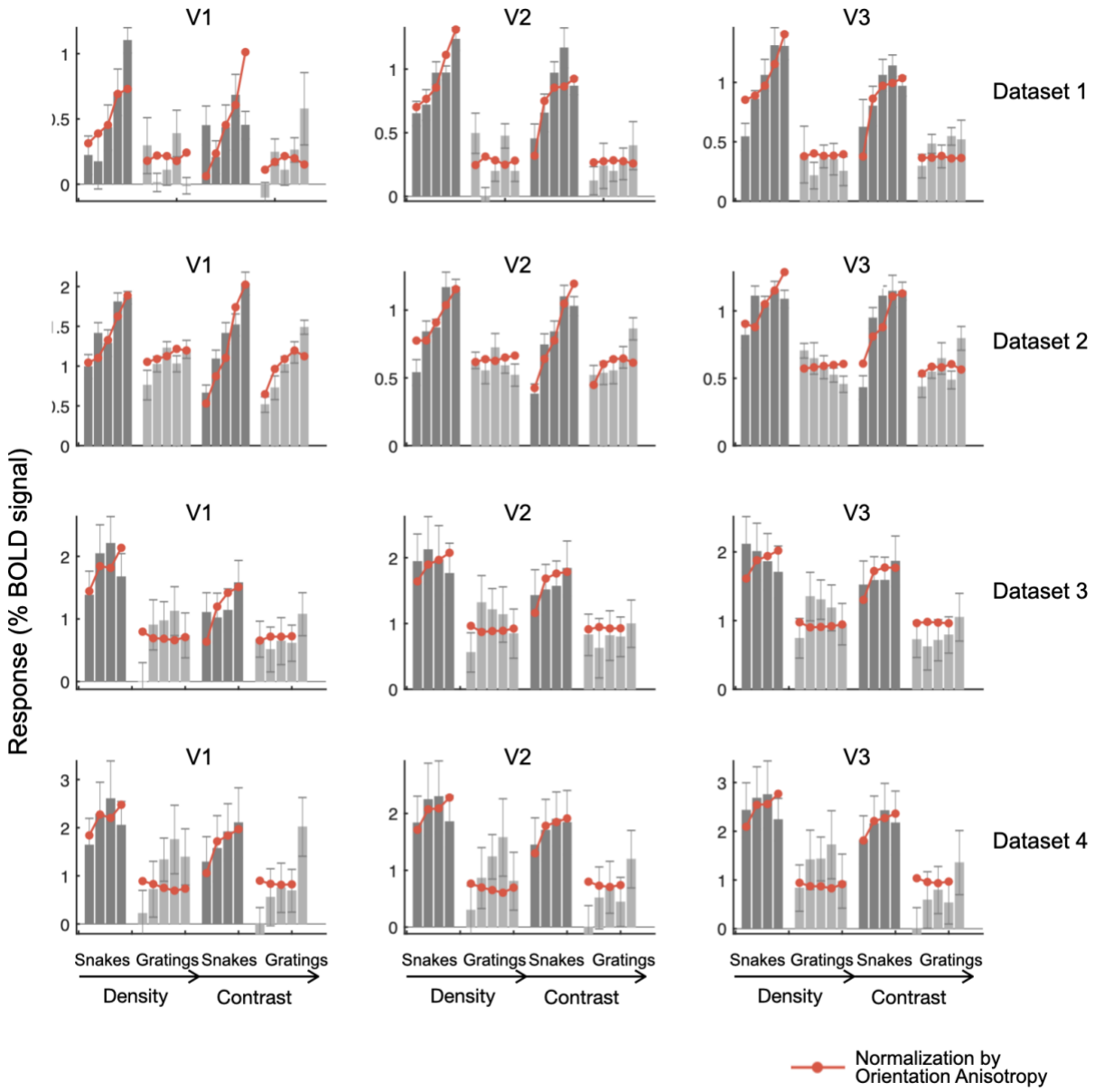


Figure S2b: Responses and contrast energy model fits for target stimuli, all datasets

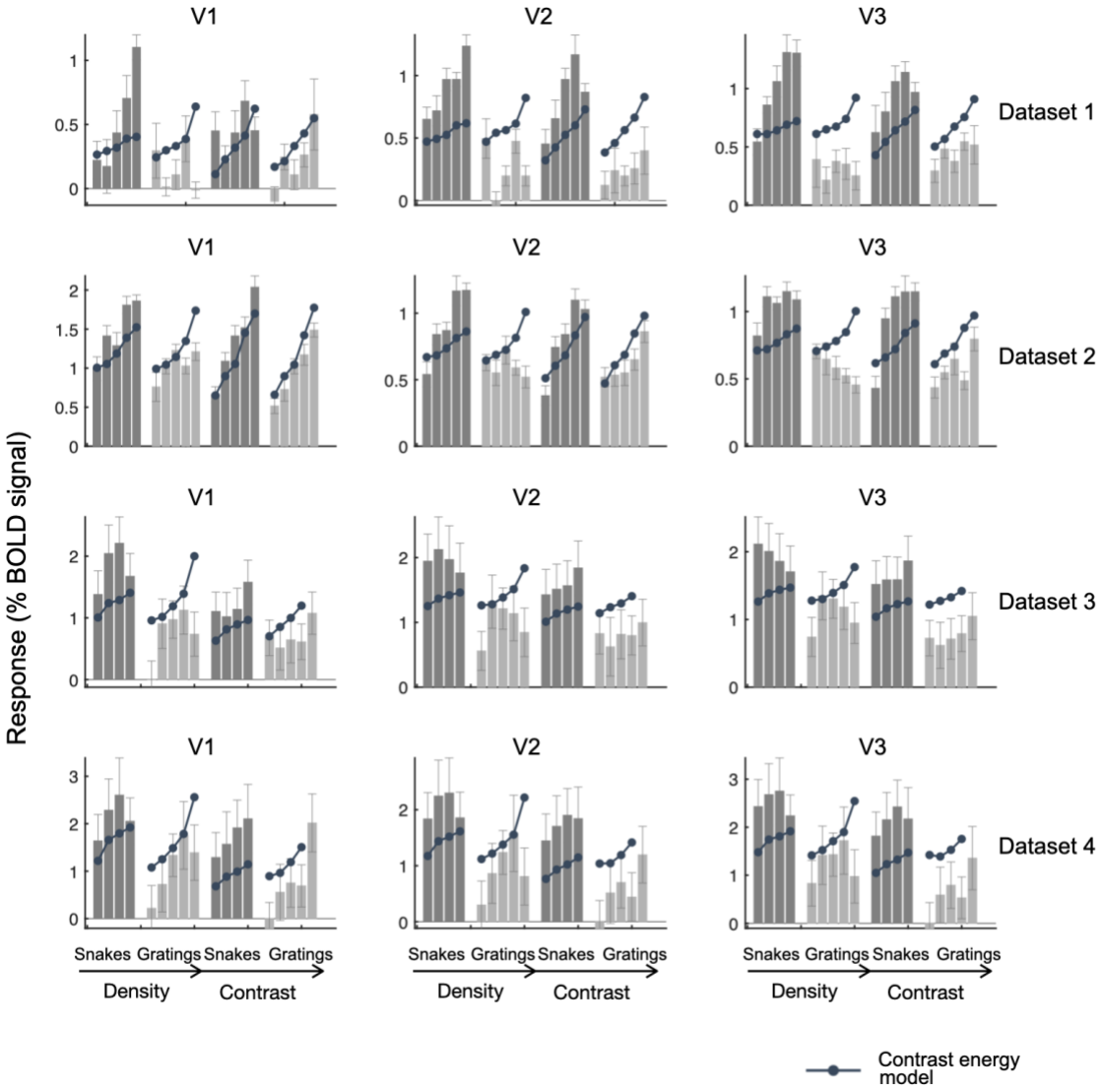


Figure S2c: Responses and Second-order contrast model fits for target stimuli, all datasets

