Perception of second- and third-order orientation signals and their interactions

Jonathan D. Victor

Department of Neurology and Neuroscience,
Weill Medical College of Cornell University,
New York, NY, USA

Daniel J. Thengone

Department of Neurology and Neuroscience,
Weill Medical College of Cornell University,
New York, NY, USA

Mary M. Conte

Department of Neurology and Neuroscience,
Weill Medical College of Cornell University,
New York, NY, USA

Orientation signals, which are crucial to many aspects of visual function, are more complex and varied in the natural world than in the stimuli typically used for laboratory investigation. Gratings and lines have a single orientation, but in natural stimuli, local features have multiple orientations, and multiple orientations can occur even at the same location. Moreover, orientation cues can arise not only from pairwise spatial correlations, but from higher-order ones as well. To investigate these orientation cues and how they interact, we examined segmentation performance for visual textures in which the strengths of different kinds of orientation cues were varied independently, while controlling potential confounds such as differences in luminance statistics. Second-order cues (the kind present in gratings) at different orientations are largely processed independently: There is no cancellation of positive and negative signals at orientations that differ by 45°. Third-order orientation cues are readily detected and interact only minimally with second-order cues. However, they combine across orientations in a different way: Positive and negative signals largely cancel if the orientations differ by 90°. Two additional elements are superimposed on this picture. First, corners play a special role. When second-order orientation cues combine to produce corners, they provide a stronger signal for texture segregation than can be accounted for by their individual effects. Second, while the object versus background distinction does not influence processing of second-order orientation cues, this distinction influences the processing of third-order orientation cues.

Introduction

Functional and physiological considerations indicate the importance of orientation signals for making sense of the visual world. At a functional level, orientation of edges and luminance gradients are primary ingredients of both shape and texture, and thus, orientation signals are critical for segregating a scene into its component objects and for analyzing the surface properties of these components. At a physiological level, selective tuning to orientation is a neurophysiological property that strikingly distinguishes cortical neurons from precortical ones.

Despite the central role played by orientation signals, there is a large gap between the richness of orientation signals that occur in natural scenes and the way that orientation processing is typically studied experimentally. In the laboratory, orientation selectivity is often studied with stimuli that target a single orientation, either over a wide area of the visual field (sine gratings, line gratings) or locally (e.g., Gabor patches). But natural scenes have components with features at multiple orientations, not just across the image but even at a single location (for example, due to transparency or the presence of multiple scales). Physiology indicates that visual cortex has the potential to capture this richness: Neurons with the entire range of orientation tunings are present at every location, so the pattern of activity in a single hypercolumn can represent multiple orientations. It is unclear as yet to

what extent this population activity pattern is "read out" by later processing or is available to perception.

These considerations suggest the need for a broad approach to the analysis of how orientation signals are processed at the perceptual level, which recognizes that multiple orientation signals may be present simultaneously and that they may interact. Perhaps the most straightforward way to study combinations of orientation signals is to use superpositions of simple stimuli, such as Gabors (Sagi, 1988) or gratings (Movshon, Adelson, Gizzi, & Newsome, 1985). With this approach, interactions can be studied by varying the relative orientations of the components. However, a potential confound arises: As the relative orientation of the components of a plaid changes, other aspects of the image change as well. For example, there are changes in the sizes of the patches of brightness and darkness that arise at the reinforcing hills and valleys, there are changes in the luminance histogram, and there are "beats"—periodicities at low spatial frequencies. So it is unclear how to separate interactions of orientation signals from these other image attributes.

Micropattern arrays, consisting of Gabors, dot clusters, or line segments whose orientations vary across space, are another way to probe aspects of orientation processing that cannot be accessed by the use of individual gratings and Gabor patches. Via experiments based on these arrays, it is possible to identify features that support effortless texture segmentation (Caelli & Julesz, 1978; Caelli, Julesz, & Gilbert, 1978; Julesz, 1981), to analyze how spatial contrasts in orientation are processed (Wolfson & Landy, 1998), to measure sensitivity to curvature (Ben-Shahar & Zucker, 2004), to determine the relationship of processing of orientation signals and contrast polarity (Motoyoshi & Kingdom, 2007), and to define the role of co-circularity and of small and large orientation differences (Motoyoshi & Kingdom, 2010).

