Functional recursion of orientation cues in figure-ground separation

Jonathan D. Victor and Mary M. Conte

Feil Family Brain and Mind Research Institute
Weill Cornell Medical College
1300 York Avenue
New York, NY 10065

Voice: 212 746 2343
Email: jdvicto@med.cornell.edu

Running head: Characteristics of Orientation Cues in Figure-Ground

Keywords: visual textures; local image statistics; texture discrimination; scene organization

Abstract

Visual texture is an important cue to figure-ground organization. While processing of texture differences is a prerequisite for the use of this cue to extract figure-ground organization, these stages are distinct processes. One potential indicator of this distinction is the possibility that texture statistics play a different role in the figure vs. in the ground. To determine whether this is the case, we probed figure-ground processing with a family of local image statistics that specified textures that varied in the strength and spatial scale of structure, and the extent to which features are oriented. For image statistics that generated approximately isotropic textures, the threshold for identification of figure-ground structure was determined by the difference in correlation strength in figure vs. ground, independent of whether the correlations were present in figure, ground, or both. However, for image statistics with strong orientation content, thresholds were up to two times higher for correlations in the ground, vs. the figure. This held equally for texture-defined objects with convex or concave boundaries, indicating that these threshold differences are driven by border ownership, not boundary shape. Similar threshold differences were found for presentation times ranging from 125 to 500 ms. These findings identify a qualitative difference in how texture is used for figure-ground analysis, vs. texture discrimination. Additionally, it reveals a functional recursion: texture differences are needed to identify tentative boundaries and consequent scene organization into figure and ground, but then scene organization modifies sensitivity to texture differences according to the figure-ground assignment.

Introduction

Texture serves multiple purposes in early visual processing (Victor, Conte, & Chubb, 2017). Texture is an indicator of surface properties (Motoyoshi, Nishida, Sharan, & Adelson, 2007; Sharan, Liu, Rosenholtz, & Adelson, 2013; Sharan, Rosenholtz, & Adelson, 2014), and therefore serves as a cue to object and material identity(Fleming, 2014); it may drive visual search (Bergen & Julesz, 1983; Chubb, Scofield, Chiao, & Sperling, 2012; Julesz & Bergen, 1983); and it may act as a cue to separate figure from ground(Schmid & Victor, 2014). Here we ask whether two aspects of the use of visual texture –figure-ground separation and texture discrimination – can be understood in terms of a similar perceptual measure of sensitivity to texture statistics.

While it is tempting to expect that a single measure of sensitivity to texture difference underlies these (and perhaps other) roles of texture, there is reason to suspect that this is not the case. In figure-ground separation, there is the possibility that texture plays distinguishable roles in figure vs. ground, as objects and backgrounds are different: not only in their angular size, but also, in the materials of which they are made. Since some kinds of texture might be more likely to be present in an object than in a background, differences in thresholds may reflect the statistics of the natural environment. Moreover, the decision to perceive one side of a boundary as figure and the other side as ground may involve the shape of the potential figure, especially with regard to convexity vs. concavity (Cheng, Walther, Park, & Dilks, 2021; Sprote & Fleming, 2013; Wilder, Feldman, & Singh, 2011).

Size, shape (Baylis & Cale, 2001), and scene organization are examples of higher-level cues that may interact with the feedforward computations that are generally considered to account for texture processing (Chubb & Landy, 1991). Such interactions would be of particular interest if they are present in circumstances in which texture provides the only low-level visual cue to figure-ground, as this would imply a functional recursion: texture differences are needed for a tentative assignment of figure and

ground, but this assignment is given greater weight if it is reinforced by the implied scene organization. The net result of this interaction would be a lower threshold for identifying figure and ground when texture cues and higher-level cues reinforce.

On the other hand, if, despite the possibility that texture could be used differently in figure vs. ground, no such difference is found, then modeling its role would be substantially simplified. One could then frame the visual computations for use of texture in figure-ground solely in terms of a perceptual distance between textures. In this framework, figure-ground thresholds, like texture discrimination thresholds, would correspond to a threshold value of this distance, where the distance is computed from coordinates in the texture space. The power of this simplification is evident from previous work in texture discrimination. In multidimensional texture domains -- textures with a range of gray levels but no spatial correlations (Chubb, Landy, & Econopouly, 2004; Silva & Chubb, 2014), and textures with spatial correlations and two (Victor, Thengone, Rizvi, & Conte, 2015) or three luminance levels (Tesileanu et al., 2020) -- a Euclidean distance in appropriately-chosen texture coordinates accurately accounts for discrimination thresholds. Note that, in contrast to the possibility that texture is used differently in figure vs. ground, texture discrimination must use texture equivalently: if texture A is discriminable from texture B, then texture B is discriminable from texture A.

To determine whether the simple distance-based framework that succeeds in accounting for texture discrimination can be extended to figure-ground separation, here we probed the use of texture differences in a figure-ground task, focusing on a set of local image statistics that provided strong texture-discrimination cues and for which discrimination thresholds have been well-characterized. We found that the overall sensitivity to different image statistics approximated the sensitivities observed for texture discrimination. For texture statistics that produced approximately isotropic power spectra, we found that texture played a similar role in figure as in ground, as expected for a distance-based

framework. However, for texture statistics that produced strongly oriented power spectra, we found threshold differences of up to a factor of two, depending on whether the local correlations were present in figure, vs. ground. We further show that this asymmetry did not depend on the concavity vs. convexity of the boundary, and we characterize its time course. These findings imply that texture cues in figure-ground discrimination are used in a manner that combines local computations with top-down influences related to scene organization.

Materials and Methods

Our goal is to analyze the contribution of local image statistics to figure/ground separation, with a focus on whether cues behave similarly when they are present in the figure, vs. in the ground. To accomplish this, we used a two-alternative forced-choice paradigm. Target stimuli consisted of several randomly-positioned shapes (usually circular disks) rendered in one texture superimposed on a background rendered in a second texture. In each trial, target stimuli were paired with non-target stimuli, in which the entire field was a uniform texture, whose statistics were equal to the spatial average of the statistics of the target. Sensitivity was then measured by independently varying the strength and nature of the local image statistics of the figure and ground textures. In all experiments, the shapes that made up the figure totaled one quarter of the stimulus area.

The textures used to render target and non-target were drawn from a space of black-and-white textures previously described (Victor & Conte, 2012), a 10-dimensional space whose parameters specify the luminance distribution and correlations of checks within a 2×2 neighborhood. Within this space, we focused on a 4-dimensional subset parameterized by pairwise correlations in horizontal, vertical, and the two diagonal dimensions, described in detail below.

These choices were motivated by several considerations. First, the texture domain encompasses two kinds of features that are strong cues for figure-ground separation -- orientation and spatial scale – and enabled probing them and their interactions in a directly comparable fashion. Second, since the textures are generated in a pixel-by-pixel fashion, the transition between figure and ground could be accomplished without the introduction of elements that were present in neither figure nor ground. Such elements – for example, the abutting edges of gratings that differ in spatial phase -- might serve as spurious cues, in addition to cues arising from the difference between figure and ground. Finally, previous studies have delineated the roles of these cues in texture discrimination and border salience (Victor, Rizvi, & Conte, 2017; Victor, Thengone, & Conte, 2013; Victor et al., 2015), serving both as a

starting point for the present work and a linkage to the informative features of natural images (Hermundstad et al., 2014; Tkacik, Prentice, Victor, & Balasubramanian, 2010).

The texture domain

The textures used to render the target and non-target were drawn from a four-parameter domain of binary textures. The four parameters of the space, denoted β_{-}, $\beta_{|}$, $\beta_{\backslash}$, and $\beta_{/}$ are second-order statistics: they specify the correlations between pairs of checks. β_{-} and $\beta_{|}$ refer to correlations between adjacent checks in the two cardinal directions, i.e., between checks that share a common edge. $\beta_{\backslash}$, and $\beta_{/}$ refer to correlations in the two diagonal directions, i.e., between checks that share a common vertex. In all cases, a value of $+1$ means that the checks are either both black or both white; a value of -1 means that they have opposite luminance values, and a value of 0 means that they are uncorrelated.

Here, we considered textures in which one or two of these parameters are varied. The strips in Figure 1A show the effect of sweeping each of the parameters through their gamut of $[-1,+1]$. For the cardinal directions (β_{-} and $\beta_{|}$), positive correlations generate textures with prominent stripes, either horizontally or vertically, while negative correlations increase the probability of contrast edges along the corresponding axis. For the diagonal directions ($\beta_{\backslash}$ and $\beta_{/}$), positive and negative correlations have analogous, but visually distinct, appearances. Because the cardinal correlations refer to checks that share an edge while diagonal correlations refer to checks that share a vertex, the diagonal correlations are not equivalent to simply rotating the cardinal correlations by 45 deg.

The square domains in Figure 1B show the effect of allowing two texture parameters to have a nonzero value. The resulting textures have correlations in more than one direction, and combining texture parameters can either increase or decrease the anisotropy of a texture. For example, when β_{-} and $\beta_{|}$ have the same sign, textures are approximately isotropic; when they have opposite sign, the

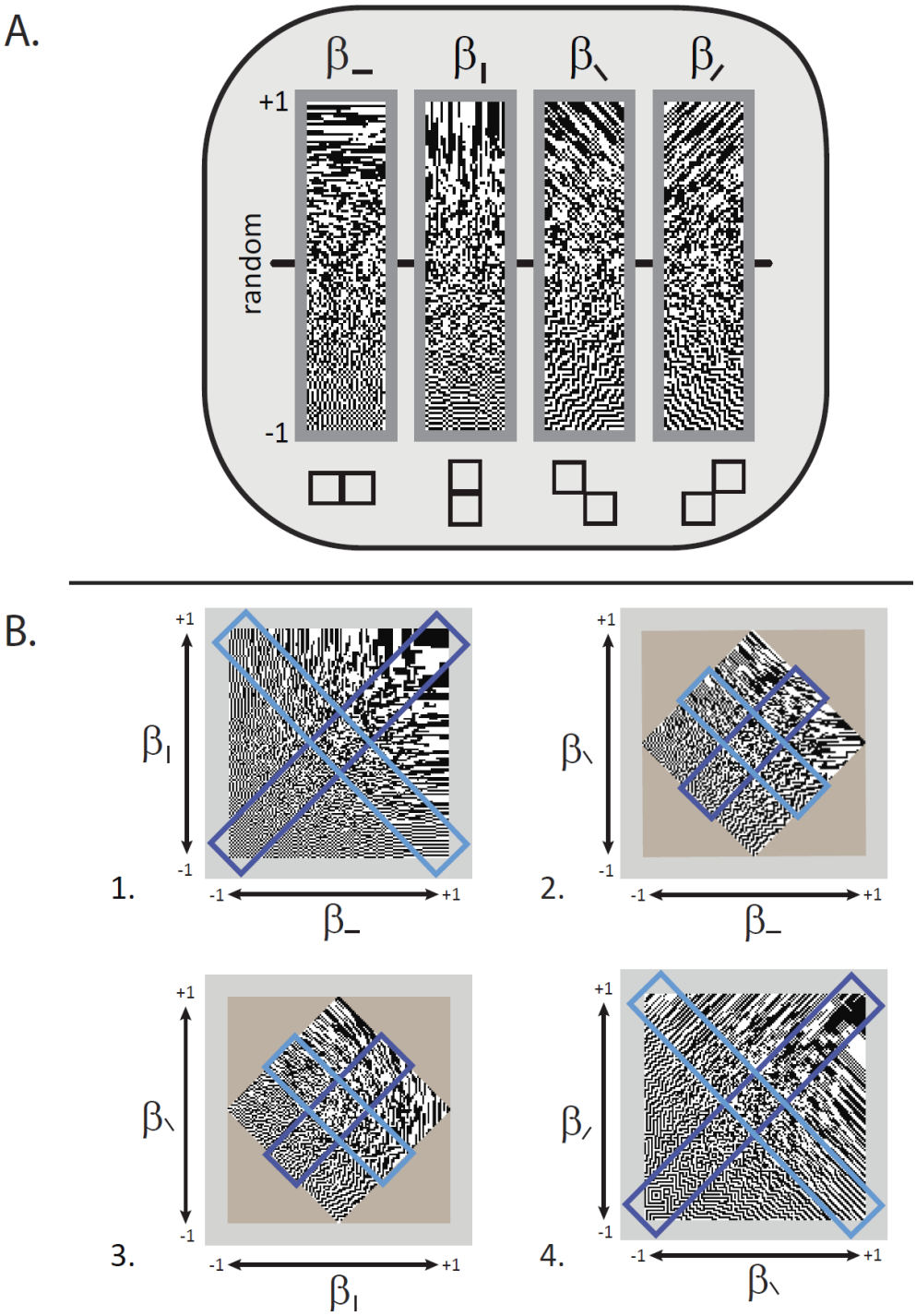

Figure 1. The textures used in these studies. A: Each of the four strips shows the effect of varying a single texture parameter through its range, from -1 to $+1$. B: The effect of jointly varying two texture parameters. The darker oblique rectangle shows the effect of varying the two texture parameters equally; the lighter oblique rectangle shows the effect of varying the two parameters equally but with opposite sign. Subpanels B1: $(\beta\_, \beta|)$; B2: $(\beta\_, \beta\backslash)$, B3: $(\beta|, \beta\backslash)$; B4: $(\beta\backslash, \beta/)$. In B2 and B3, the beige regions indicate parameter values that cannot be accessed by the stimulus generation algorithms of (Victor & Conte, 2012). In all cases, $+1$ indicates maximal correlation, -1 indicates maximal anticorrelation, and 0 indicates no correlation (random).

textures are strongly oriented. In all cases, the textures generated have an equal probability of black and white checks, and are maximum-entropy: that is, the textures are as random as possible, given the specified pairwise correlations.