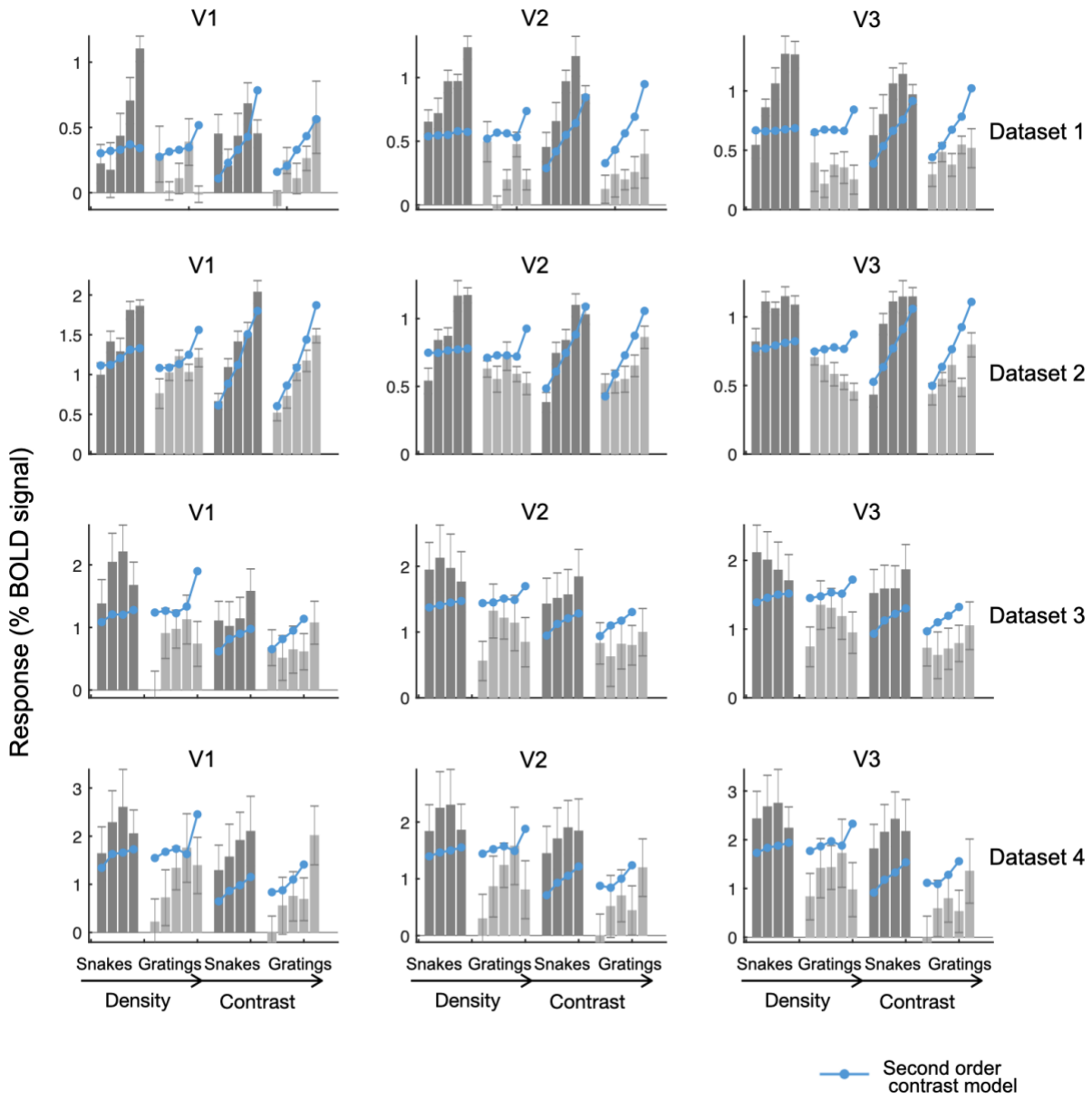


Figure S2d: Responses and Orientation-tuned surround model fits for target stimuli, all datasets

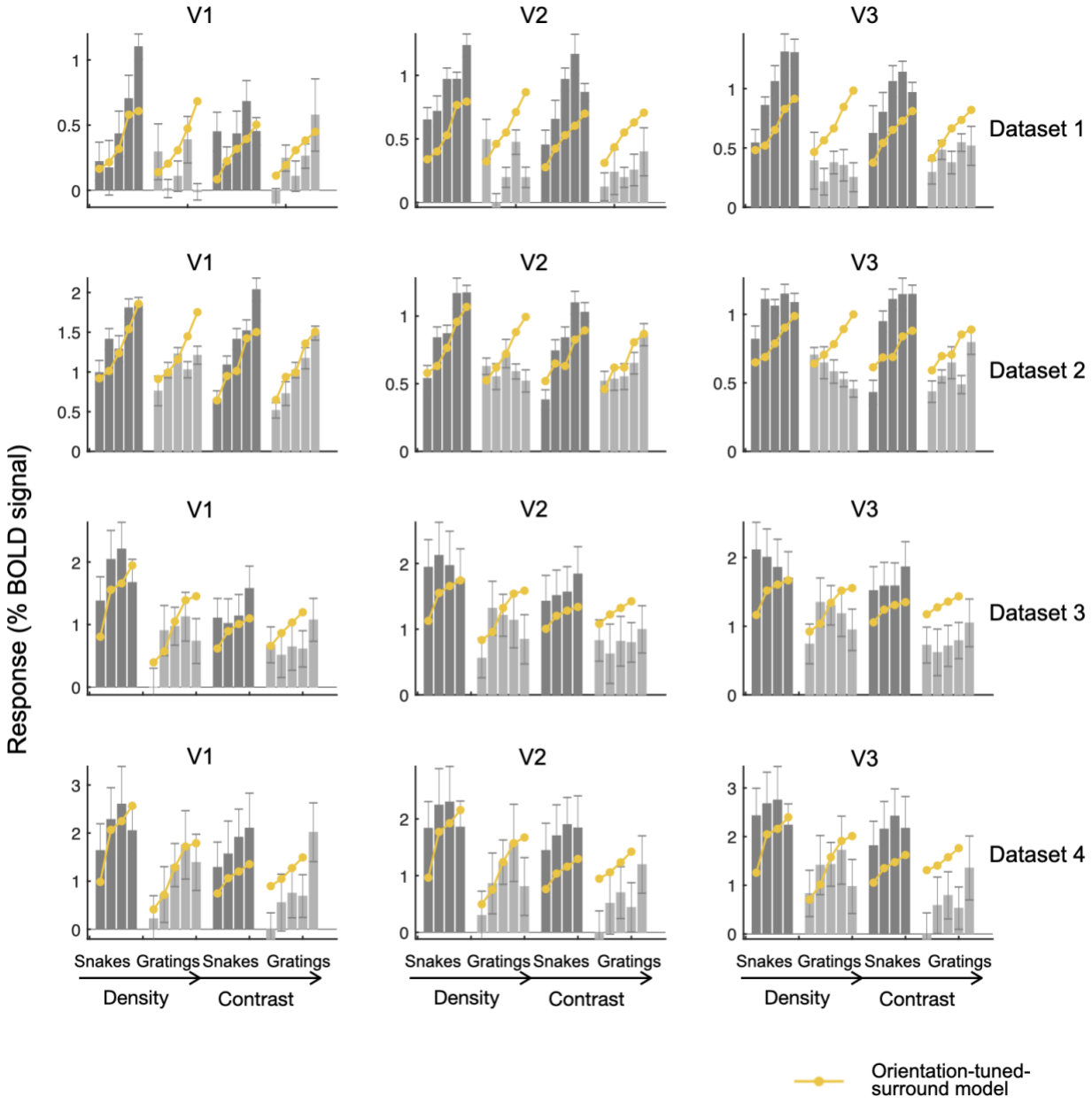


Figure S3a: Responses and normalization by orientation anisotropy model fits for all stimuli, data set 2