Contrast-modulated gratings and noises (Baker & Mareschal, 2001; Landy & Henry, 2007; Landy & Oruc, 2002; Larsson, Landy, & Heeger, 2006; Lu & Sperling, 1995; Motoyoshi & Kingdom, 2003; Schofield, Rock, Sun, Jiang, & Georgeson, 2010) represent yet another class of stimuli with complex orientation characteristics. They have been used, for example, (Motoyoshi & Kingdom, 2003) to define the orientation bandwidth and opponent nature of orientation channels. However, all of these approaches have an intrinsic limitation: They only probe one basic kind of orientation signal. For lines, gratings, and Gabors, orientation can be determined from the spatial frequency content of their components (i.e., its power spectrum). For contrast-modulated noises and related stimuli, orientation cannot be determined from the power spectrum of the raw stimulus, but it can be determined from the power spectrum of a stimulus that has been transformed by the action of a simple local nonlinearity. But there are yet other kinds of orientation signals: Orientation information can also be carried by correlations of order three and higher. As we describe below, these may be present even when the above cues are absent. The implication of this observation is that the orientation signals that are present in one region of an image can be quite complex: They include not only a distribution of the orientation cues that can be understood in terms of local spatial frequency content, but of others as well.

The origin of third-order orientation cues is straightforward. A third-order correlation is calculated from the average product of the intensities of an image at triplets of points. As this configuration of points is rotated, the value of this product changes. For example, the intensities among three points arranged in a ⌐-shaped region may be correlated, while the intensities among points arranged in a ⌐-shaped region are independent. Put another way, the value of the third-order correlation depends on the orientation of the "aperture" through which it is calculated. Because of this dependence, it follows that third-order correlations can carry information about orientation.

Before proceeding further, it may be helpful to clarify the "order" terminology used here. In the sense used here, the "order" of an orientation cue refers to the minimum number of points in the image that need to be simultaneously inspected to determine the orientation. While this usage is mathematically natural, it differs from the traditional usage in vision research with regard to orientation and motion. In these domains, the term "first-order" refers to cues (such as gratings or lines) that can be extracted by a simple linear filter, and the term "second-order" refers to cues that require some form of local nonlinear preprocessing (such as filter and rectify), prior to extraction of orientation or motion by a second linear filter (Baker & Mareschal, 2001; Landy & Henry, 2007; Landy & Oruc, 2002; Larsson et al., 2006; Lu & Sperling, 1995; Motoyoshi & Kingdom, 2003; Schofield et al., 2010). But in the sense used here, gratings and lines are said to contain "second-order" cues, since a pair of points needs to be inspected to determine their orientation. Visual stimuli traditionally described as "second-order" (i.e., the ones in the preceding references) contain *fourth*-order cues in the sense used here, since a pair of points need to be inspected to extract the relevant local feature (such as contrast), and two features—and thus four points—need to be analyzed at the later filter stage. The shift in terminology is required because orientation can also be carried by correlations among *triplets* of points—third-order in the sense used here—and the traditional terminology has no obvious parallel.

Informal evidence that these third-order orientation cues are actually used can be gained by visual inspection of images that contain them. Examples of such images are shown in Figure 1 and Figure 4 (below).