Within the target, figure and ground were delineated by assigning different values of the texture parameters to these image components. To synthesize images in which the values of the texture parameters vary across space, we used the approach of (Victor & Conte, 2012), which provides a mapping from the texture parameters to a specification of the probability of the various configurations of 2×2 blocks of checks. For locations within the figure (to be precise, when the center of a check lies within the figure), the figure's parameters are used to specify these probabilities; otherwise, the ground's parameters are used. At the boundary between figure and ground, the probabilities assigned to the 2×2 blocks change abruptly. However, since every kind of 2×2 block can occur both in the figure and in the ground, there are no edge artifacts that occur at the boundaries that can act as boundary cues.

Power spectra and orientation content of the textures

To help interpret the psychophysical findings, we calculated the spatial power spectra and orientation content of the textures parametrized by one or two of the second-order statistics. The power spectrum of each texture was computed by Fourier transformation of its autocorrelation function $X(I,J)$, defined by

$$X(I,J)=\left\langle T(x,y)T(x+I,y+J)\right\rangle_{x,y}, \tag{1}$$

where $T(x,y)=-1$ for a black check, and $T(x,y)=+1$ for a white check, I increases from left to right, and J increases from top to bottom. For textures defined by a single statistic, the autocorrelation functions have simple analytic forms. Specifically, since the textures are generated by one-dimensional Markov processes (Victor & Conte, 2012), the autocorrelations decline exponentially in the direction in

which neighboring checks are correlated, and are zero in directions that are not parallel to this direction. Thus,

$$X_{\beta\_}(I,J)=\begin{cases}(\beta\_)^{|I|}, J=0\\ 0, J\neq 0\end{cases}, \tag{2}$$

$$X_{\beta|}(I,J)=\begin{cases}(\beta|)^{|J|}, I=0\\ 0, I\neq 0\end{cases}, \tag{3}$$

$$X_{\beta\backslash}(I,J)=\begin{cases}(\beta\backslash)^{|I|}, I=J\\ 0, I\neq J\end{cases}. \tag{4}$$

and

$$X_{\beta/}(I,J)=\begin{cases}(\beta/)^{|I|}, I=-J\\ 0, I\neq -J\end{cases}. \tag{5}$$

For pairs of texture statistics, the constructions of (Victor & Conte, 2012) lead to simple analytic forms when both local correlations are in cardinal directions or both are in diagonal directions:

$$X_{\beta\_,\beta|}(I,J)=(\beta\_)^{|I|}(\beta|)^{|J|}, \tag{6}$$

and

$$X_{\beta\backslash,\beta/}(I,J)=\begin{cases}(\beta\backslash)^{\frac{|I+J|}{2}}(\beta/)^{\frac{|I-J|}{2}}, |I-J|\ even\\ 0, |I-J|\ odd\end{cases}. \tag{7}$$

For pairings of one cardinal and one diagonal statistic (e.g, $X_{\beta\_,\beta\backslash}(I,J)$), a simple form for the entire autocorrelation is elusive, though the following special cases follow from the constructions of (Victor & Conte, 2012):

$$
\begin{aligned}
X_{\beta\_,\beta\backslash}(I,0) &= (\beta\_)^{|I|} \\
X_{\beta\_,\beta\backslash}(I,I) &= (\beta\backslash)^{|I|} \\
X_{\beta\_,\beta\backslash}(I,1) &= I(\beta\_)^{I-1}\beta\backslash,\ I \geq 1 \\
X_{\beta\_,\beta\backslash}(1,J) &= (\beta\backslash)^{J}\left(-\frac{\beta\_}{1-\beta\_^{2}}\right)^{J-1},\ J \geq 1
\end{aligned}
\qquad (8)
$$

To estimate the full autocorrelation functions for these cases, we estimated the autocorrelations from 16 samples of 256×256 instances of the texture (and also used this approach to verify eqs. (2) through(8)).

To compute the spatial power spectrum $S(f_x, f_y)$ from the autocorrelation, we subsampled the autocorrelation functions in steps of 0.25, windowed by a squared cosine bell, and zero-padded with an apron equal to the size of the texture in each dimension to eliminate edge effects.

Oriented power at each angle θ, $P(\theta)$, was then determined from $S(f_x, f_y)$ by integration over spatial frequencies whose periods ranged from 2 to 8 checks (approximately 0.75 c/d to 3 c/d in these experiments), and weighting power at spatial frequencies with nearby orientations φ by a von Mises function $w(\varphi)$:

$$P(\theta) = \frac{1}{2\pi}\int_0^{2\pi}\int_{f_{lo}}^{f_{hi}} S(f\cos\varphi, f\sin\varphi)\, w(\theta-\varphi)\, df\, d\varphi, \text{ where} \qquad (9)$$

$$w(\varphi) = K\exp\left(B\left(\cos(\varphi)-1\right)\right) \qquad (10)$$

and K is chosen so that w integrates to 1. We used a von Mises function that had a full width at half maximum of $\pi/4$ (45 deg), which corresponds to $B = \ln 2 / \left(1-\cos(\pi/8)\right) \approx 9.1059$.

Organization of experiments and analysis

Our goal was to measure figure-ground thresholds for positive and negative correlations specified by the texture parameters, and to compare these thresholds when the correlations were present in the figure, or the ground, or both. To do this, we organized the experiments along "rays", i.e., a set of stimuli in which the texture parameters had a fixed ratio but varied in overall strength. We further

grouped the rays that explored a single image statistic or image statistic pair into a figure-ground plane, as illustrated in Figure 2.

For each ray, the target stimuli are determined by a choice of one or two of the β's, along with a multiplier c_{fig} for the figure, and a multiplier c_{gnd} for the ground. For example, for a ray that probed the statistics $\beta_{\_}$ and $\beta_{|}$, the figure texture was specified by $(\beta_{\_,fig}, \beta_{|,fig}) = c_{fig}(\beta_{\_}{}^{max}, \beta_{|}^{max})$, and the ground texture by $(\beta_{\_gnd}, \beta_{|gnd}) = c_{gnd}(\beta_{\_}{}^{max}, \beta_{|}^{max})$. The parameters that specified the texture of the non-target stimulus was taken to be the spatial average of the parameters for the target stimulus: $(\beta_{\_,nontarg}, \beta_{|,nontarg}) = c_{dis}(\beta_{\_}{}^{max}, \beta_{|}^{max})$, where

$$c_{nontarg} = \frac{1}{4} c_{fig} + \frac{3}{4} c_{gnd}, \tag{11}$$

since the figure occupied one-quarter of the stimulus area.

The following figure-ground planes were examined: $\beta_{\_}{}^{max} = +1$, $\beta_{|}^{max} = +1$, $\beta_{\backslash}^{max} = +1$, $\beta_{/}^{max} = +1$ (the latter in one subject only), $(\beta_{\_}{}^{max}, \beta_{|}^{max}) = (+1, +1)$, $(\beta_{\_}{}^{max}, \beta_{|}^{max}) = (+1, -1)$, $(\beta_{\_}{}^{max}, \beta_{\backslash}^{max}) = (+1, +1)$, $(\beta_{\_}{}^{max}, \beta_{\backslash}^{max}) = (+1, -1)$, $(\beta_{|}^{max}, \beta_{\backslash}^{max}) = (+1, +1)$, $(\beta_{|}^{max}, \beta_{\backslash}^{max}) = (+1, -1)$, $(\beta_{\backslash}^{max}, \beta_{/}^{max}) = (+1, +1)$, $(\beta_{\backslash}^{max}, \beta_{/}^{max}) = (+1, -1)$. This provides for a complete exploration of the four second-order statistics in isolation and their pairwise interactions, along with a verification in one subject (MC) that $\beta_{\backslash}$ and $\beta_{/}$, which are related by left-right mirror symmetry, behave similarly.

In all cases except for the $(\beta_{\_}{}^{max}, \beta_{|}^{max})$ figure-ground planes, the unspecified β's were zero; for this case, $\beta_{\backslash} = \beta_{/} = \beta_{\_}\beta_{|}$, as required by the maximum-entropy construction (Victor & Conte, 2012). However, as shown below, the thresholds for the cardinal β's were sufficiently low so that at

threshold, the diagonal pairwise correlations introduced by $\beta_{\backslash} = \beta_{/} = \beta\_\beta_{|}$ were well below their thresholds.

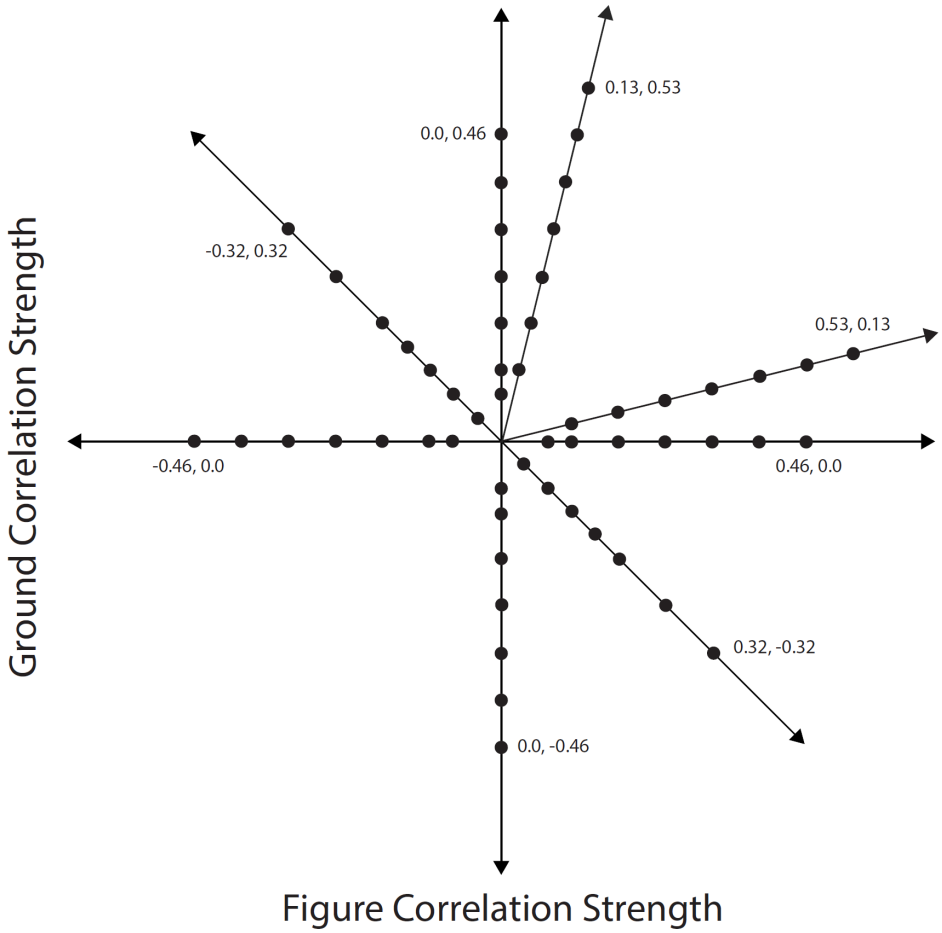

Figure 2. Experimental design for measuring thresholds in a figure-ground plane. The abscissa and ordinate indicate, respectively, the overall correlation strength in the figure and in the ground. All experiments selected from these rays; texture statistics in figure and ground had a constant ratio along each ray. The seven points along each ray correspond to the stimuli used to determine the threshold along that ray; their locations were chosen to cover the range from near-chance to near-perfect performance as determined by previous pilot experiments. Here we show the correlation strengths used to probe the figure-ground planes defined by $(\beta\_^{max}, \beta_{|}^{max}) = (+1, +1)$ in which horizontal and vertical correlations ($\beta\_$ and $\beta_{|}$) were combined with the same sign, and by $(\beta\_^{max}, \beta_{|}^{max}) = (+1, -1)$ in which these correlations $\beta\_$ and $\beta_{|}$ were combined with the opposite sign. For each stimulus, the texture statistics used to render figure and ground textures are determined by $(\beta\_{fig}, \beta_{|fig}) = c_{fig}(\beta\_^{max}, \beta_{|}^{max})$ and $(\beta\_{gnd}, \beta_{|gnd}) = c_{gnd}(\beta\_^{max}, \beta_{|}^{max})$. Points along the horizontal axis correspond to stimuli for which only the figure contained correlations ($c_{fig} \neq 0$, $c_{gnd} = 0$); points along the vertical axis correspond to stimuli for which only the ground contained correlations ($c_{fig} = 0$, $c_{gnd} \neq 0$).