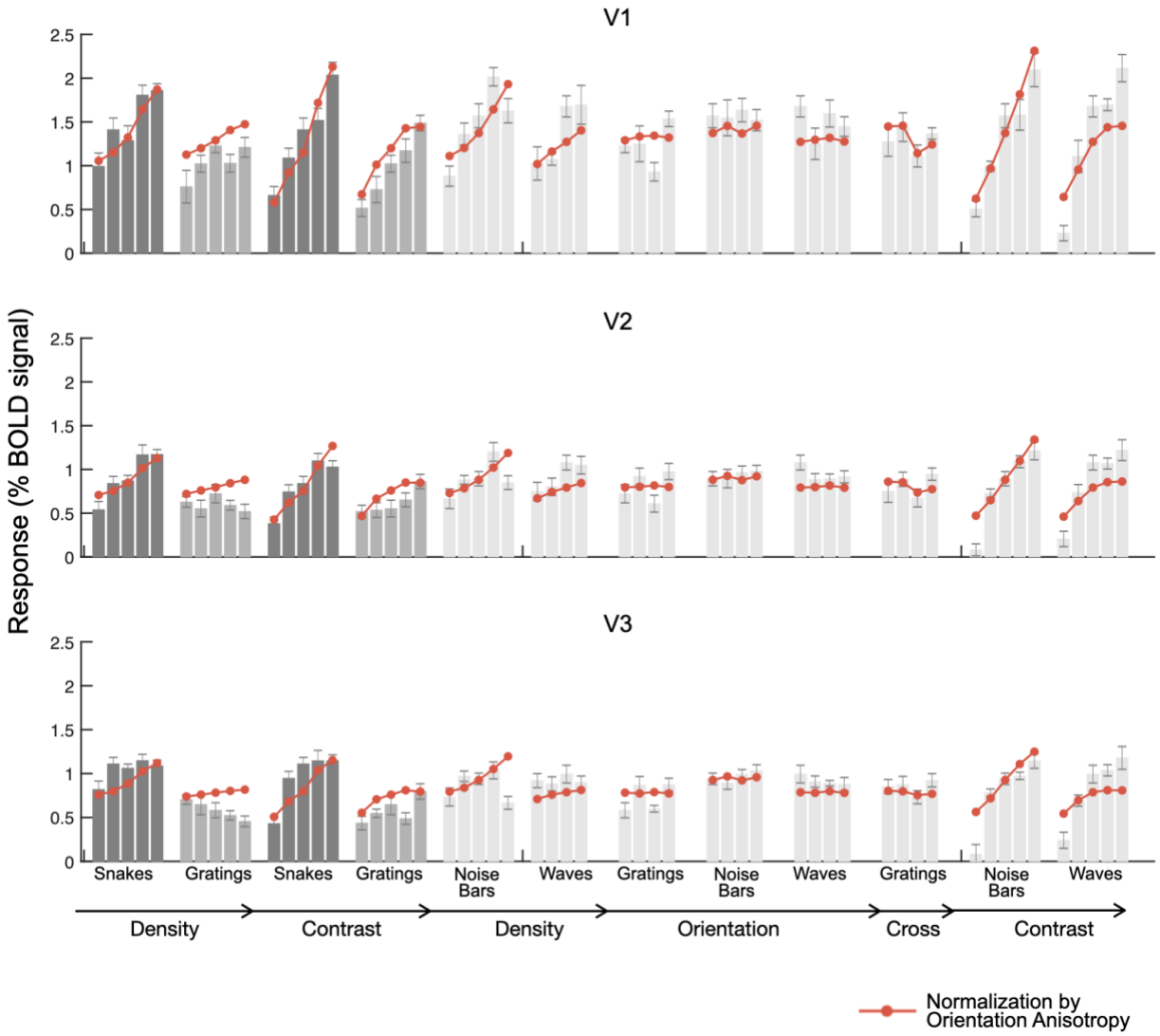


Figure S3b: Responses and NOA model fits for all stimuli, data set 3

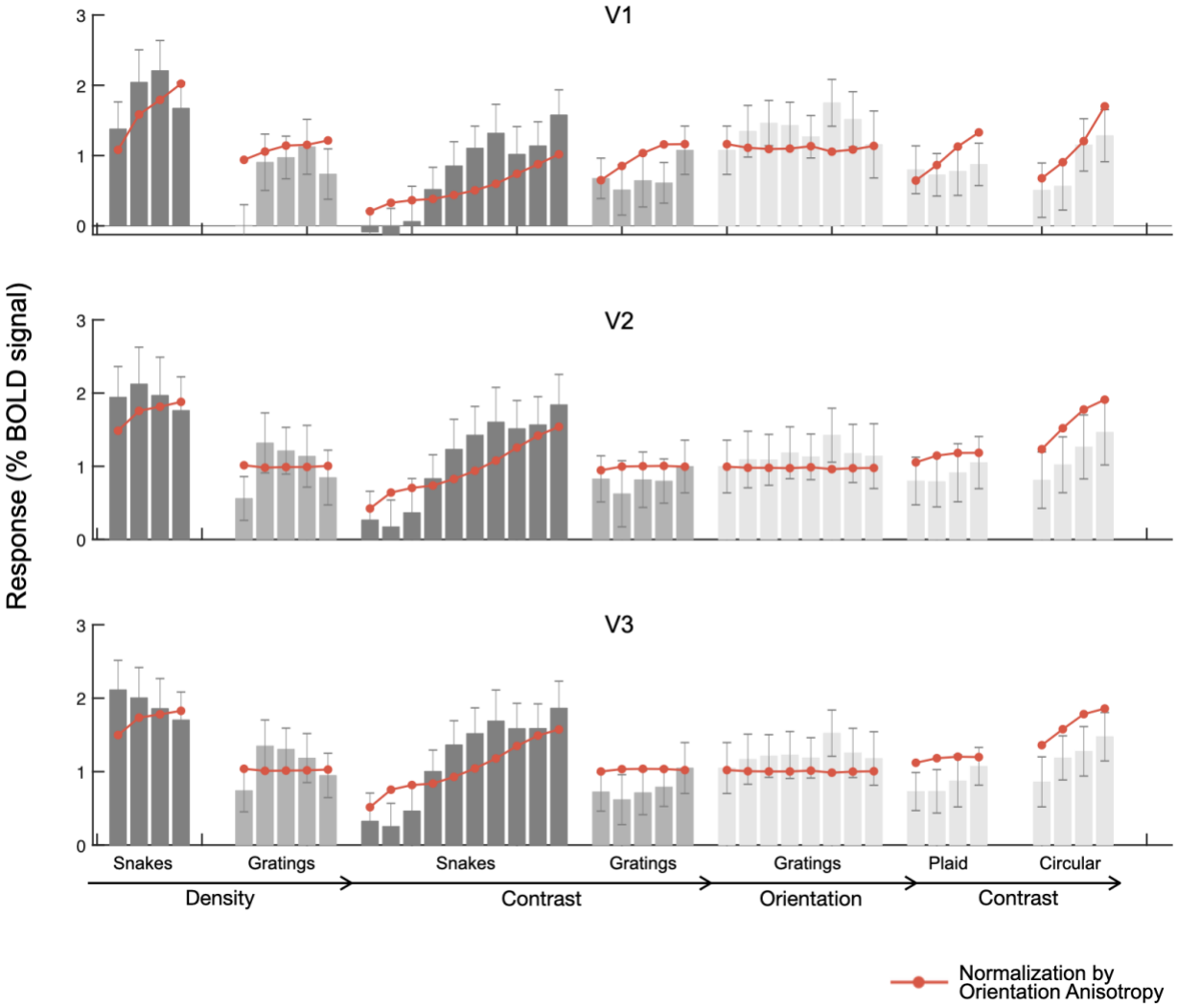


Figure S3c: Responses and NOA model fits for all stimuli, data set 4

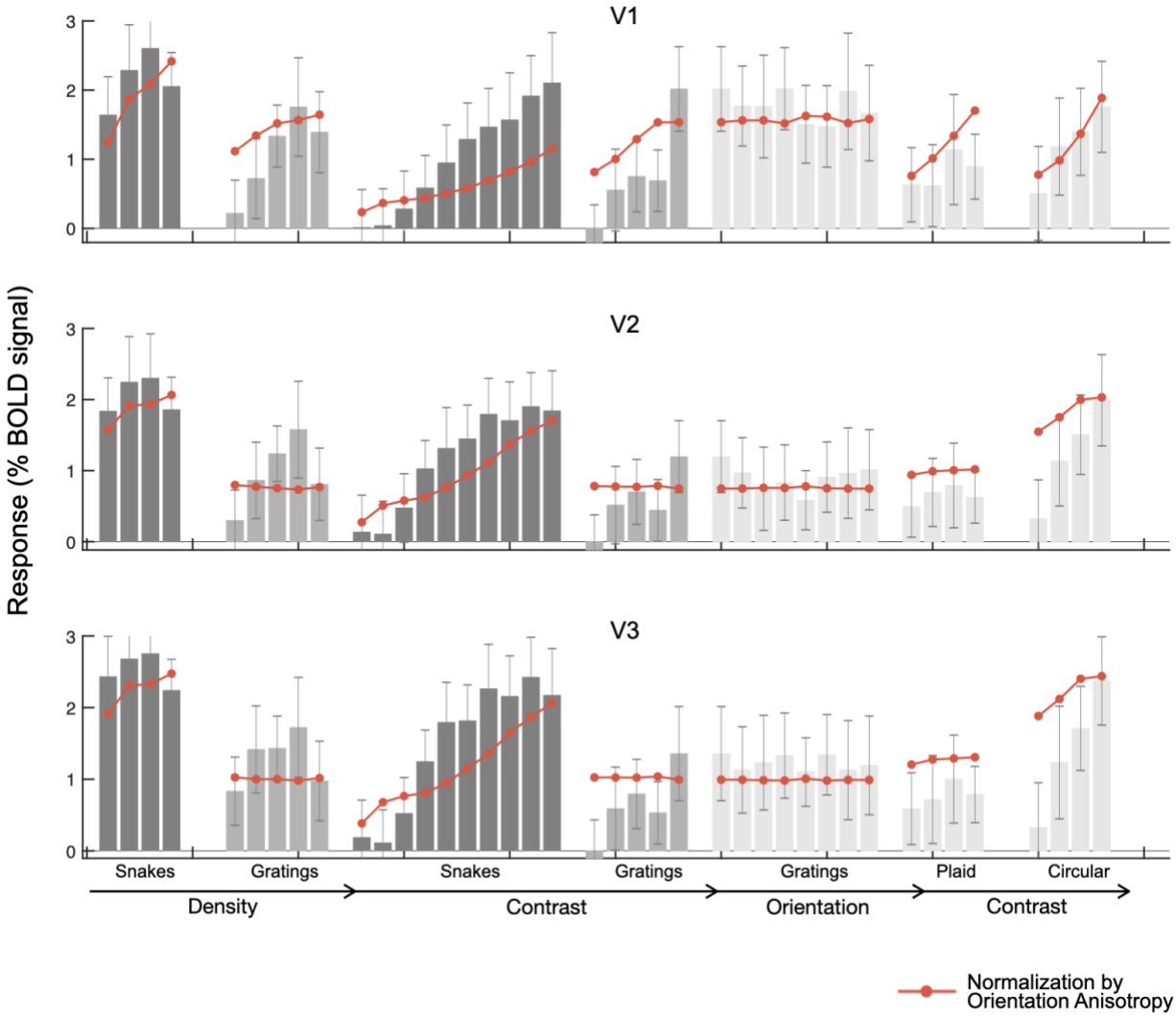