In each case, three image intensities are strongly correlated when the aperture consists of an ⌐-shaped arrangement of checks, but only when the aperture is in a particular orientation (i.e., ⌐ but not ⌐, ⌐, or ⌐). Correspondingly, the images all have prominent regions in the shape of right triangles, all of which share the same orientation. This shared orientation is apparent on visual inspection of the images. Less obvious, but readily verified, is that these images do not contain any pairwise correlations (for details, see Victor & Conte, 2012)—so the entire visual impression of an

oriented texture is conveyed by third-order (and possibly higher-order) orientation cues. The basic reason that pairwise correlations are absent is the rule for coloring the checks within the ⌐-shaped template: The parity of the number of white checks is biased to be either typically even, or typically odd. Since the coloring rule is based on the parity of all three checks as a group, individual pairs of checks can be colored independently. As an example, at the highest level of correlation in a "white triangle" texture (Figure 1), all ⌐-shaped regions contain either three white checks or one white check and two black checks. The four possible configurations occur with equal probability: three white checks or a single white check in any of the three positions along with black checks in the other two positions. As a consequence, pairs of checks match and mismatch with equal frequency, so there are no second-order correlations. Additionally, there is no difference in the frequency of white and black checks.

The orientation cues present in these images are closely related to specific kinds of "non-Fourier" motion cues: For motion cues, the regions of correlation are slanted in a space-time plane; here, the regions of correlation are slanted in a plane with two spatial coordinates. In both cases, the designation "non-Fourier" is applicable, since the slanted regions of correlation involve more than two points and are therefore not captured by the power spectrum. (One of the "glider" non-Fourier motion stimuli of Hu and Victor [2010] is the space-time version of the third-order orientation stimuli described here.) Many illusory contours—such as the contour defined by abutting gratings of different orientations—can also be considered fourth-order orientation cues, since the illusory contour is defined by a relationship between a pair of regions, each of which is in turn defined by second-order statistics. As is the case with second-order orientation cues, higher-order orientation cues can also be present in combination with each other (examples in Figure 4 and Figure 7, below); they can also be present in combination with second-order cues (Figure 5, below).

Thus, orientation cues can be quite complex: They can be carried by pairwise correlations (as in gratings), or by higher-order correlations, and cues of either kind can be present alone or in combination.

The approach taken here attempts to address at least a portion of this complexity. To do this, we make use of a set of "maximum-entropy" stimuli (detailed in Victor & Conte, 2012) consisting of visual textures specified by their local correlations. This stimulus set has 10 coordinates; eight of them—described below—control the strength of orientation cues. Four of these coordinates are second-order (the β's, below), and four are third-order (the θ's, below). Each of these cues can be introduced independently or in combination, and
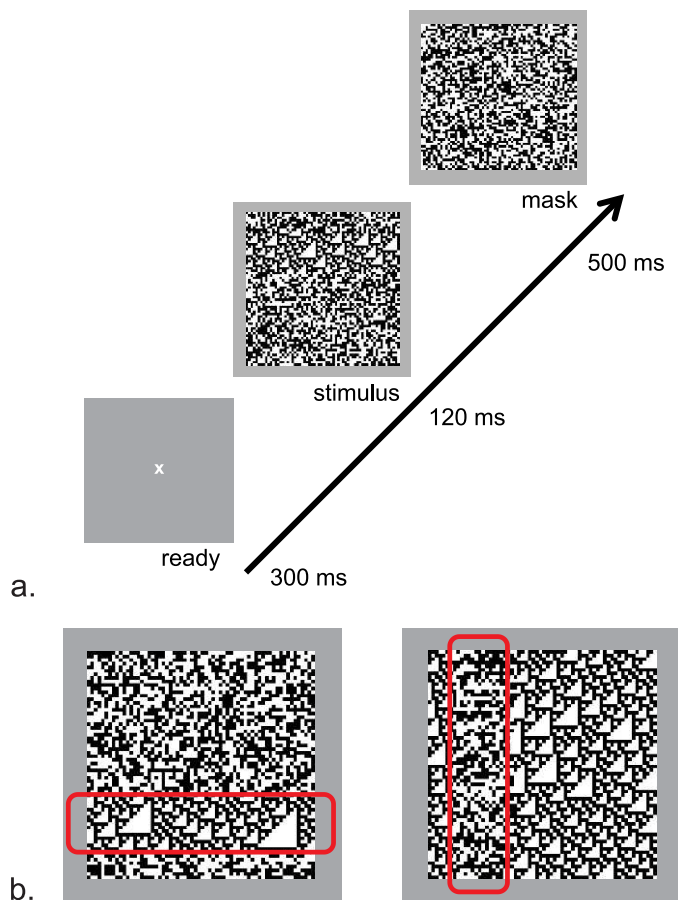