As shown in Figure 2, within each figure-ground plane, we explored multiple rays emanating from the origin (the random texture). Along four of the rays – the coordinate axes in the figure-ground plane -- either figure or ground was structured: $(c_{fig}, c_{gnd}) = (+1, 0)$, $(c_{fig}, c_{gnd}) = (-1, 0)$, $(c_{fig}, c_{gnd}) = (0, +1)$, $(c_{fig}, c_{gnd}) = (0, -1)$. Along four other rays – oblique directions in the figure-ground plane -- both figure and ground contained correlations: $(c_{fig}, c_{gnd}) = (+1, +0.25)$, $(c_{fig}, c_{gnd}) = (+0.25, +1)$, $(c_{fig}, c_{gnd}) = (+1, -1)$ and $(c_{fig}, c_{gnd}) = (-1, +1)$. For all rays, we carried out pilot experiments to determine a range of contrasts for which performance (discrimination of target from non-target) was likely to vary from near-chance to near-perfect, and then chose seven points c_m to measure a psychometric function. These values were multiplied by the maximal texture contrasts specified by the ray ((c_{fig}, c_{gnd})) and the plane (the β^{max}'s). Thus, the texture parameters in a target stimulus at position c_m along a ray were specified by $\beta_{s,fig} = c_m c_{fig} \beta_s^{max}$ and $\beta_{s,gnd} = c_m c_{gnd} \beta_s^{max}$, where β_s runs over the one or two second-order statistics that specify the figure-ground plane. Note that at least one of c_{fig} or c_{gnd} had an absolute value of 1, and at least one of β_s^{max} also had an absolute value of 1. The correlation strengths were thus controlled by the choice of values of c_m determined in preliminary experiments. For the example figure-ground plane $(\beta\_^{max}, \beta|^{max}) = (+1, +1)$ shown in Figure 2, the values of c_m for rays with the lowest thresholds (e.g., the ray $(c_{fig}, c_{gnd}) = (+1, -1)$) ranged from 0.035 to 0.318; the values of c_m for rays with intermediate thresholds (e.g., $(c_{fig}, c_{gnd}) = (+1, 0)$) ranged from 0.071 to 0.460; and the values of c_m for rays with the highest thresholds (e.g., $(c_{fig}, c_{gnd}) = (+1, +0.25)$ ranged from 0.106 to 0.530). Note that while the specific values of c_m differed across the figure-ground planes, we always used the same values in figure-ground planes that differed in terms of whether the texture parameters were combined with the

same sign (e.g., $(\beta\_^{max}, \beta|^{max}) = (+1,+1)$) or with opposite sign (e.g., $(\beta\_^{max}, \beta|^{max}) = (+1,-1)$), as this was a critical comparison.

For each of the rays r in a figure-ground plane, thresholds were determined by fitting Weibull functions to the observed fraction correct (FC):

$$FC(c_m) = \frac{1}{2} + \frac{1}{2}\left(1 - 2^{-(c_m/a_r)^{b_r}}\right), \tag{12}$$

where c_m is the contrast multiplier, a_r is the value of the contrast multiplier for which FC=0.75 (halfway between chance (0.5), and perfect (1.0)), and b_r is the Weibull shape parameter. As in previous studies, (Victor, Rizvi, et al., 2017; Victor et al., 2013; Victor et al., 2015), we used a maximum-likelihood procedure(Victor, Chubb, & Conte, 2005) in two stages, summarized here. First, we fitted the Weibull threshold and shape parameters separately on each ray. Then, since the shape parameters typically had similar values along each ray (almost always in the range 2 to 3), we refined the estimates of the threshold parameters by refitting the data in each figure-ground plane with a common value of the shape parameter. By reducing the number of degrees of freedom from $2n_{rays}$ to $n_{rays}+1$ (where n_{rays} is the number of rays in a figure-ground plane), this strategy increased the precision of threshold estimation. The threshold contrasts along each ray were then taken to be $(a_r c_{fig}, a_r c_{gnd})$. As in the above previous studies, we determined 95% confidence intervals for the threshold value via 1000-sample bootstraps. As in (Victor, Rizvi, et al., 2017), we computed averages across subjects via the geometric mean.

Task and procedure

Stimuli were grouped into blocks of 140 trials (20 examples at each of 7 multiplier values), along a single ray. Blocks were grouped into sets containing one block for each ray, and four such sets were presented, accumulating 560 unique trials per ray. The order of the rays within sets was randomized

within and across subjects. The order in which planes were tested was balanced across subjects. Prior to the presentation of each block, subjects were shown paper examples of target and non-target stimuli to familiarize them with the stimulus combination and task. Thus, they were aware of the shape and size of the target figures but had no prior knowledge of their location.

We used a two-alternative forced-choice paradigm, with self-paced trials. Unless otherwise specified, target and non-target stimuli were presented for 500 ms each (in random order), followed by a 500 ms mask consisting of an equally-sized array of random black and white checks. Figure 3 illustrates the paradigm, with two examples of suprathreshold targets in panel A, and representative target and non-target pairs in panel B. In experiments in which stimulus duration was shortened to 125 or 250 ms, an intertrial interval was added so that the time from the onset of the first stimulus to the offset of the second stimulus remained 1000 ms. This allowed subjects to maintain a similar pacing throughout the trials.

Stimuli appeared abruptly from the background, which was at the mean luminance of the array. As in previous studies (Victor et al., 2005; Victor, Rizvi, et al., 2017; Victor et al., 2013; Victor et al., 2015), subjects were given practice (100 to 200 trials), without feedback, to become familiar with stimulus timing. After these initial practice trials, we did not observe any improvement either in the time to complete a block or overall fraction correct over time.

Target, non-target, and masks consisted of 64×64 arrays of checks, each measuring 9.8×9.8 min (array size, 10.5×10.5 deg), presented on a Dell Model 1707FPV LCD monitor, 1152 x 864 pixel resolution, 60 Hz refresh. Contrast was 1.0 and mean luminance was 81 cd/m^2. Viewing was binocular at 1m. Unless otherwise noted, the figure component of the target consisted of five randomly-placed non-overlapping circular disks, in aggregate accounting for one-quarter of the area of the display, and thus had a diameter of approximately 16 checks (2.6 deg). In one set of experiments, the figure shape

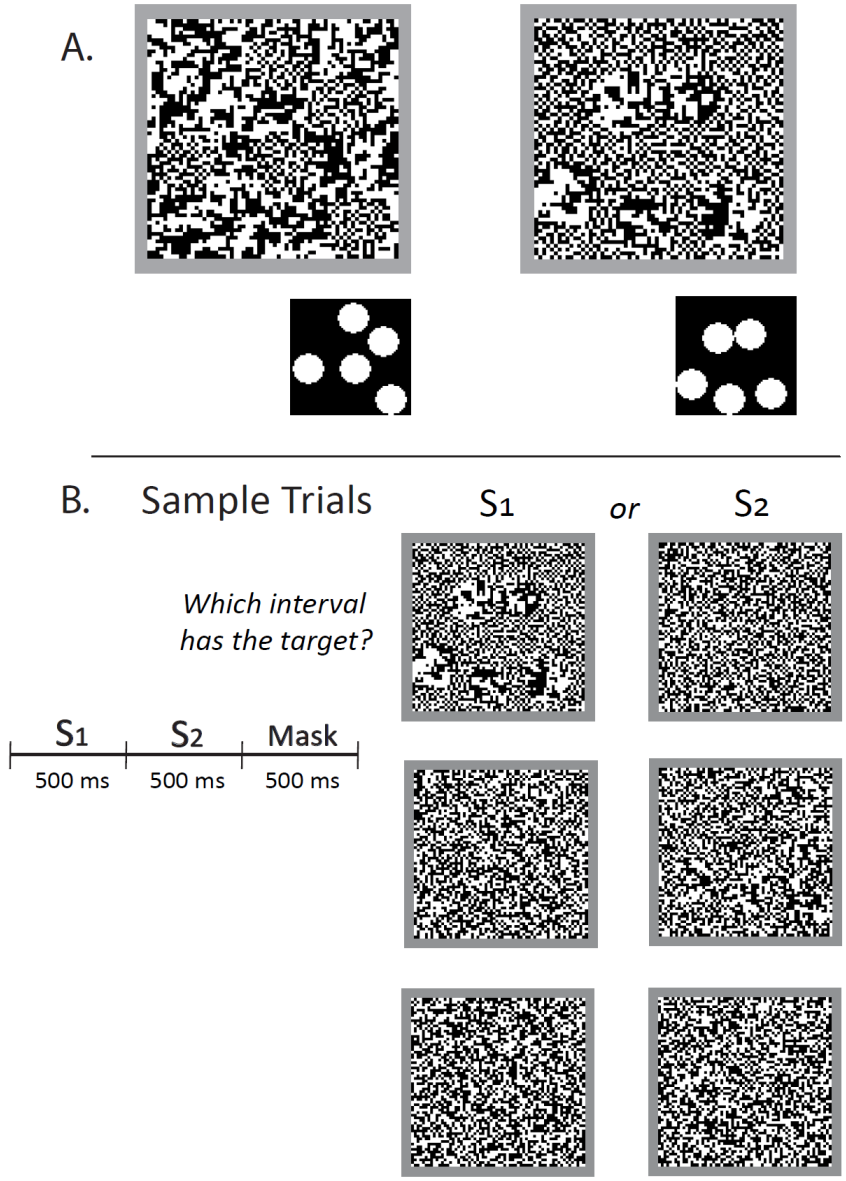

Figure 3. A: Examples of target stimuli along two rays in the $(\beta_{\_}^{max}, \beta_{|}^{max}) = (+1, +1)$-plane. Left, $(\beta_{\_fig}, \beta_{|fig}) = (-0.32, -0.32)$ and $(\beta_{\_gnd}, \beta_{|gnd}) = (0.32, 0.32)$; right: $(\beta_{\_fig}, \beta_{|fig}) = (0.32, 0.32)$ and $(\beta_{\_gnd}, \beta_{|gnd}) = (-0.32, -0.32)$, the maximum correlation strengths tested for this ray. Insets below each stimulus (not shown during the experiment) indicate locations of the five circular disks. Note that these examples are markedly suprathreshold, and therefore do not serve to indicate the psychophysical results. B: Left, experimental timeline; right, examples of two-alternative forced-choice trials showing targets and non-targets along the ray in Panel A, right. From top to bottom, targets are in S1, S2, and S1. Target examples cover the range of correlation strengths used along this ray: $(\beta_{\_fig}, \beta_{|fig}) = (0.32, 0.32)$ and $(\beta_{\_gnd}, \beta_{|gnd}) = (-0.32, -0.32)$ (top, same as in panel A right), $(\beta_{\_fig}, \beta_{|fig}) = (0.14, 0.14)$ and $(\beta_{\_gnd}, \beta_{|gnd}) = (-0.14, -0.14)$ (middle), and $(\beta_{\_fig}, \beta_{|fig}) = (0.04, 0.04)$ and $(\beta_{\_gnd}, \beta_{|gnd}) = (-0.04, -0.04)$ (bottom). Non-targets are uniform textures with statistics equal to the space-averaged value in the corresponding target: -0.16 for the non-target in the top row, -0.07 for the non-target in the middle row, and -0.02 for the non-target in the bottom row.

was modified to a shape consisting of the interior of a region bounded by four concave circular arcs. To

maintain a constant area, these figures had a diameter of approximately 31 checks (5.0 deg). All stimuli were created and presented with in-house MATLAB (MathWorks, Natick, MA) software, which also recorded the subjects' responses.