Figure S4a: Responses and CE model fits for all stimuli, data set 1

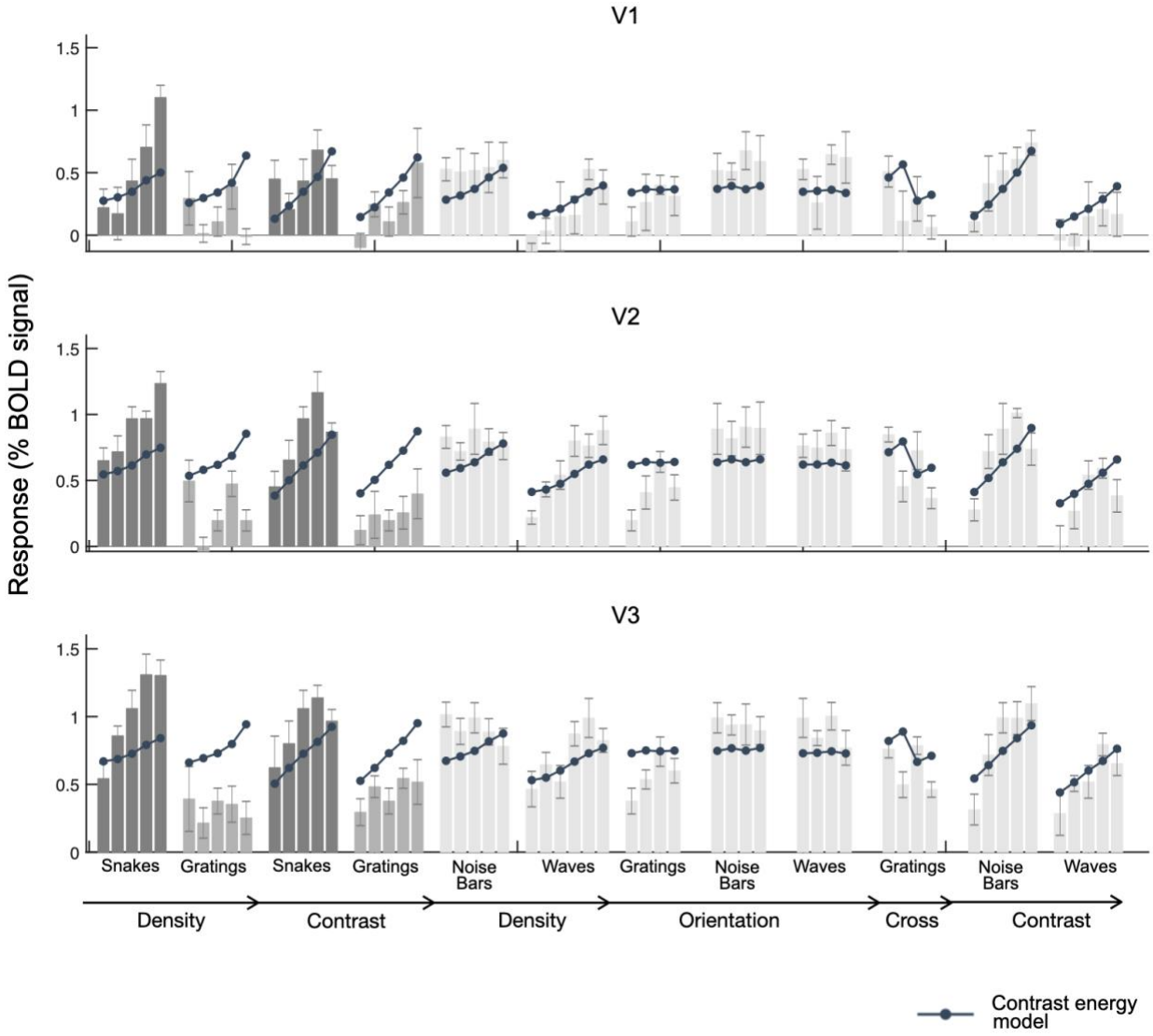


Figure S4b: Responses and CE model fits for all stimuli, data set 2

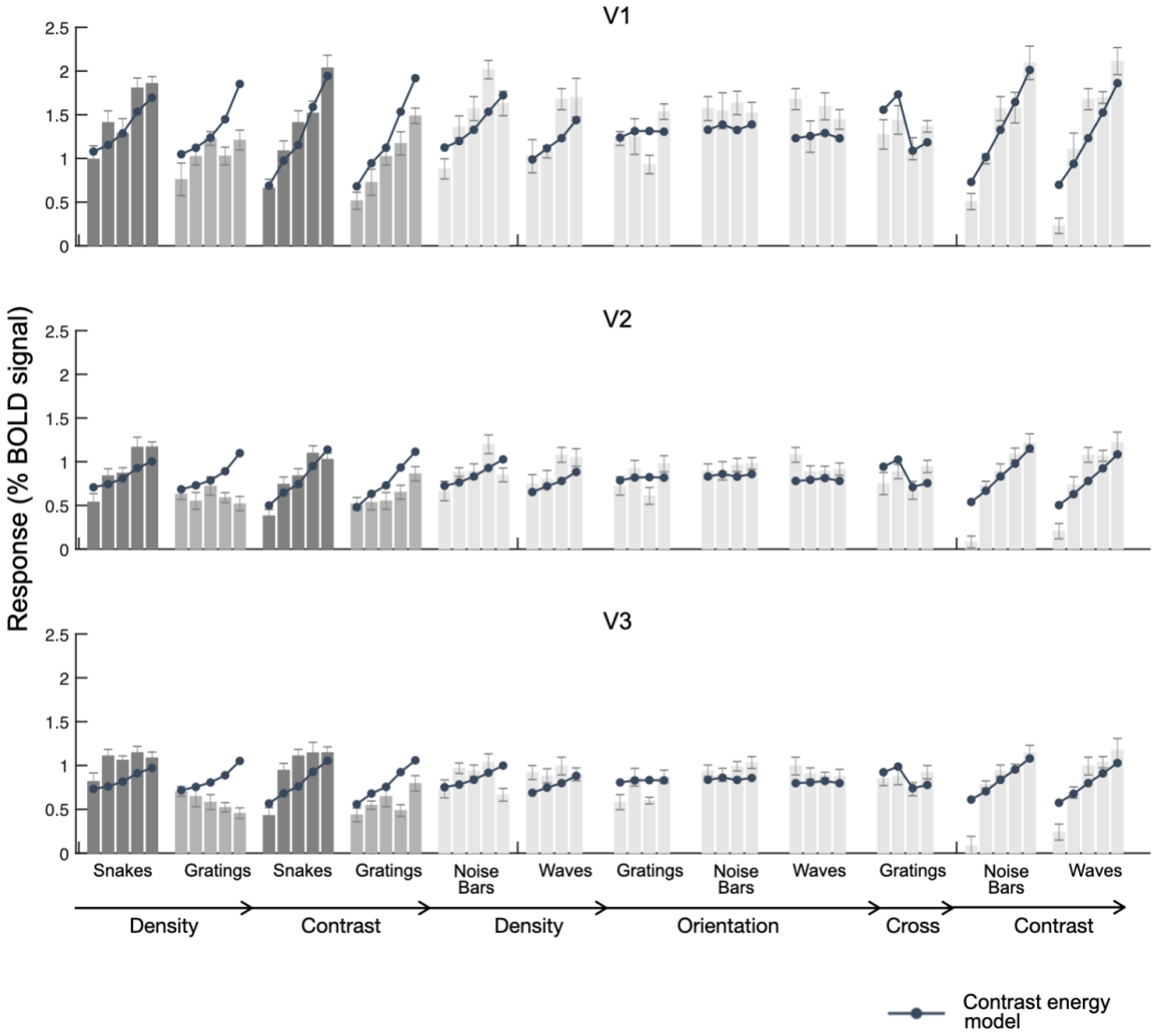


Figure S4c: Responses and CE model fits for all stimuli, data set 3