Figure 1. (A) The experimental paradigm. Trials were subject-initiated and began with a fixation spot (300 ms), followed by a stimulus (120 ms), followed by a random mask (500 ms). (B) Example stimuli. Stimuli consisted of a 64 × 64 array of checks, in which a 16 × 64 target region is positioned eight checks from one of the four edges of the array. Stimuli were of two types: *left*, the background is random, and the target is structured (here, $\theta_{\lrcorner} = 1$), or *right*, the background is structured and the target is random. The red rectangles indicate the location of the target and were not visible during the trials.

without changing the distribution of intensities—thus eliminating a key confound.

We use these stimuli in a texture segmentation task to determine human sensitivity to orientation cues individually, how they interact in pairs (both within and across orders), and in more complex combinations. A picture that is relatively simple and highly consistent across subjects emerges.

The "maximum-entropy" property of these textures guarantees that they are as random as possible, other than the spatial correlations that we explicitly specify. This is a key advantage in these experiments, as it allows us to focus on how the different kinds of image statistics interact at the level of processing, rather than at the level of how they might interact via producing alterations in the luminance distribution, or other changes in local statistics. Additionally, since it provides for control over long-range correlations, the maximum-entropy property enables analysis of the extent to which emergent properties play a role in visual performance.

Other classes of textures are also maximum-entropy, but they achieve this goal at the expense of certain limitations that are not shared by the stimuli used here. Specifically, textures based on independent discrete elements (e.g., micropattern textures, or "IID" textures composed of checks whose luminances are independently chosen from the same distribution [Chubb, Landy, & Econopouly, 2004]) achieve maximum entropy. While they can be used to probe sensitivities to statistics beyond second order, the method of construction requires that adjacent elements are uncorrelated. This restricts the kinds of image statistics that can be explored to those that depend on single elements. Conversely, filtered Gaussian noises achieve maximum entropy and include correlations across space, but the correlations are limited to second order. In contrast, the texture stimuli used here are maximum-entropy stimuli that combine both attributes: They include correlations beyond second order and also that extend across space. This enables them to serve as probes of a wide variety of statistics that might provide orientation cues.

The advantages of the present approach do not come without cost. The main limitation is that the checker-board lattice plays an essential role in defining the correlations and constructing the stimuli. Because of this, these stimuli cannot be used to probe interactions among orientations that are close, and cannot be used to make inferences about perceptual differences between cardinal orientations and oblique ones—since the lattice itself leads to differences in these cues. As a good example of the complementary advantages of stimuli based on Gabor patches, see Motoyoshi and Kingdom (2010), who used this approach to determine the interactions of orientation signals ("second-order"

in the terminology used here) across a wide range of large and small orientation differences.

Finally, it is worth emphasizing that although the psychophysical judgments of course are based on individual images, the viewpoint taken is a statistical one. That is, the images are considered to be representatives of an *ensemble*, and the statistical properties described above are rigorously applicable to the *ensemble*, not to individual images (Victor, 1994). We take the statistical view for several reasons. First, it allows for completely independent manipulation of the coordinates of interest. Rigorous control of correlations is possible at the ensemble level, but not for individual images: Since individual images are finite, the correlations estimated from the individual images will differ from that of the ensemble. Fortunately, for images of the size used, this difference is expected to be minor (Maddess, Nagai, Victor, & Taylor, 2007), so that individual images serve as good surrogates for the ensemble. Second, we are asking the subject to perform a statistical task: Even in principle, the "correct" choice on any given trial is simply the most likely choice— though the statistical evidence available to an ideal observer is quite strong. But most importantly, ensembles defined by characteristic statistics play a critical role in the normal function of the visual system. Intuitively, one recognizes a texture not by identifying a particular exemplar, but by recognizing the class to which it belongs. Experimentally, categorization of visual images into statistical classes is rapid, robust, and highly conserved across subjects (Julesz, Gilbert, & Victor, 1978; Victor & Conte, 1991), and the image statistics that support these classifications are demonstrably among the "tokens" of visual working memory (Victor & Conte, 2004).