Subjects

Studies were conducted in 3 normal subjects (all female), ages 28-62, all of whom were experienced psychophysical observers; one was naïve to the purposes of the experiment. All had visual acuities (corrected if necessary) of 20/20 or better.

This work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), following approval of the Institutional Review Board of Weill Cornell and consent of the individual subjects.

Results

Figure-ground thresholds for single texture parameters

In the basic experiment, we determined threshold for identification of figure-vs-ground structure when the figure-ground distinction is determined by a single texture parameter. As described in Methods, a texture parameter value of zero corresponded to spatial randomness, while nonzero values corresponded to positive or negative correlation along a cardinal ($\beta_{\_}$ or $\beta_{|}$) or diagonal ($\beta_{\backslash}$, and $\beta_{/}$) direction. Of particular interest was whether thresholds depended on the sign of the correlation, and on whether the correlation was present in the figure vs. the ground.

Figure 4 shows example psychometric functions from a typical experiment. Each psychometric function corresponds to a ray in the figure-ground plane for $\beta_{\_}$ (see Figure 2), i.e., a set of trials in which figure and ground were defined by modulation of $\beta_{\_}$, and in which the values of $\beta_{\_}$ in figure and ground had a constant ratio. For example, along the ray designated by $(c_{fig}, c_{gnd}) = (+1, 0)$ (Figure 4A), the figure was defined by selecting $\beta_{\_,fig} > 0$, and ground was random ($\beta_{\_,gnd} = 0$). In Figure 4B, measurements were made along the ray designated by $(c_{fig}, c_{gnd}) = (+1, -1)$, where the figure was again defined by selecting $\beta_{\_,fig} > 0$ but the ground was defined by $\beta_{\_,gnd} = -\beta_{\_,fig}$. Note that the presence of a contrasting (i.e., opposite-sign) correlation in the ground lowered the threshold: 0.19 in Figure 4B, compared to 0.28 in Figure 4A. In all cases, the non-target stimulus consisted of a uniform texture defined by $\beta_{\_,nontarg} = \frac{1}{4}\beta_{\_,fig} + \frac{3}{4}\beta_{\_,gnd}$, the area-weighted average of the correlation strength in the figure-ground target. As detailed in Methods, thresholds were determined by the correlation strength required to achieve a fraction correct of 0.75, obtained by fitting a Weibull function (eq. (12)) to the data.

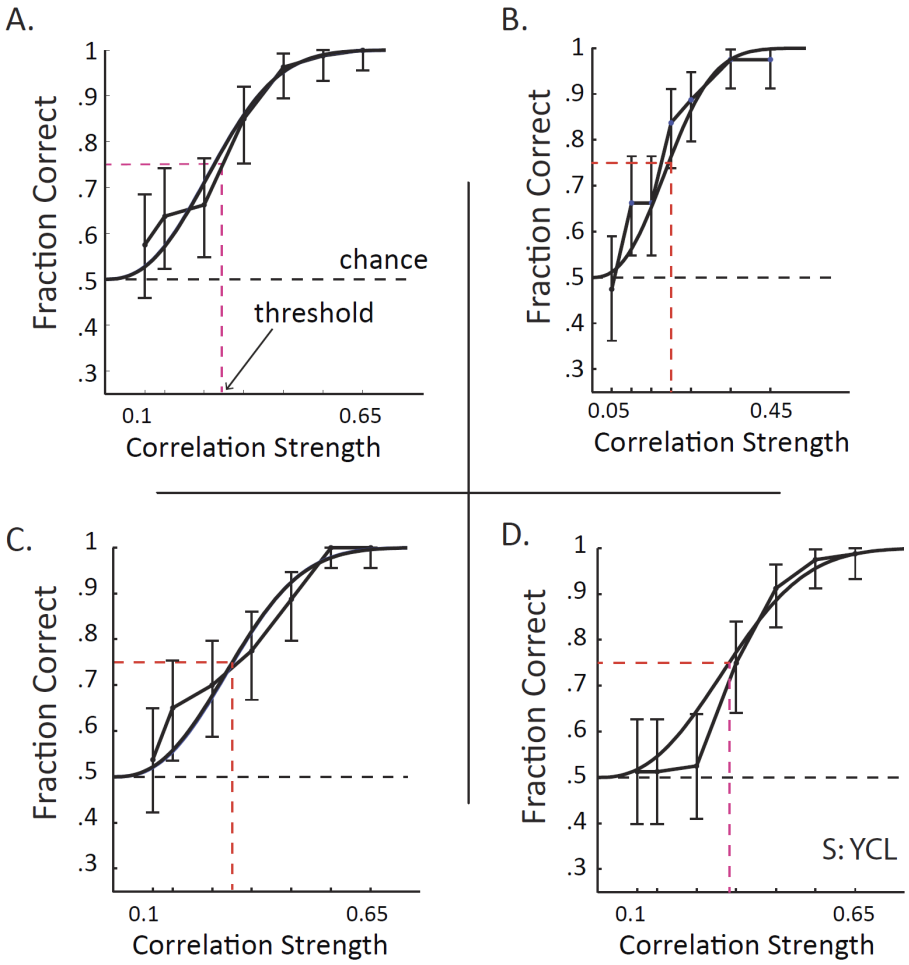

Figure 4. Psychophysical curves for representative rays in the figure-ground plane $\beta\_$. A, $(c_{fig}, c_{gnd}) = (+1, 0)$; B, $(c_{fig}, c_{gnd}) = (+1, -1)$; C, $(c_{fig}, c_{gnd}) = (-1, 0)$; D, $(c_{fig}, c_{gnd}) = (0, +1)$. Smooth curves are fits with Weibull functions, with a common value of the shape parameter $b_r = 2.49$ (eq. (12)). Error bars indicate 95% confidence intervals. Subject: YCL.

For this plane, thresholds for positive and negative correlations were similar (0.28 vs. 0.30, Figure 4A vs. Figure 4C, consistent with previous findings for a texture discrimination task (Victor, Rizvi, et al., 2017; Victor et al., 2015). Interestingly, however, thresholds were slightly higher when the correlation was present in the ground (0.33, Figure 4D), vs. in the figure (0.28, Figure 4A).

Thresholds in the figure-ground planes corresponding to $\beta\_$, $\beta|$, and $\beta\backslash$ are shown in Figure 5ABC, and demonstrate that this difference in thresholds was consistent across all subjects and directions of correlation. Specifically, when the figure was structured by pairwise correlation (or anti-

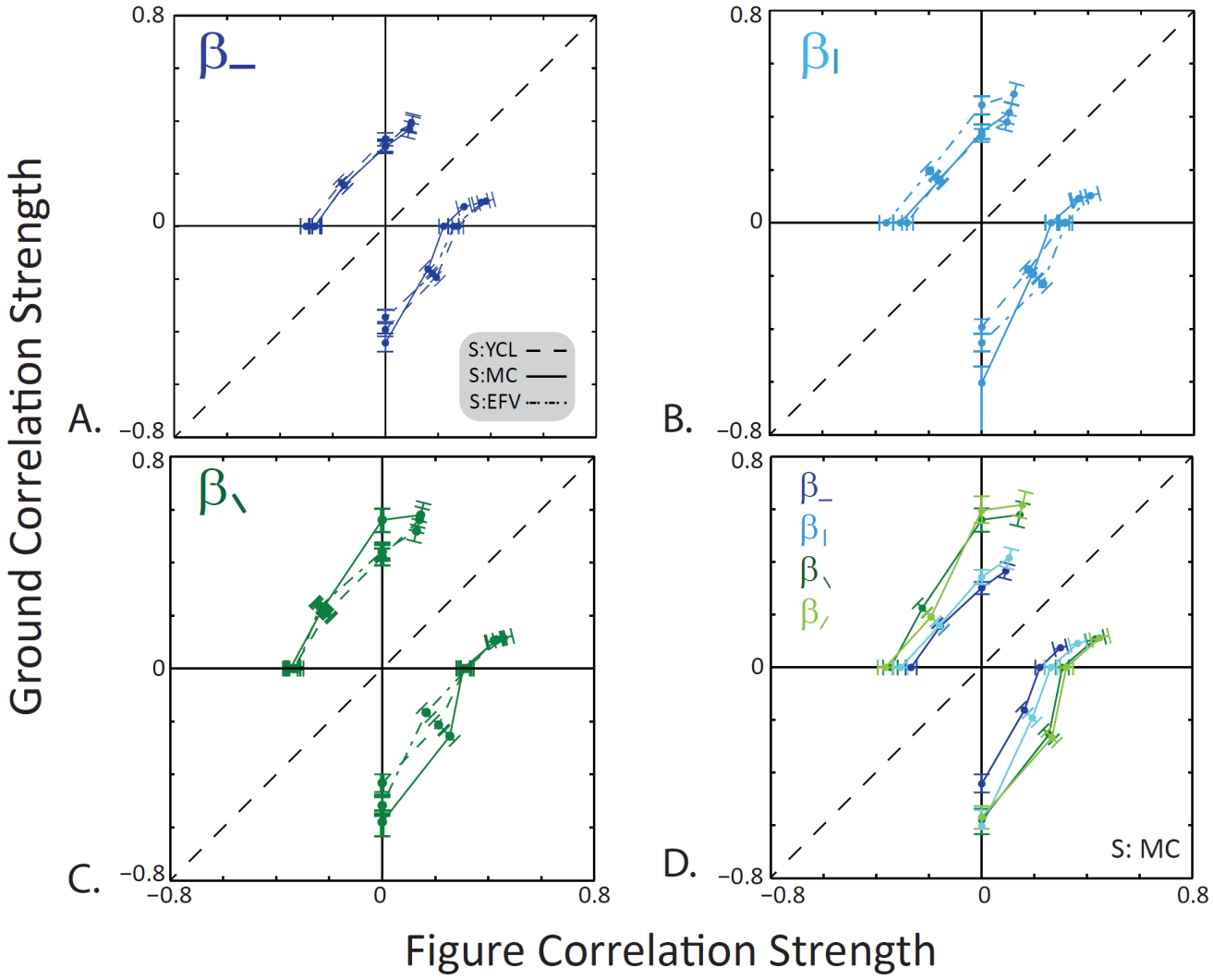

Figure 5. Thresholds in single-parameter figure-ground planes for $N = 3$ subjects: A, $\beta_{\_}$; B, $\beta_{|}$; C, $\beta_{\diagdown}$. D: Thresholds in four single-parameter figure-ground planes ($\beta_{\_}$, $\beta_{|}$, $\beta_{\diagdown}$, and $\beta_{/}$) in subject MC. Error bars indicate 95% confidence intervals. The psychophysical curves in Figure 4 correspond to four of the data points in panel A for subject YCL.

correlation) and the ground was random, thresholds were, on average, 30% lower than when the ground was structured and the figure was random. Subject MC was also tested with stimuli structured according to $\beta_{/}$, and, as expected, (Figure 5D) results were nearly identical to findings for $\beta_{\diagdown}$.