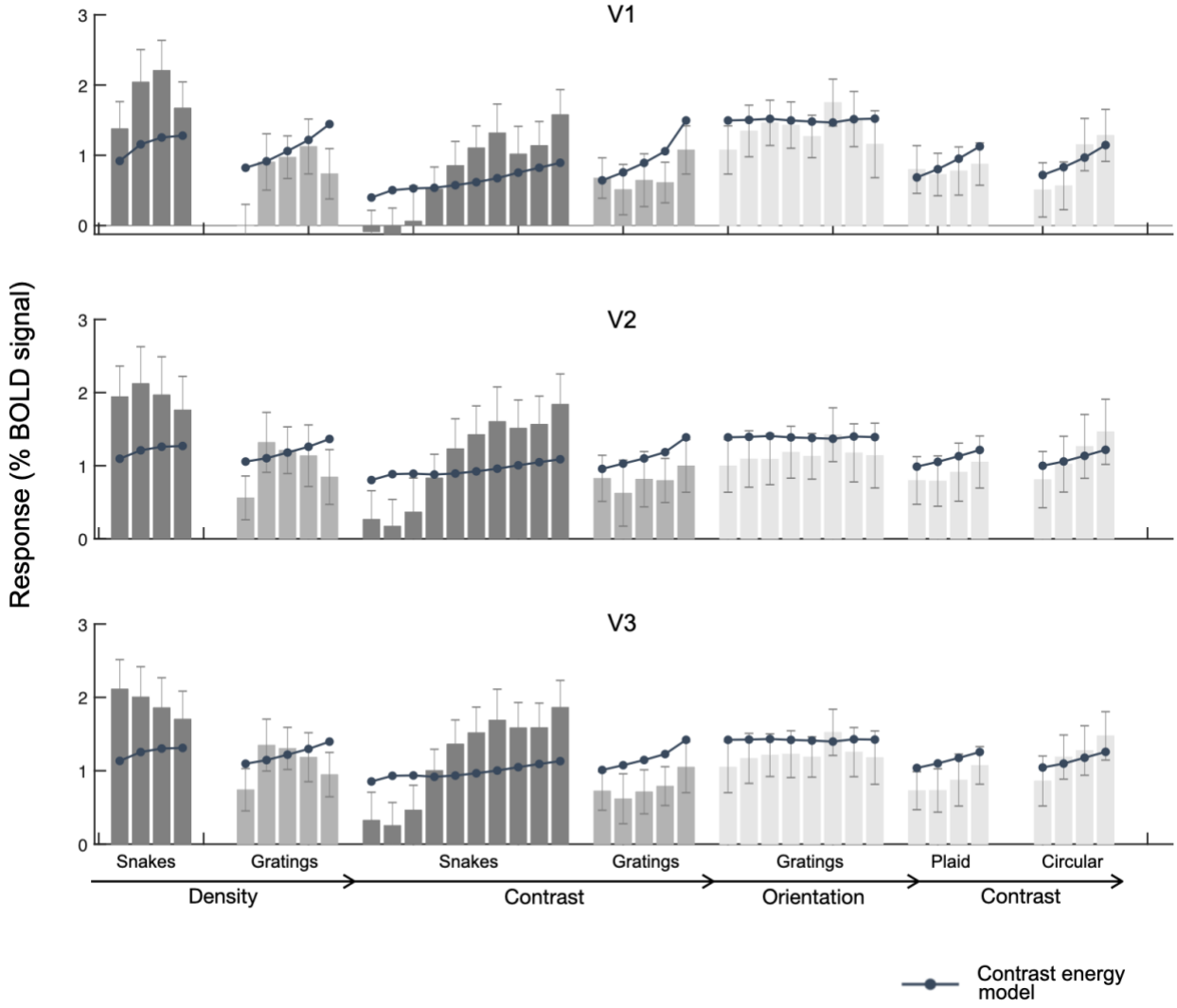


Figure S4d: Responses and CE model fits for all stimuli, data set 4

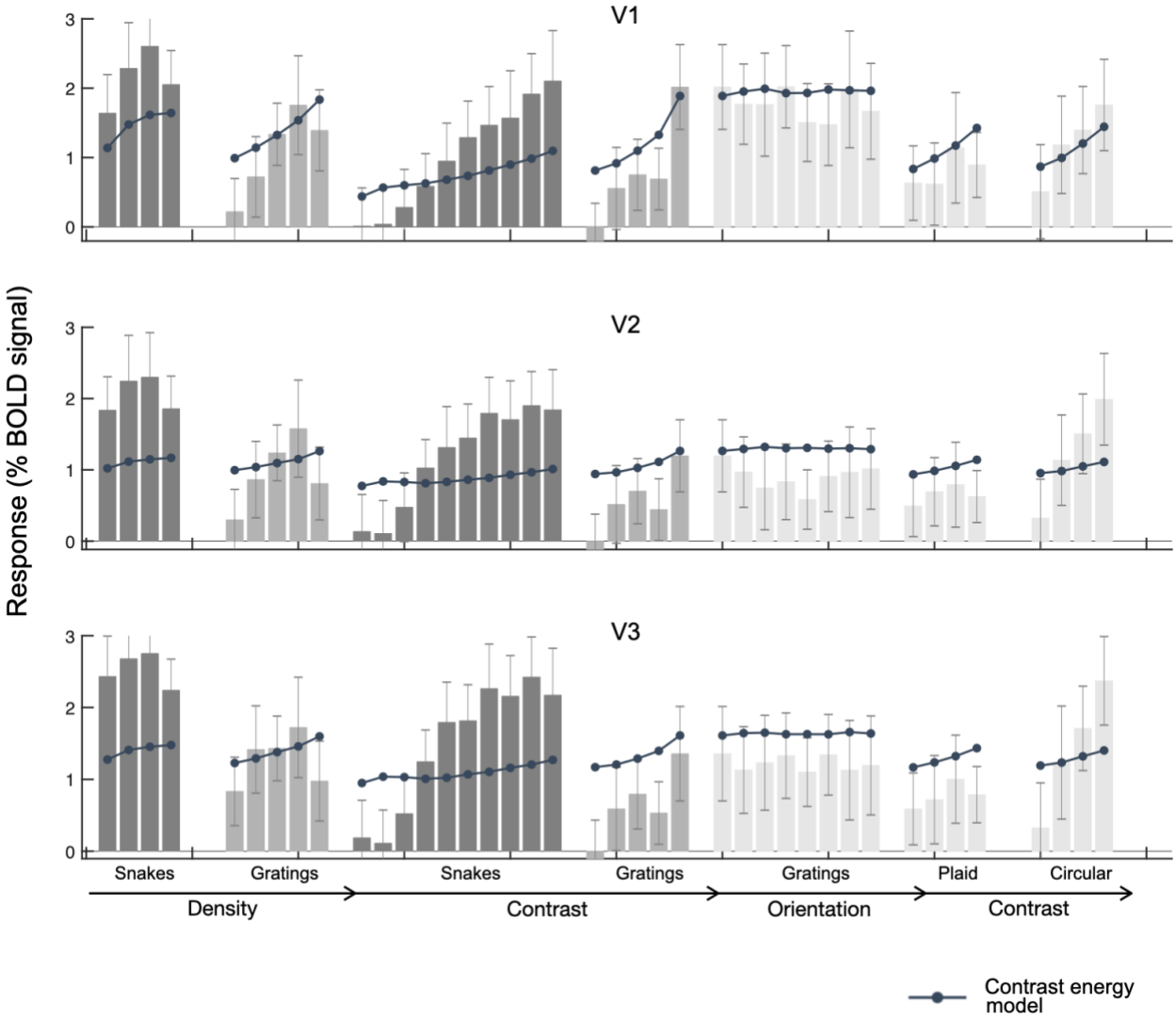


Figure S5a: Responses and SOC model fits for all stimuli, data set 1

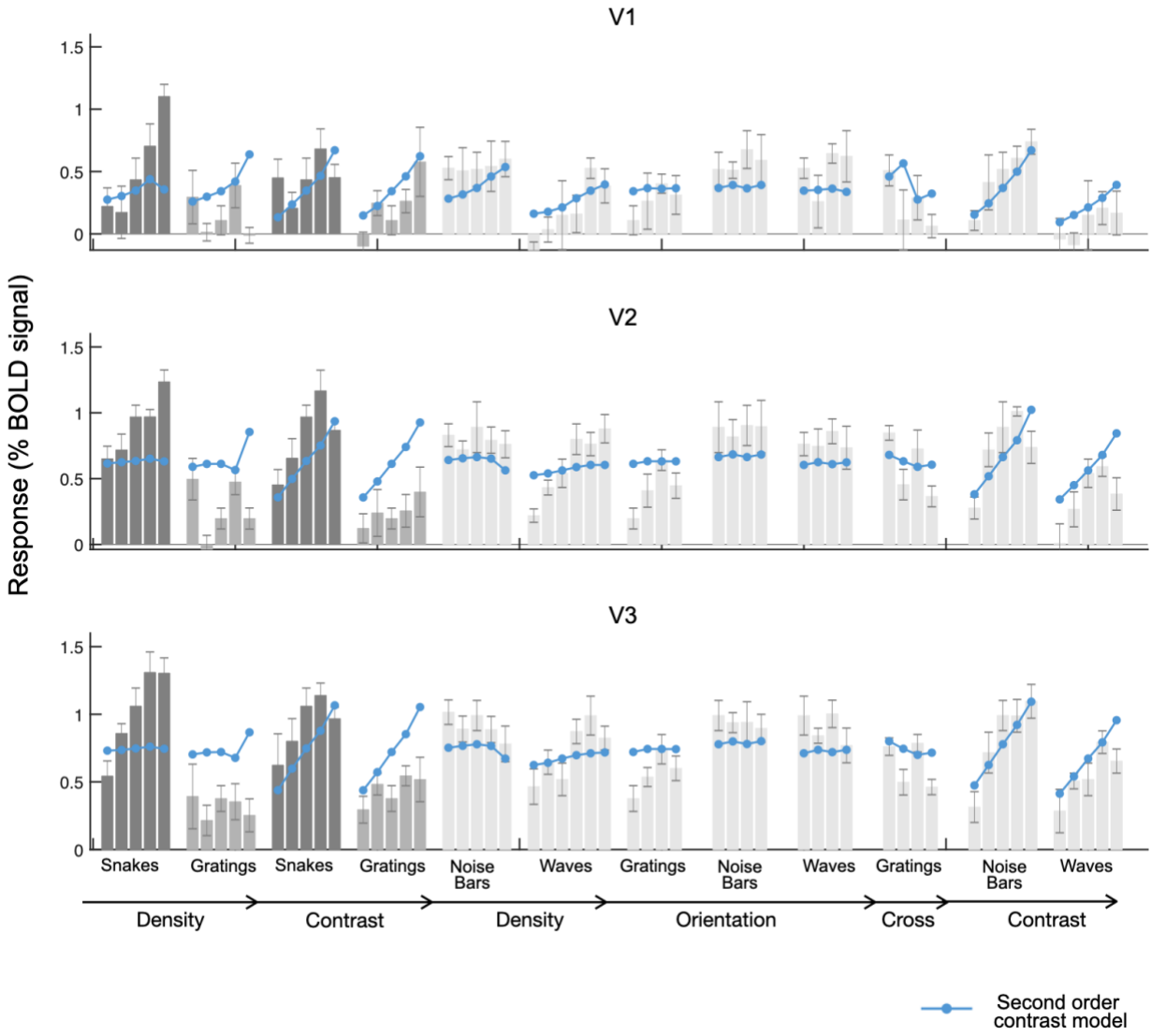


Figure S5b: Responses and SOC model fits for all stimuli, data set 2

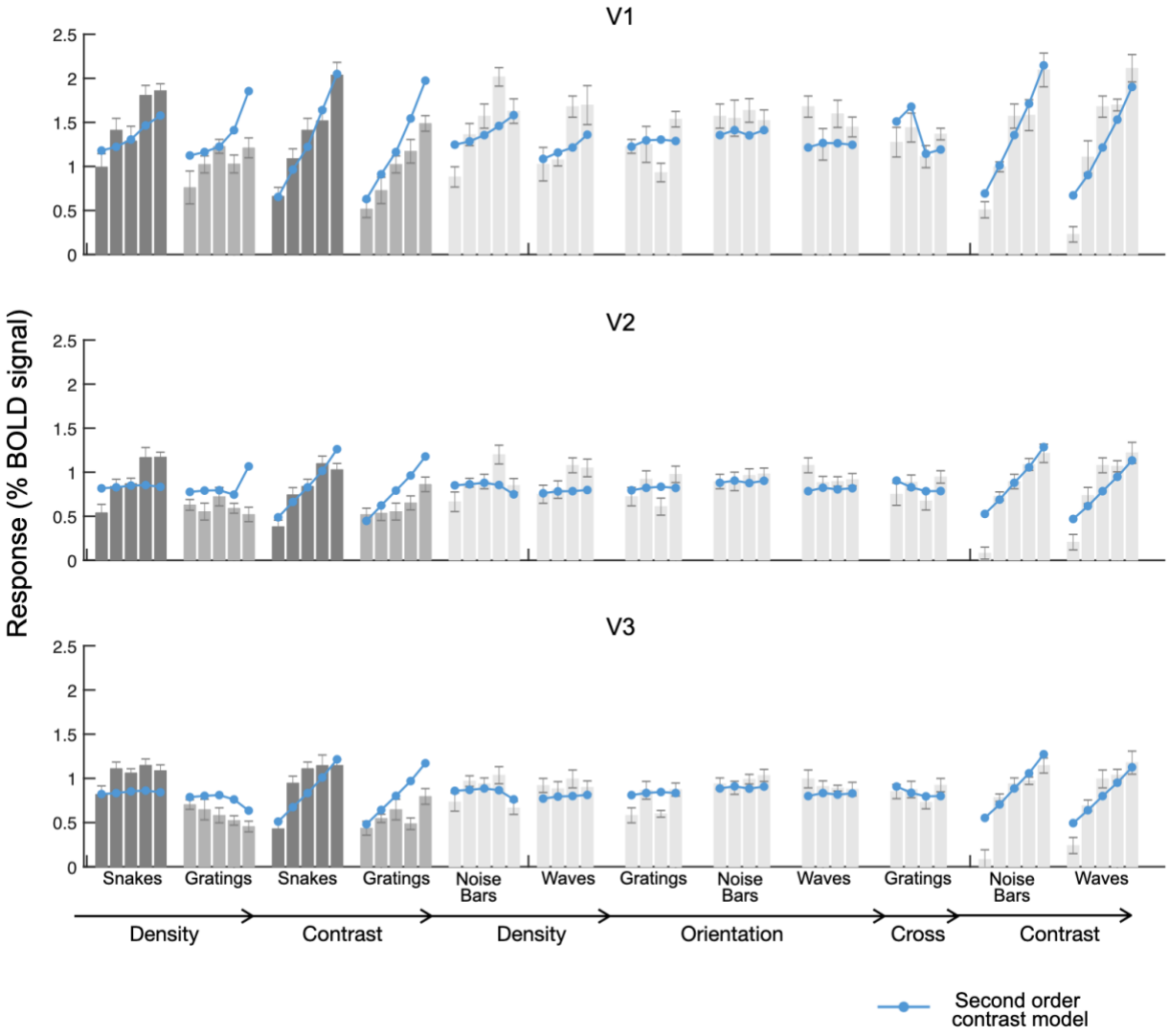


Figure S5c: Responses and SOC model fits for all stimuli, data set 3

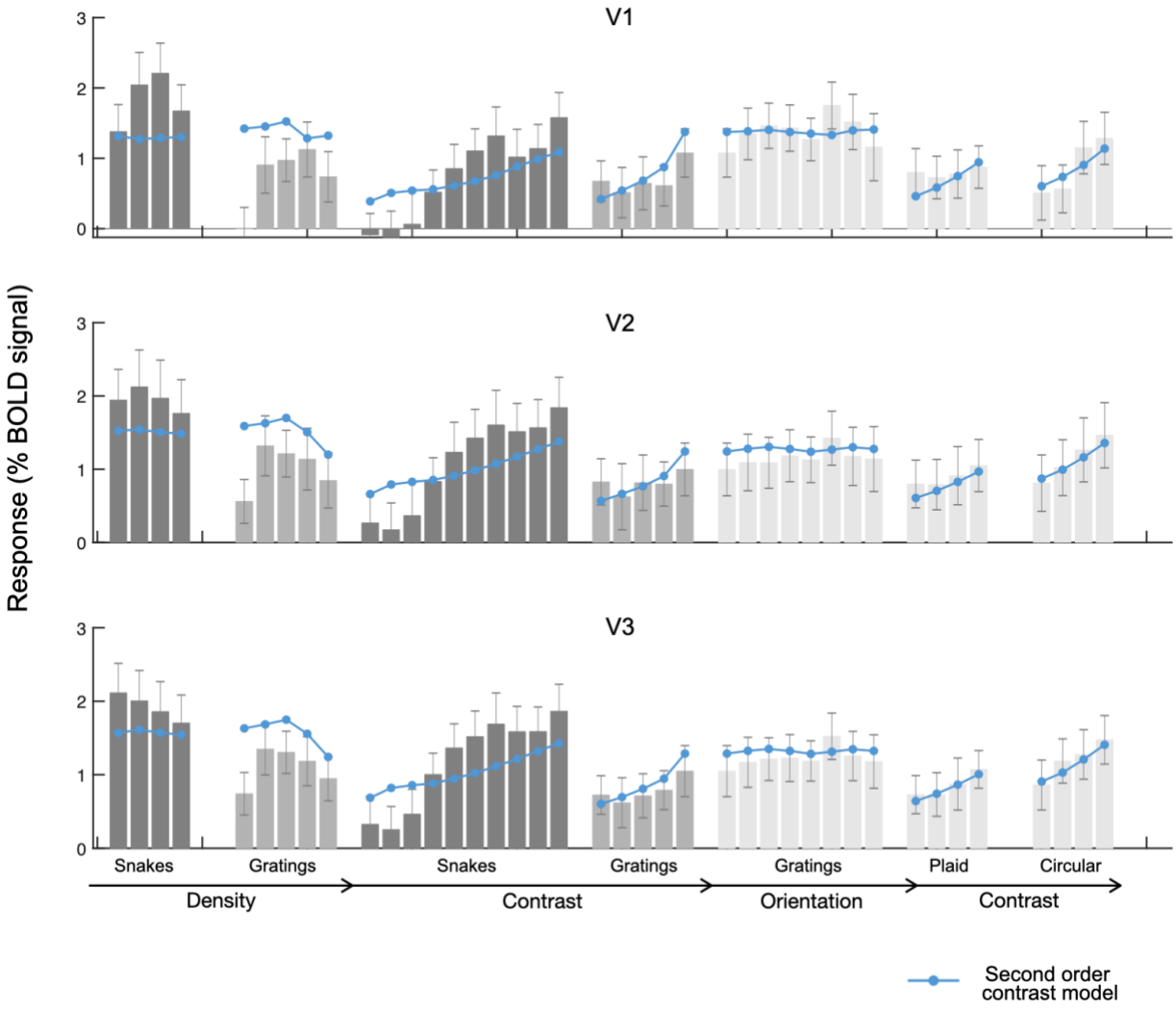