We quantified the difference in thresholds for figure and ground by computing the ratio

$$R = \frac{a_{gnd,+} + a_{gnd,-}}{a_{fig,+} + a_{fig,-}}, \tag{13}$$

Table 1

I. One texture parameter

	subject	R	c_{lo}	c_{hi}
β_-	EFV	1.24	1.15	1.34
	MC	1.52	1.40	1.64
	YCL	1.26	1.17	1.35
	AVG	1.33	1.27	1.39
$\beta_\mid$	EFV	1.33	1.23	1.43
	MC	1.66	1.51	2.11
	YCL	1.23	1.14	1.33
	AVG	1.39	1.32	1.50
$\beta_\backslash$	EFV	1.49	1.38	1.60
	MC	1.75	1.62	1.90
	YCL	1.32	1.22	1.42
	AVG	1.51	1.44	1.58
$\beta_/$	MC	1.70	1.57	1.84

II. Two texture parameters

		same sign			opposite sign		
	subject	R	c_{lo}	c_{hi}	R	c_{lo}	c_{hi}
$(\beta_-, \beta_\mid)$	EFV	0.98	0.91	1.06	1.78	1.64	1.92
	MC	1.06	0.98	1.15	2.64	2.41	2.88
	YCL	1.08	1.00	1.17	2.02	1.86	2.19
	AVG	1.04	1.00	1.09	2.12	2.02	2.22
$(\beta_-, \beta_\backslash)$	EFV	1.11	1.03	1.19	1.06	0.98	1.14
	MC	1.37	1.26	1.49	1.32	1.21	1.44
	YCL	1.26	1.16	1.36	1.30	1.21	1.40
	AVG	1.24	1.19	1.30	1.23	1.17	1.29
$(\beta_\mid, \beta_\backslash)$	EFV	1.15	1.07	1.23	1.07	0.99	1.15
	MC	1.34	1.21	1.47	1.49	1.37	1.62
	YCL	1.00	0.93	1.07	1.20	1.12	1.30
	AVG	1.15	1.10	1.21	1.24	1.19	1.30
$(\beta_\backslash, \beta_/)$	EFV	1.01	0.93	1.10	2.38	2.02	2.74
	MC	1.33	1.23	1.44	1.98	1.71	2.50
	YCL	0.91	0.83	1.00	1.91	1.77	2.07
	AVG	1.07	1.02	1.13	2.08	1.93	2.28

III. Two texture parameters, concstar

		same sign			opposite sign		
	subject	R	c_{lo}	c_{hi}	R	c_{lo}	c_{hi}
$(\beta_-, \beta_\mid)$	EFV	1.03	0.95	1.11	1.68	1.54	1.85
	MC	0.93	0.86	1.02	2.60	2.21	3.08
	YCL	0.94	0.86	1.02	1.95	1.80	2.12
	AVG	0.97	0.92	1.02	2.05	1.91	2.19

Table 1. Analysis of thresholds in figure-ground planes: I, planes defined by a single texture parameter; II, planes defined by pairs of texture parameters, III, planes defined by pairs of texture parameters, concstar shape. R is the ratio of thresholds for structured ground, vs. structured figure; see eq. (13). Confidence limits on R estimated from the measured thresholds and their confidence limits, using Gaussian approximations. For the rows labeled "AVG", thresholds are first averaged across subjects prior to calculating R.

where $a_{fig,+}$ is the threshold along the ray $(c_{fig}, c_{gnd}) = (+1, 0)$ in which figure correlations are positive,

$a_{fig,-}$ is the threshold along the ray $(c_{fig}, c_{gnd}) = (-1, 0)$ in which figure correlations are negative, and similarly for $a_{gnd,+}$ and $a_{gnd,-}$. This ratio isolates the difference in sensitivities to figure and ground, without regard to whether the correlations are positive or negative. The results of this analysis are shown in part I of Table 1, which indicates that for individual image statistics, thresholds for structure in the ground are 30 to 50% higher than thresholds for structure in the figure.

Figure-ground thresholds for pairs of texture parameters

The findings of Figure 5 and part I of Table 1 suggest that it is generally easier to identify figure-ground organization when the figure is correlated and the ground is random, than for the alternative organization of a random figure in a correlated ground. However, the next set of experiments shows that the situation is more complicated. Specifically (Figure 6), when figure or ground are defined by certain combinations of texture parameter pairs, this advantage may disappear, and for others, it may be increased.

Figure 6A and B, left panel, examine thresholds in figure-ground planes corresponding to combinations of the cardinal-direction image statistics, β_{-} and $\beta_{|}$. In Figure 6A, they are presented with the same sign in figure or ground (both positive or both negative), in Figure 6B, they have opposite sign. The resulting textures (see dark and light blue rectangles in Figure 1B1) are quite different: qualitatively, the same-sign textures vary in coarseness but have little orientation content, while the opposite-sign textures are strongly oriented. For the same-sign condition, there is essentially no asymmetry. That is, thresholds for a target with a structured figure and a random background (on the abscissa) are identical to the thresholds for a structured ground and a random figure (on the ordinate). In contrast, for the opposite-sign condition, asymmetry is greater than for the individual texture parameters. This finding is quantified by the ratio of thresholds for structured-ground vs. structured-figure, R (eq. (13)): it ranged from 0.98 to 1.08 for individual subjects for the same-sign case (cross-subject average,

1.04), and from 1.78 to 2.64 (cross-subject average, 2.12) for the opposite-sign case (part II of Table 1). In both cases, however, overall thresholds were lower when the two image statistics were combined than for individual statistics, by approximately a factor of $\sqrt{2}$, suggesting similar degrees of cue combination.

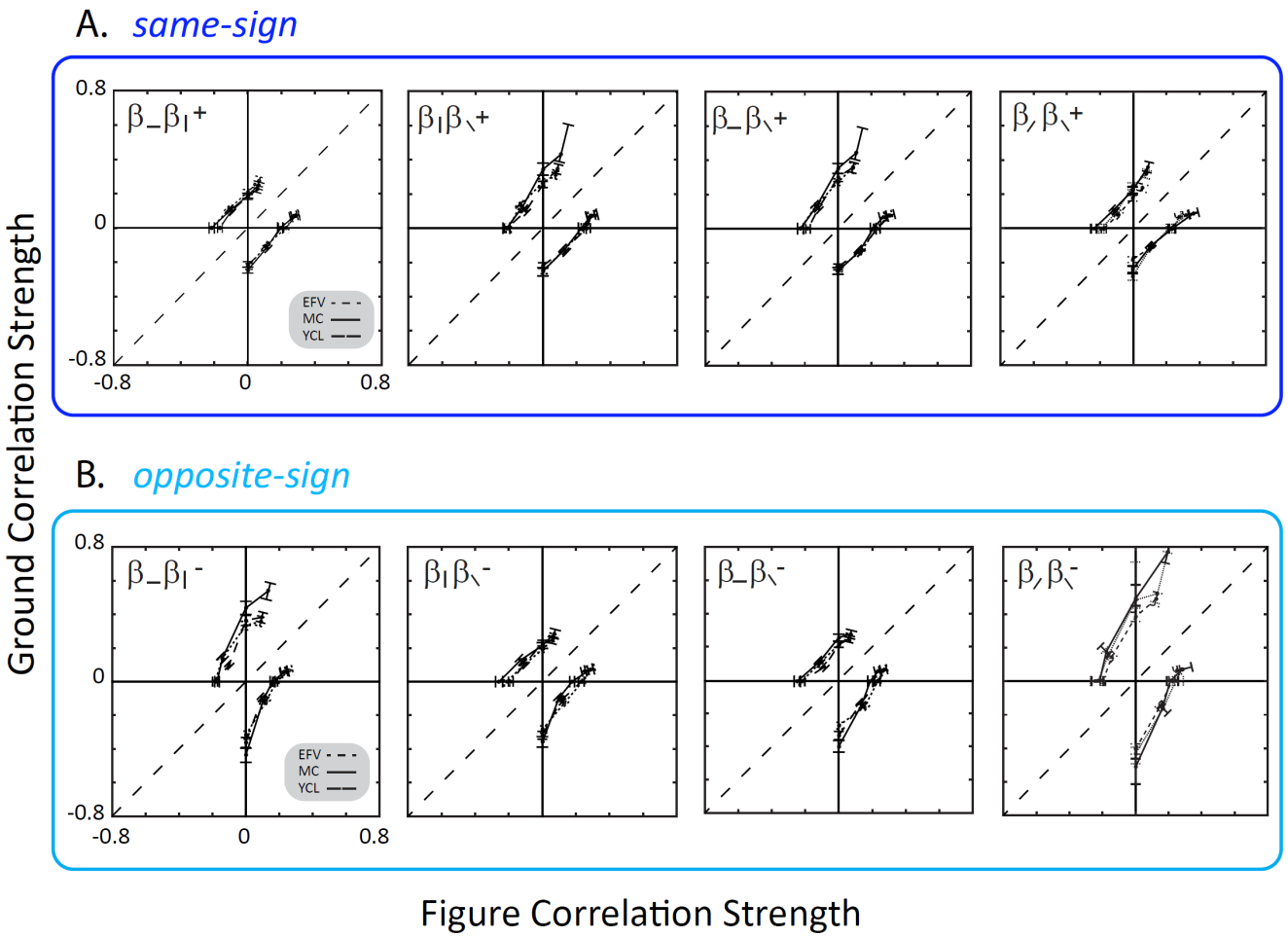

Figure 6. Thresholds in figure-ground planes for pairs of texture parameters for $N = 3$ subjects. A: Texture parameters have same sign. B: Texture parameters have opposite sign. Figure-ground planes, from left to right, are $(\beta\_, \beta_|)$, $(\beta_|, \beta_\backslash)$, $(\beta\_, \beta_\backslash)$, and $(\beta_/, \beta_\backslash)$. Error bars indicate 95% confidence intervals. (Note that one error bar is partly off-scale in the rightmost panel of B.)

Findings were similar for the two statistics that parametrized diagonal correlations $\beta_\backslash$ and $\beta_/$ (Figure 6A and B, right panels) : little or no difference in threshold when they were used with the same

sign to render figure or ground (cross-subject average, $R = 1.07$) but substantially higher threshold when they were used with opposite sign to render ground, vs. figure (cross-subject average, $R = 2.08$). However, for stimuli in which correlations along a cardinal axis (e.g., $\beta_{\_}$) axis and a diagonal axis (e.g., $\beta_{\backslash}$) are mixed (Figure 6A and B, middle panels), there was a small but consistent asymmetry in both same-sign and opposite-sign conditions ($R = 1.15$ to $R = 1.24$). These intermediate values indicate that figure-ground asymmetry is not merely a consequence of whether the statistics have the same sign or not.

Concave borders

While the above findings show that image statistics within the two regions of the stimulus designated as figure and ground are treated differently, the possibility remains that this difference is not driven by the figure-ground distinction *per se*, but by the shape of the contour that forms their border(Schmidtmann, Jennings, & Kingdom, 2015). In particular, the convex side of a contour tends to be assigned to objects, while the concave side tends to be assigned to ground(Zhaoping, 2005). Thus, it is possible that the asymmetry we observe is due to the shape of the border, rather than figure-ground assignment *per se*.

To examine this possibility, we dissociated figure from convexity by replacing the convex figure (a circular disk) by a concave figure ("concstar", consisting of four circular arcs). As shown in Figure 7, when the figure border was concave, we again found that for opposite-sign pairs of statistics (in this case, $\beta_{\_}$ and $\beta_{|}$), thresholds were lower when the figure was structured, vs. when the ground was structured. Had the asymmetry been driven by convexity of the border, then the asymmetry would have been reversed. The degree of asymmetry (Table 1, part III) was similar to what was seen for circular figures (cross-subject average, $R = 2.05$). Also, as was seen for circular figures, the asymmetry was absent when the statistics were of the same sign (cross-subject average, $R = 0.97$).