Figure S5d: Responses and SOC model fits for all stimuli, data set 4

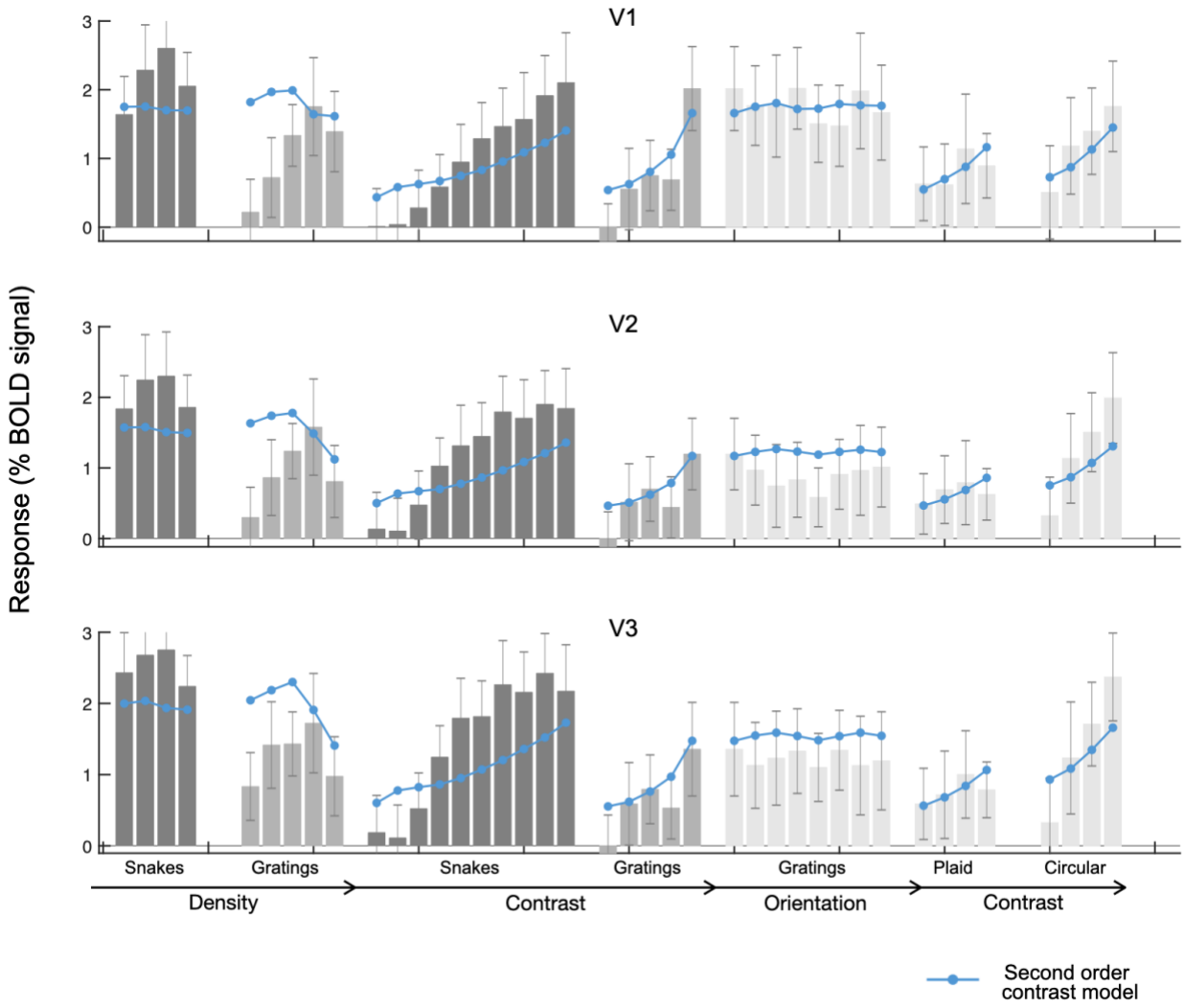


Figure S6a: Responses and OTS model fits for all stimuli, data set 1

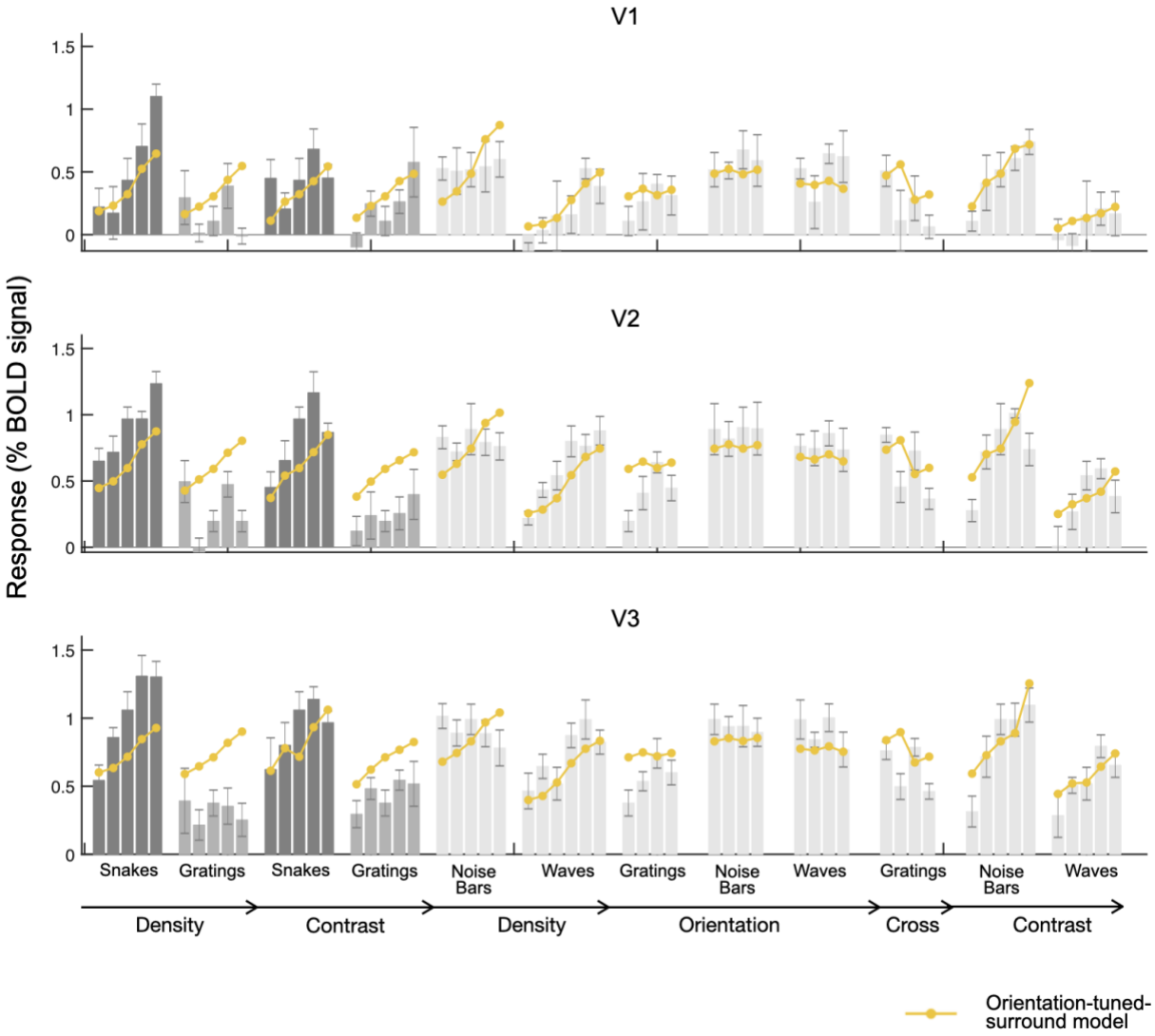


Figure S6b: Responses and OTS model fits for all stimuli, data set 2