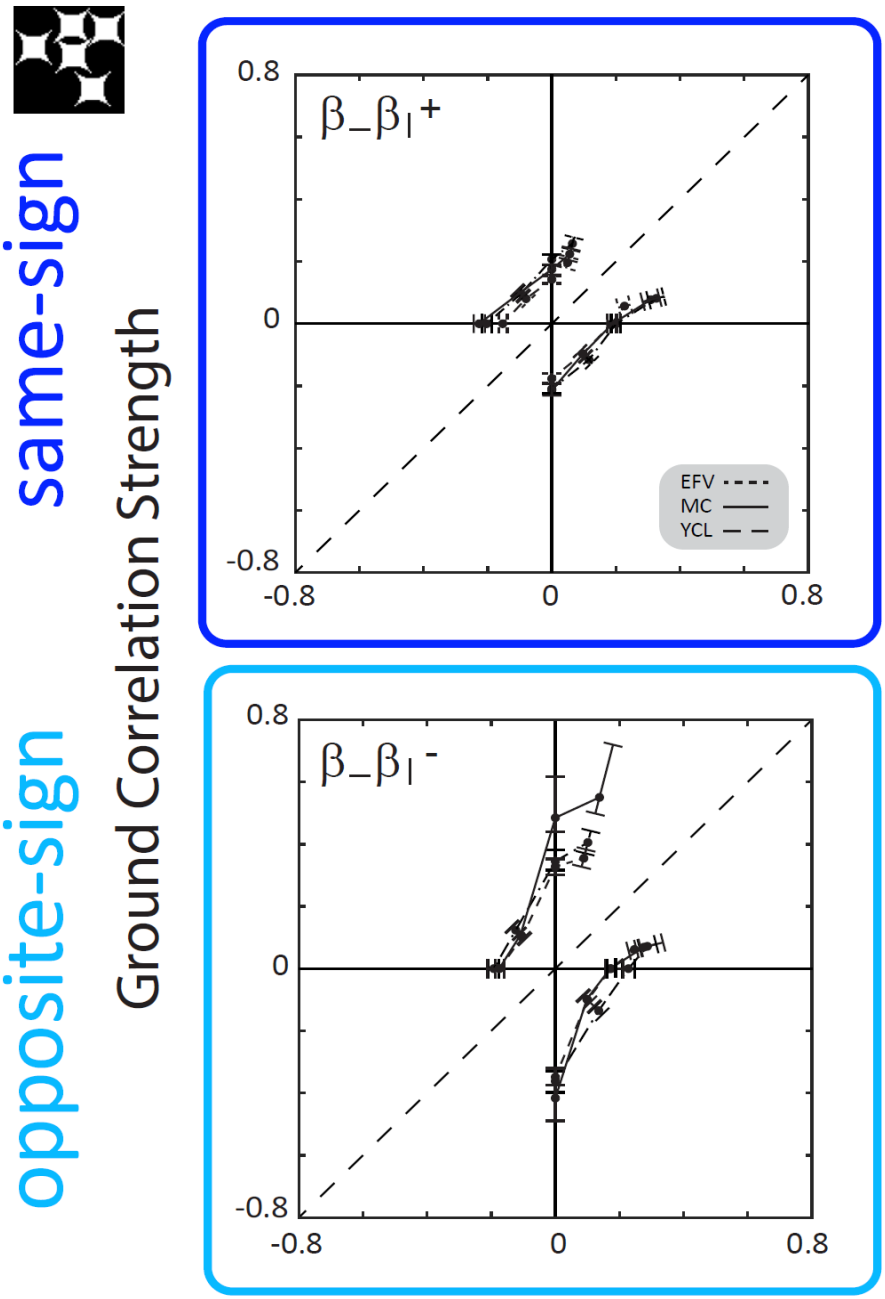

Figure 7. Thresholds in figure-ground planes for $(\beta\_, \beta|)$ for the concstar object shape. Error bars indicate 95% confidence intervals. Note similarity to left panels of Figure 6, obtained with the same parameters but circular objects.

Oriented power

The results of Figure 5,Figure 6, and Figure 7 indicate that the advantage of a structured figure depends in an apparently complex manner on the specific texture parameters. It is present for all

individual second-order parameters, it is eliminated when two cardinal or diagonal statistics are combined with the same sign, it is reinforced when they are combined with opposite sign, and it is

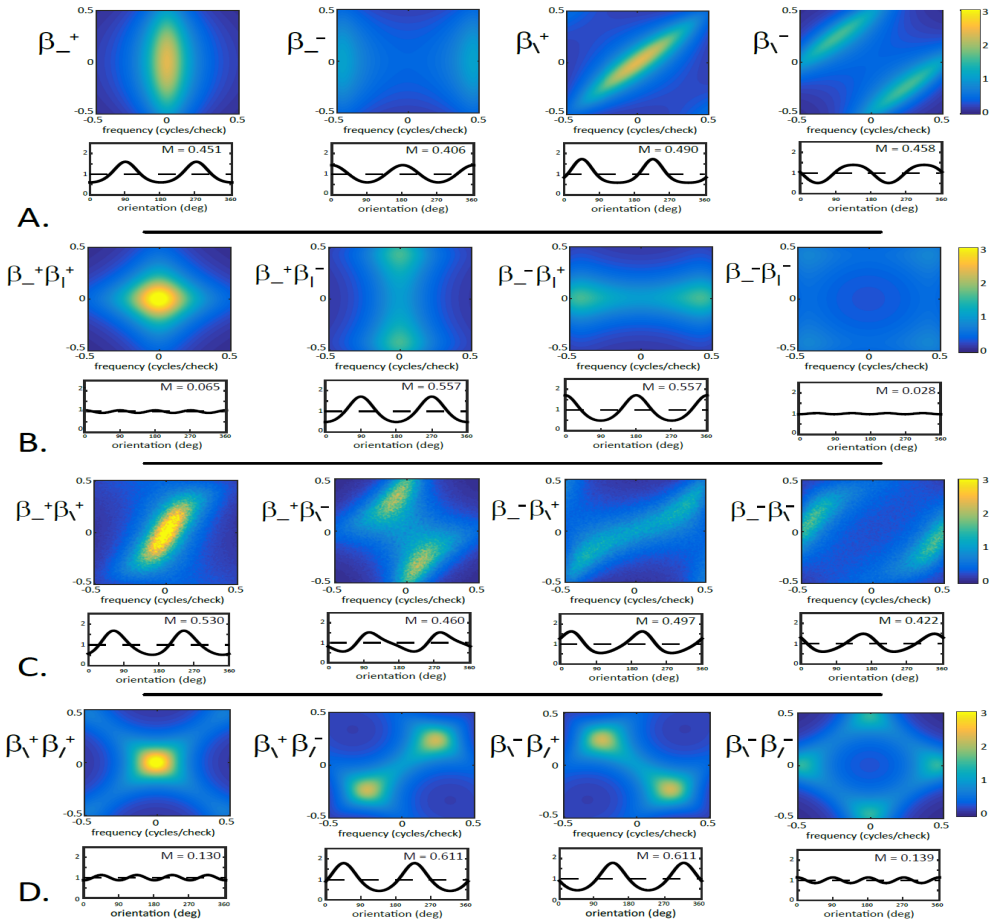

Figure 8. Power spectra and oriented power for selected textures. Heatmaps show power spectra up to 0.5 cycles per check; the same scale is used across all panels. Line graphs show power as a function of orientation, $P(\theta)$, calculated from the spectra via eq. (9). Michelson contrast for oriented power, (M, eq. (14)), is inset into each plot. A: $\beta\_ = +0.4$, $\beta\_ = -0.4$, $\beta\backslash = +0.4$, $\beta\backslash = -0.4$. B: $(\beta\_, \beta|) = (+0.28, +0.28)$, $(\beta\_, \beta|) = (+0.28, -0.28)$, $(\beta\_, \beta|) = (-0.28, +0.28)$, $(\beta\_, \beta|) = (-0.28, -0.28)$. C: $(\beta\_, \beta\backslash) = (+0.28, +0.28)$, $(\beta\_, \beta\backslash) = (+0.28, -0.28)$, $(\beta\_, \beta\backslash) = (-0.28, +0.28)$, $(\beta\_, \beta\backslash) = (-0.28, -0.28)$. D: $(\beta\backslash, \beta/) = (+0.28, +0.28)$, $(\beta\backslash, \beta/) = (+0.28, -0.28)$, $(\beta\backslash, \beta/) = (-0.28, +0.28)$, $(\beta\backslash, \beta/) = (-0.28, -0.28)$. Spectra in A, B, and D calculated analytically from eqs. (2)-(7); spectra in C calculated empirically as described in Methods.

reduced but not eliminated when cardinal and diagonal statistics are mixed, independent of their relative sign. Here we test the hypothesis that this behavior has a simple underlying explanation: that the asymmetry is driven by the orientation content of the texture defined by the statistics.

To analyze this possibility, we first quantified the orientation content of the texture stimuli,

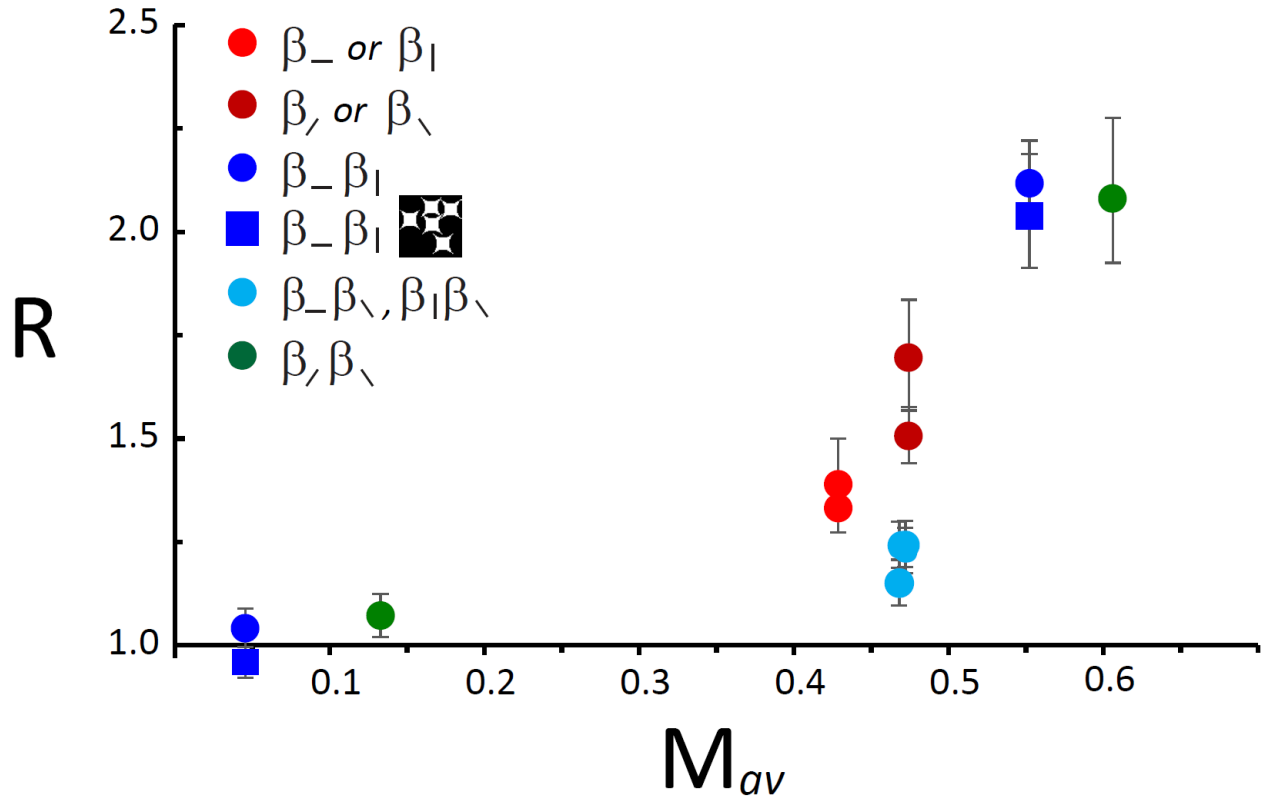

Figure 9. Correspondence between the figure-ground asymmetry of thresholds (R , eq. (13)) and oriented power, M_{av}. R is determined from the average thresholds across subjects (Table 1). M_{av}, the oriented power in each figure-ground plane, is calculated as described in the text and eq. (14). Symbol shape indicates the shape of the figures: circular or concstar. $r = 0.73$. For thresholds in conditions defined by pairs of cardinal-direction or diagonal-direction statistics, the same-sign combination has the lower value of M_{av} (less than 0.15). Error bars indicate 2 SEM for the average across subjects.

recognizing that, in contrast to a standard grating stimulus, multiple orientations are simultaneously present (Victor et al., 2013). To do this, we computed the power spectra of homogeneous samples of the textures, and, from the power spectra, the amount of power as a function of orientation, $P(\theta)$, eq. (9). Figure 8 shows these calculations, with total correlation strength set to 0.4 ($|\beta|$ for single-parameter textures, $\sqrt{\beta_1^2 + \beta_2^2}$ for two-parameter textures), comparable to the psychophysical thresholds. We then used the Michelson contrast of $P(\theta)$, namely,

$$M = \frac{\max(P(\theta)) - \min((P(\theta))}{\max(P(\theta)) + \min((P(\theta))}, \quad (14)$$

as an overall measure of the extent to which the texture's power was oriented (0 for isotropic, 1 for maximally anisotropic). Finally, to obtain an average measure M_{av} of oriented power for a figure-ground plane, we averaged the Michelson contrasts M for positive and negative polarities (i.e., the pairs shown in Figure 8).

Figure 9 shows the relationship between oriented power M_{av} and the asymmetry of thresholds for figure vs. ground, R (eq. (13)). These two quantities were strongly correlated ($r = 0.73$). This relationship encompasses all the figure-ground planes studied. It includes figure-ground planes defined by single texture parameters and by texture parameter pairs, and those defined by texture parameters of matched or opposite polarity. The data obtained with convex (circular disk) and concave (concstar) objects are also combined in this analysis. Similar results were obtained for other choices of texture contrast and the parameters in eq. (9) used to quantify orientation content.

Time course

The findings so far indicate that processing of local orientation differs in figure vs. ground. Since, in these stimuli, the distinction between figure and ground is solely driven by the statistics of local orientation, these findings imply (see Discussion) that the separation of figure and ground relies on a recurrent process. That is, identification of texture differences is necessary to generate a tentative organization of the image into figure and ground, and the observed difference between figure and ground can only arise once a tentative scene organization is generated. This interdependence suggests that the figure-ground asymmetry might evolve slowly in time.

We therefore sought to resolve the evolution of the figure-ground asymmetry by varying the presentation time of the stimuli. We restricted measurements to key conditions: the rays along the

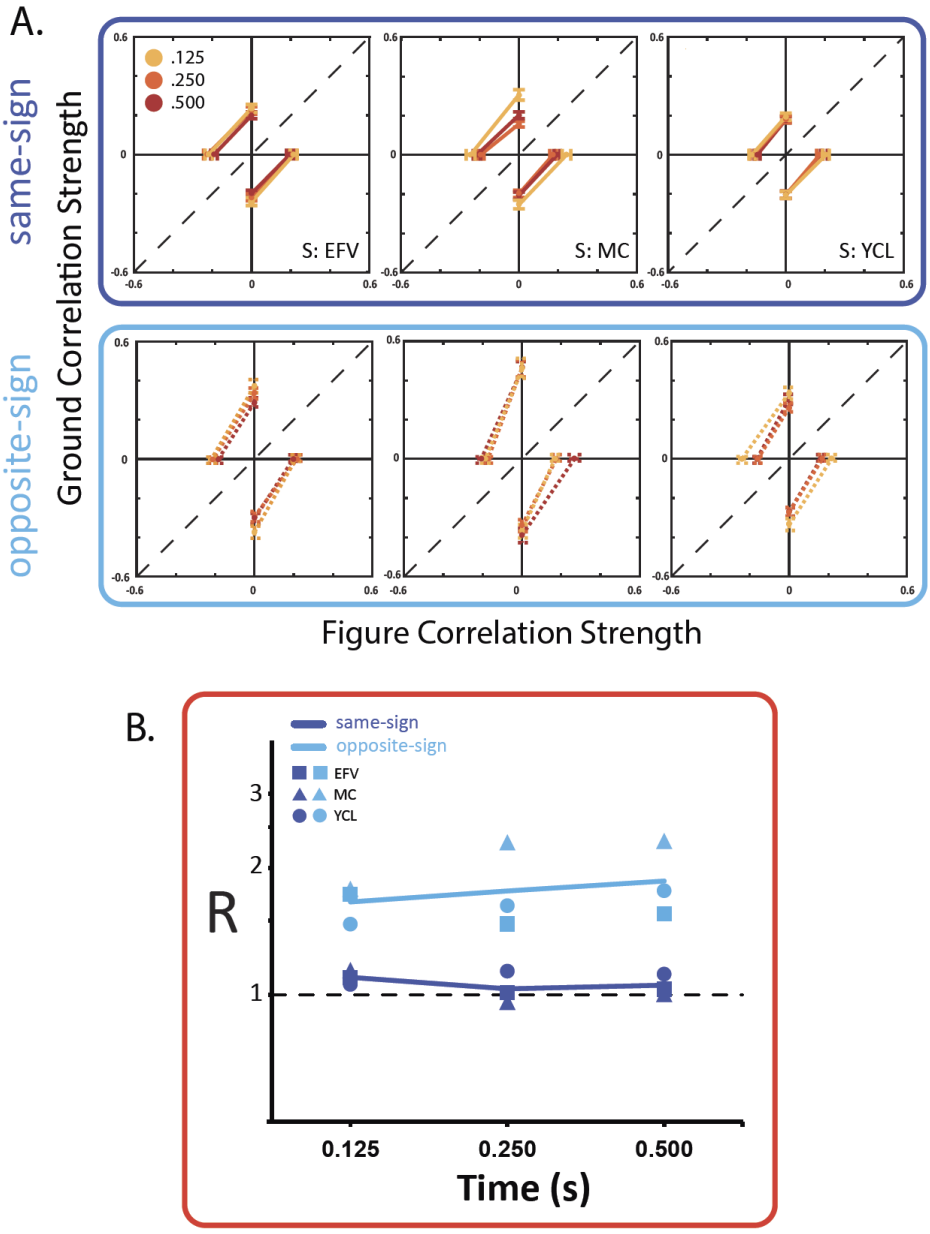

Figure 10. A: Figure-ground asymmetry for a range of presentation times for same-sign modulation of β_{-} and $\beta_{|}$ (upper row) and opposite-sign modulation of β_{-} and $\beta_{|}$ (lower row). Each plot shows data from a single subject (top and bottom row) at three presentation times. Color code for presentation time is identical in all subpanels. B: Asymmetry index (R , eq. (13)) as a function of presentation time.

cardinal axes in the figure-ground planes for β_{-} and $\beta_{|}$ -- so that either figure or ground was structured, and the other was random. Stimuli were presented for durations of 125, 250, or 500 ms. The 500 ms presentation time duplicated the conditions reported above, but we carried out a replication in blocks interleaved with the shorter presentation times. For the shorter presentation times, a blank interstimulus interval was inserted between the two alternative images, to allow subjects to keep the same pacing.

Figure 10A shows the findings for the three subjects. As expected, no asymmetry was present for the non-oriented (same-sign modulation of β_{-} and $\beta_{|}$), at all presentation times. For the oriented condition (opposite-sign modulation of β_{-} and $\beta_{|}$), a strong asymmetry was already present at 125 ms. Figure 10B, which summarizes the difference between figure and ground via the threshold ratio (13), suggests an increase in the asymmetry from 125 ms to 500 ms, but this increase is minimal.

Discussion

Separation of figure and ground is a critical part of mid-level visual processing, and visual texture is well-recognized as an important contributor to this process. Since a texture cue can only drive figure-ground separation if the figure and ground textures themselves can be discriminated, a natural hypothesis is that thresholds for figure-ground discrimination are determined by the thresholds for texture discrimination. However, this need not be the case: in a texture discrimination task, the two textures, by definition, play equivalent roles. In contrast, figure and ground are not interchangeable parts of typical natural images.

Here we investigated the role of texture in figure-ground separation, examining thresholds for figure and ground textures that differed in spatial scale and the extent to which their local features were oriented. This texture space is parameterized by pairwise correlations between adjacent checks. A key technical advantage of this set of textures is that, because textures are specified by purely local statistics and generated by a local algorithm, it is possible to create images with abrupt transitions between figure and ground without introducing edge artifacts at the figure-ground boundary. For example, even though figure and/or ground may have different orientation contents (e.g.,Figure 1), the figure-ground boundary does not have an excess of terminators, which would induce an illusory contour(Peterhans & von der Heydt, 1989). Such terminators would have been present, for example, had the oriented textures been generated from patches of gratings. These edge artifacts can be reduced, but not eliminated, by smoothing procedures(Hunt & Meinhardt, 2021).

The space has conceptual advantages as well. It is a subspace of the 10-parameter texture space previously introduced (Victor & Conte, 2012) that also includes higher-order correlations; as in the larger space, textures are as random as possible given the specified local correlations, and texture parameters can be specified independently as positive, negative, or zero. Texture discrimination

thresholds in this space have been well-characterized, and are closely-approximated by a simple Euclidean distance (Victor et al., 2015). In particular, thresholds for positive and negative correlations are essentially identical, a finding we reproduce here: the thresholds in Figure 5 and Figure 6 are symmetric about the origin. Additionally, although these textures are entirely synthetic and entirely composed of black and white checks, sensitivity to differences in image statistics correspond closely to the informativeness of the image statistic in natural images(Hermundstad et al., 2014).

Our main finding is that, while the threshold for figure-ground discrimination is independent of whether the statistical structure is present in the figure vs. in the ground for some kinds of image statistics, there is a substantial departure from this behavior when the figure or ground is defined by a strongly oriented structure. Specifically, when the ground texture is oriented and the figure is random, thresholds can be twice as high as in the opposite scenario: an oriented figure texture and a random ground texture. This asymmetry is closely correlated with the degree of orientation anisotropy in the texture that makes up the figure or the ground (Figure 9).

This difference between figure and ground reflects an increase in threshold for the oriented-ground configuration, rather than a decrease in threshold for the oriented-figure configuration. The evidence is seen in Figure 6: the threshold for the oriented-figure configuration in the $(\beta_{\_}^{max}, \beta_{|}^{max}) = (+1, -1)$ -plane is comparable to the threshold for a configuration in which same local image statistics combine to produce an approximately isotropic texture $(\beta_{\_}^{max}, \beta_{|}^{max}) = (+1, +1)$, while the threshold for the oriented-ground configuration in the $(\beta_{\_}^{max}, \beta_{|}^{max}) = (+1, -1)$ -plane is twice as high. Moreover, when correlations are present in the figure, thresholds in either the $(\beta_{\_}^{max}, \beta_{|}^{max}) = (+1, +1)$ -plane or the $(\beta_{\_}^{max}, \beta_{|}^{max}) = (+1, -1)$ -plane are reduced by approximately a factor of $\sqrt{2}$ from the thresholds for either statistic presented separately. In contrast, the threshold for

the combined correlations $(\beta_{\_}^{max}, \beta_{|}^{max}) = (+1, -1)$ in the ground is *larger* than for either statistic presented separately (compare Figure 5AB with Figure 6AB left panels).

Our interpretation of the results as evidence of a difference in processing of oriented local features in figure and ground rests on the exclusion of a range of alternative explanations. Figure and ground constitute, respectively, 0.25 and 0.75 of the total area of the display, but this ratio is constant across all experimental conditions. Hence, the difference in area cannot account for the specific pattern of asymmetry: present for some combinations of image statistics and absent for others.

A related possibility is that asymmetry arises because (hypothetically) the visual system has mechanisms that can detect increments of oriented power relative to the spatial average, but not decrements – a possibility suggested by the asymmetries of the field-capture mechanisms proposed for discrimination of luminance statistics in spatially uncorrelated textures(Silva & Chubb, 2014). If present, a predominance of increment-detection mechanisms could contribute to our findings: in the structured-figure condition, the correlation strength in the figure, c_{fig}, is four times the mean correlation strength ($c_{fig} = 4c_{nontarg}$ in eq. (11) when $c_{gnd} = 0$), while in the structured-ground condition, the correlation strength in the ground, c_{gnd}, is only 33% more than the mean correlation strength ($c_{gnd} = \frac{4}{3}c_{nontarg}$ in eq. (11) when $c_{fig} = 0$). However, a re-analysis of previous studies from our laboratory (Victor et al., 2015) shows that the contribution of any such asymmetry in visual mechanisms is too limited to account for our findings. The previous studies required that subjects perform a segmentation task, in which they localized a texture-defined strip that ran from one edge of the stimulus to the other. As in the present study, the strip accounted for 25% of the stimulus area, and there were separate trials in which the strip was structured and the background was random, and trials with the opposite assignment. The original publication reported the thresholds from responses pooled across trial

types. When those two trial types are analyzed separately, the difference in thresholds was approximately one-third of the asymmetry in the figure-ground task reported here. Specifically, for the conditions with the greatest asymmetry (opposite-sign conditions in Figure 6B), which yielded an asymmetry of more than a factor of 2 here (part II of Table 1), the asymmetry in the segmentation task was no more than 1.32:1. Finally, the distinction between the strong asymmetry found here, vs. the minimal asymmetry seen when the target contained strips (Victor et al., 2015), cannot be due to a hypothetical matching of center-surround "structure detectors" with the circular disks that constitute the figure. Had that been the case, then the asymmetry would invert when the disks are replaced by concave objects – but we find instead that the asymmetry retains its sign and magnitude (Figure 7 and Table 1, part III).