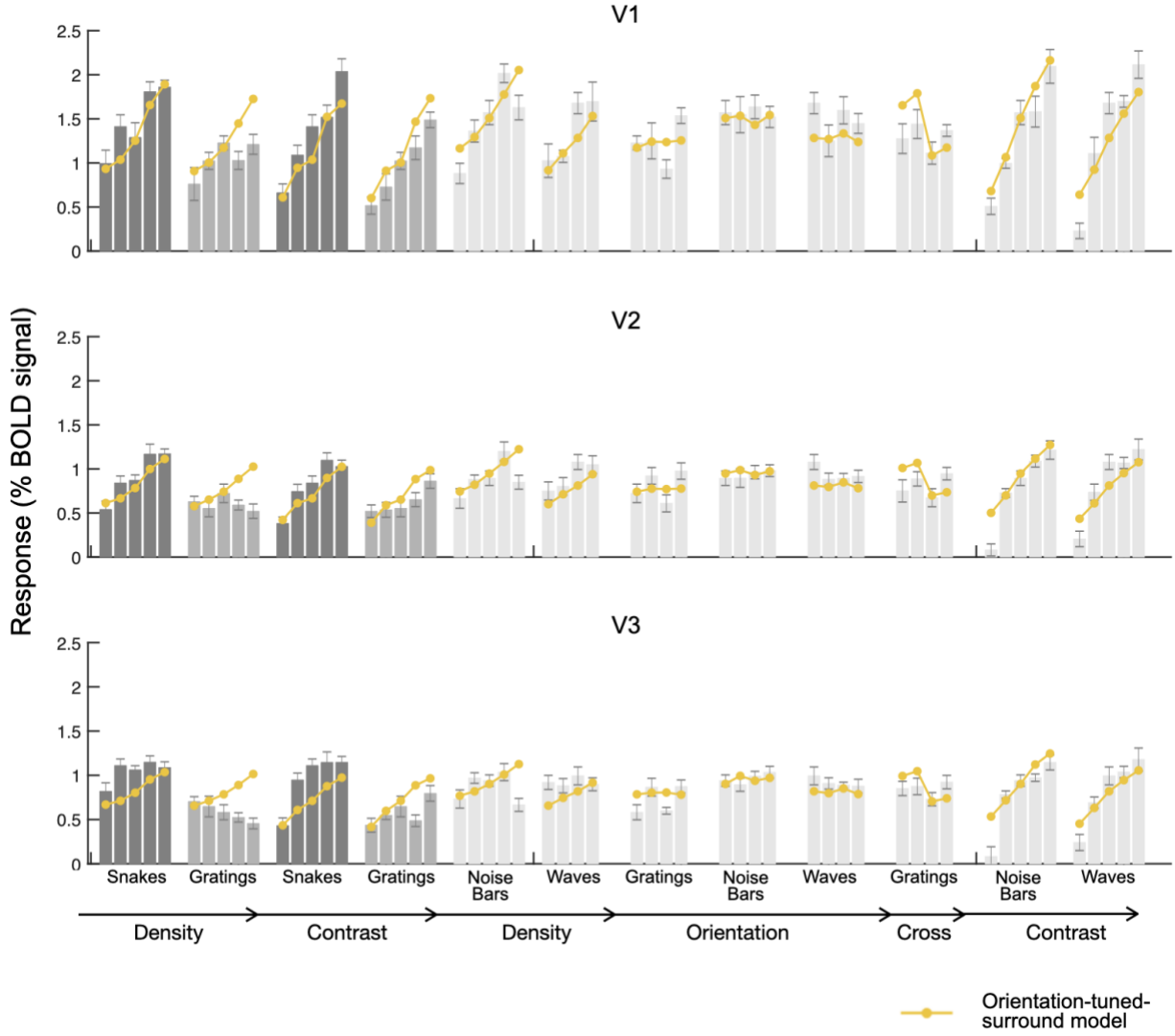


Figure S6c: Responses and OTS model fits for all stimuli, data set 3

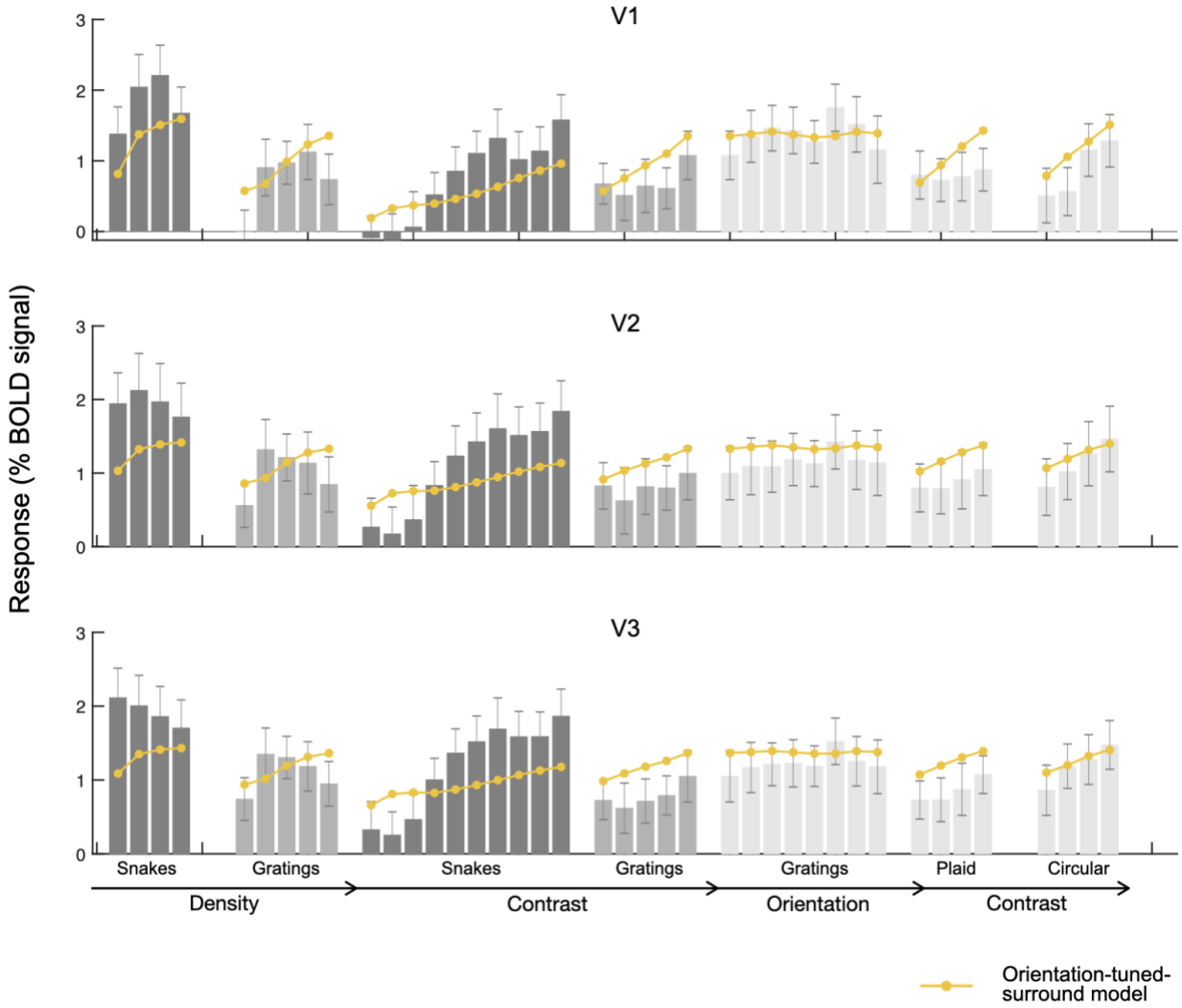
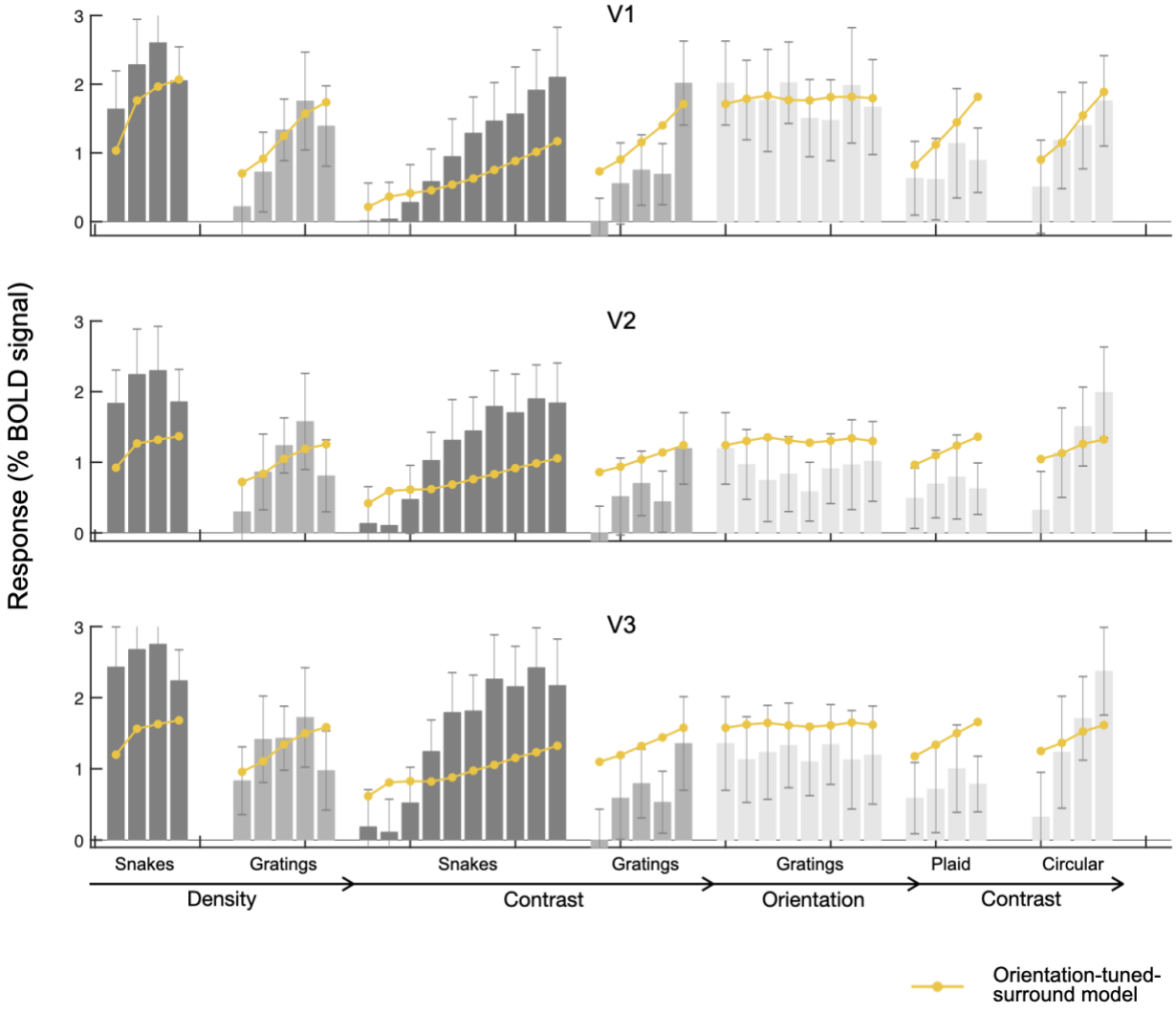



Figure S6d: Responses and OTS model fits for all stimuli, data set 4