Independent evidence against the hypothesis that the observed asymmetry is due to limitations of underlying detection mechanisms is that, as mentioned above, thresholds *increase* when oriented statistics are combined in ground but *decrease* when they are combined in the figure (Figure 5AB vs. left panels of Figure 6AB).

Our data also exclude other classes of explanation. While the largest asymmetries are seen for a combination of a positive correlation along one axis and a negative correlation along another axis, neither a negative correlation nor a combination of two image statistics are needed to drive the asymmetry: substantial asymmetries are seen when correlations are present along just one axis (Figure 5), as well as for combinations of positive correlations along cardinal and diagonal directions (Figure 6). These observations, along with the correspondence between the amount of oriented power and the size of the figure-ground asymmetry (Figure 9), support the simple and concise interpretation that the observed asymmetry is driven by oriented features.

As we show, these features are processed differently, depending on whether they are on one side or the other of a curved boundary. The possibility that this processing difference is driven by convexity vs. concavity, rather than by border ownership, is excluded: the same asymmetry is present when the figure boundary is concave (Figure 7), as when it is convex.

Thus, we are forced to the conclusion that our results require a functional recursion involving processes that extract local features, and processes that infer scene organization. The asymmetry in processing local features depends on whether they are considered to be part of the figure, vs. part of the ground – but figure and ground can only be separated by processing these local features. We speculate that this functional recursion corresponds to feedback of border ownership signals emerging in V2 or V4 (Chen et al., 2014; Hu, von der Heydt, & Niebur, 2019; Poort, Self, van Vugt, Malkki, & Roelfsema, 2016; Qiu & von der Heydt, 2005; Roelfsema, Lamme, Spekreijse, & Bosch, 2002; Zhou, Friedman, & von der Heydt, 2000) to neurons that extract oriented power in V1 (David & Gallant, 2005; Hubel & Wiesel, 1977; Ringach, Shapley, & Hawken, 2002; Rust, Schwartz, Movshon, & Simoncelli, 2005; Tao, Shelley, McLaughlin, & Shapley, 2004). Physiological studies in macaque indicate that such feedback is rapid (Chen et al., 2014; Poort et al., 2016; van Kempen et al., 2021), and would account for our findings that the asymmetry is nearly maximal at 125 ms (Figure 10). We also note that this postulated role for extrastriate areas is separate from their role (Freeman, Ziemba, Heeger, Simoncelli, & Movshon, 2013; Yu, Schmid, & Victor, 2015) in processing higher-order local image statistics: the findings reported here reflect an interaction between simple pairwise statistics (i.e., spectral content), and overall scene organization.

Whatever their anatomical substrate, the computations required to account for our findings are difficult to square with a purely feedforward model for texture analysis (Chubb & Landy, 1991; DiMattina & Baker, 2021; Hunt & Meinhardt, 2021; Wilson, 1993), as such models would not be able to

account for a difference in sensitivity that depends on global scene organization(Groen, Ghebreab, Prins, Lamme, & Scholte, 2013). We speculate that this interaction between scene organization and sensitivity to oriented textures is a way that visual computations incorporate priors about natural images: that textures of objects tend to be more oriented than textures of backgrounds.

While functional recurrence is a prominent aspect of many models for figure-ground segregation (Chen et al., 2014; Craft, Schutze, Niebur, & von der Heydt, 2007; Grossberg & Mingolla, 1985; Grossberg & Pessoa, 1998; Hu et al., 2019; Poort et al., 2016; Roelfsema et al., 2002), our findings require a form of recurrence that is distinctive and specific. It is notable that the effect of assignment to figure vs. ground on processing of local statistics is present only when an oriented structure is created (e.g., horizontal and vertical correlations of opposite sign), but not when the statistics merely lead to larger blobs (e.g., horizontal and vertical correlations of the same sign). In both cases, however, when these combined correlations are present in the ground, thresholds are nearly identical, and are lower than when correlations are present along only one orientation. So, the performance we observe requires both orientation-selective pooling (when the correlations have the opposite sign) and pooling across orientations (when the correlations have the same sign). The effect of a scene organization that assigns a region to the ground leads to a selective elevation of the threshold in the opposite-sign case, indicating that orientation-selective pooling is attenuated by top-down influences, while pooling across orientations is not.

Acknowledgements

Portions of this work were presented at the 2020 and 2021 meetings of the Vision Sciences Society (St. Petersburg, FL). This work was supported by NIH NEI EY7977.

References

Baylis, G. C., & Cale, E. M. (2001). The figure has a shape, but the ground does not: evidence from a priming paradigm. *J Exp Psychol Hum Percept Perform, 27*(3), 633-643. doi:10.1037//0096-1523.27.3.633

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*(303), 696-698.

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., & Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron, 82*(3), 682-694. doi:10.1016/j.neuron.2014.03.023

Cheng, A., Walther, D. B., Park, S., & Dilks, D. D. (2021). Concavity as a diagnostic feature of visual scenes. *Neuroimage, 232*, 117920. doi:10.1016/j.neuroimage.2021.117920

Chubb, C., & Landy, M. (1991). Orthogonal distribution analysis: a new approach to the study of texture perception. In M. S. Landy, and Movshon, J.A. (Ed.), *Computational models of visual processing* (pp. 291-301). Cambridge, MA: MIT Press.

Chubb, C., Landy, M. S., & Econopouly, J. (2004). A visual mechanism tuned to black. *Vision Res, 44*(27), 3223-3232.

Chubb, C., Scofield, I., Chiao, C.-C., & Sperling, G. (2012). A method for analyzing the dimensions of preattentive visual sensitivity. *J. Math. Psychol., 56*(6), 427-443.

Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *J Neurophysiol, 97*(6), 4310-4326.

David, S. V., & Gallant, J. L. (2005). Predicting neuronal responses during natural vision. *Network, 16*(2-3), 239-260.

DiMattina, C., & Baker, C. L., Jr. (2021). Segmenting surface boundaries using luminance cues. *Sci Rep, 11*(1), 10074. doi:10.1038/s41598-021-89277-2

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Res, 94*, 62-75. doi:10.1016/j.visres.2013.11.004

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nat Neurosci, 16*(7), 974-981. doi:10.1038/nn.3402

Groen, II, Ghebreab, S., Prins, H., Lamme, V. A., & Scholte, H. S. (2013). From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *J Neurosci, 33*(48), 18814-18824. doi:10.1523/JNEUROSCI.3128-13.2013

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Percept Psychophys, 38*(2), 141-171.

Grossberg, S., & Pessoa, L. (1998). Texture segregation, surface representation and figure-ground separation. *Vision Res, 38*(17), 2657-2684.

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., & Tkacik, G. (2014). Variance predicts salience in central sensory processing. *Elife, 3*. doi:10.7554/eLife.03722

Hu, B., von der Heydt, R., & Niebur, E. (2019). Figure-Ground Organization in Natural Scenes: Performance of a Recurrent Neural Model Compared with Neurons of Area V2. *eNeuro, 6*(3). doi:10.1523/ENEURO.0479-18.2019

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci, 198*(1130), 1-59.

Hunt, C., & Meinhardt, G. (2021). Synergy of spatial frequency and orientation bandwidth in texture segregation. *J Vis, 21*(2), 5. doi:10.1167/jov.21.2.5

Julesz, B., & Bergen, J. R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *The Bell System Technical Journal, 62*(6), 1619-1645.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447*(7141), 206-209.

Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J Neurosci, 9*, 1731-1748.

Poort, J., Self, M. W., van Vugt, B., Malkki, H., & Roelfsema, P. R. (2016). Texture Segregation Causes Early Figure Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual Cortex. *Cereb Cortex, 26*(10), 3964-3976. doi:10.1093/cercor/bhw235

Qiu, F. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: v2 combines stereoscopic cues with gestalt rules. *Neuron, 47*(1), 155-166.

Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation selectivity in macaque V1: diversity and laminar dependence. *J Neurosci, 22*(13), 5639-5651.

Roelfsema, P. R., Lamme, V. A., Spekreijse, H., & Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. *J Cogn Neurosci, 14*(4), 525-537. doi:10.1162/08989290260045756

Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2005). Spatiotemporal elements of macaque v1 receptive fields. *Neuron, 46*(6), 945-956.

Schmid, A. M., & Victor, J. D. (2014). Possible functions of contextual modulations and receptive field nonlinearities: pop-out and texture segmentation. *Vision Res, 104*, 57-67. doi:10.1016/j.visres.2014.07.002

Schmidtmann, G., Jennings, B. J., & Kingdom, F. A. (2015). Shape recognition: convexities, concavities and things in between. *Sci Rep, 5*, 17142. doi:10.1038/srep17142

Sharan, L., Liu, C., Rosenholtz, R., & Adelson, E. H. (2013). Recognizing Materials using Perceptually Inspired Features. *Int J Comput Vis, 103*(3), 348-371. doi:10.1007/s11263-013-0609-0

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Accuracy and speed of material categorization in real-world images. *J Vis, 14*(9). doi:10.1167/14.9.12

Silva, A. E., & Chubb, C. (2014). The 3-dimensional, 4-channel model of human visual sensitivity to grayscale scrambles. *Vision Res, 101*, 94-107. doi:10.1016/j.visres.2014.06.001

Sprote, P., & Fleming, R. W. (2013). Concavities, negative parts, and the perception that shapes are complete. *J Vis, 13*(14). doi:10.1167/13.14.3

Tao, L., Shelley, M., McLaughlin, D., & Shapley, R. (2004). An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proc Natl Acad Sci U S A, 101*(1), 366-371. doi:10.1073/pnas.2036460100

Tesileanu, T., Conte, M. M., Briguglio, J. J., Hermundstad, A. M., Victor, J. D., & Balasubramanian, V. (2020). Efficient coding of natural scene statistics predicts discrimination thresholds for grayscale textures. *Elife, 9*. doi:10.7554/eLife.54347

Tkacik, G., Prentice, J. S., Victor, J. D., & Balasubramanian, V. (2010). Local statistics in natural scenes predict the saliency of synthetic textures. *Proc Natl Acad Sci U S A, 107*(42), 18149-18154. doi:10.1073/pnas.0914916107

van Kempen, J., Gieselmann, M. A., Boyd, M., Steinmetz, N. A., Moore, T., Engel, T. A., & Thiele, A. (2021). Top-down coordination of local cortical state during selective attention. *Neuron, 109*(5), 894-904 e898. doi:10.1016/j.neuron.2020.12.013

Victor, J. D., Chubb, C., & Conte, M. M. (2005). Interaction of luminance and higher-order statistics in texture discrimination. *Vision Res, 45*(3), 311-328.

Victor, J. D., & Conte, M. M. (2012). Local image statistics: maximum-entropy constructions and perceptual salience. *J Opt Soc Am A Opt Image Sci Vis, 29*(7), 1313-1345. doi:10.1364/JOSAA.29.001313

Victor, J. D., Conte, M. M., & Chubb, C. F. (2017). Textures as Probes of Visual Processing. *Annu Rev Vis Sci, 3*, 275-296. doi:10.1146/annurev-vision-102016-061316

Victor, J. D., Rizvi, S. M., & Conte, M. M. (2017). Two representations of a high-dimensional perceptual space. *Vision Res, 137*, 1-23. doi:10.1016/j.visres.2017.05.003

Victor, J. D., Thengone, D. J., & Conte, M. M. (2013). Perception of second- and third-order orientation signals and their interactions. *J Vis, 13*(4), 21. doi:10.1167/13.4.21

Victor, J. D., Thengone, D. J., Rizvi, S. M., & Conte, M. M. (2015). A perceptual space of local image statistics. *Vision Res, 117*, 117-135. doi:10.1016/j.visres.2015.05.018

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition, 119*(3), 325-340. doi:10.1016/j.cognition.2011.01.009

Wilson, H. R. (1993). Nonlinear processes in visual pattern discrimination. *Proc Natl Acad Sci U S A, 90*(21), 9785-9790.

Yu, Y., Schmid, A. M., & Victor, J. D. (2015). Visual processing of informative multipoint correlations arises primarily in V2. *Elife, 4*. doi:10.7554/eLife.06604

Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area v2. *Neuron, 47*(1), 143-153. doi:10.1016/j.neuron.2005.04.005

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J Neurosci, 20*(17), 6594-6611.