Methods

The stimulus space

The overall goal of these experiments is to determine how the visual system processes local image statistics that carry information about orientation, and how these statistics interact. To do this, we draw stimuli from a space of images in which local statistics are individually specified, and long-range statistics are as random as possible given the local specifications. The stimulus space is described by 10 coordinates (corresponding to the number of independent local image statistics for a 2×2 array of binary checks); of these, eight are potential carriers of orientation information. We focus on these, but for the reader's convenience we also summarize the relevant aspects of the full stimulus space. For a complete description of the space and the

construction of stimuli within it, see Victor and Conte (2012).

All images consist of binary (black and white) images on a checkerboard lattice. The local statistics used to specify a stimulus consist of the probabilities of all 2×2 blocks, i.e., the frequency with which each way of coloring a 2×2 "windowpane" of checks occurs in the stimulus. Although there are $16 = 2^{2\times2}$ kinds of blocks, there are only 10 free parameters. The reduction in the number of degrees of freedom occurs because the blocks can be placed in overlapping fashion, and they must match where they overlap. Linear combinations of these 16 probabilities provide 10 independent coordinates, which, together, fully specify the stimulus space. The coordinates fall into categories based on their order, i.e., the number of checks that must be simultaneously inspected to determine their values. As detailed below, there is one first-order coordinate (denoted γ), four second-order coordinates (denoted β_-, $\beta_|$, $\beta_\backslash$, and $\beta_/$), four third-order coordinates (denoted $\theta_\lrcorner$, $\theta_\llcorner$, $\theta_\ulcorner$, and $\theta_\urcorner$), and one fourth-order coordinate (denoted α). All coordinates range from -1 to 1, and a completely random binary image corresponds to all 10 coordinates having the value 0. Our focus is on the four β's and the four θ's, since they are the potential carriers of orientation information. To describe these and the two nonoriented coordinates γ and α in detail, we use the convention that white checks are denoted by 1, and black checks are denoted by 0.

The overall luminance bias of the image is captured by γ: It is the difference between the probability of a white check and the probability of a black check. If $\gamma = 1$, all checks are white; if $\gamma = -1$, all checks are black; and if $\gamma = 0$, both colors are equally likely.

The β's capture the pairwise (second-order) statistics: They are the difference between the probability that two neighboring checks match (i.e., both are white or both are black), and the probability that they do not match (i.e., one is white and one is black). If $\beta = 1$, all checks match their nearest neighbor (in the direction indicated by the subscript), and if $\beta = -1$, they all mismatch. The four subscripts (β_-, $\beta_|$, $\beta_\backslash$, and $\beta_/$) correspond to the direction that is relevant to the match. For example, $\beta_- = 1$ means that all 1×2 blocks are either (0 0) or (1 1) and none are (0 1) or (1 0); such images will be dominated by horizontal stripes. If $\beta_- = -1$, all 1×2 blocks are either (0 1) or (1 0) and none are (0 0) or (1 1); in such images, horizontal rows will have alternating black and white checks. Values of β_- between 0 and 1 indicate a partial bias towards matching neighbors, while values between -1 and 0 indicate a partial bias towards mismatching neighbors, and $\beta_- = 0$ means that matching and mismatching neighbors are equally likely. Similarly, $\beta_|$, $\beta_\backslash$, and $\beta_/$ capture the pairwise correlations in the vertical

direction and the two oblique directions. Image patches with $\beta = \pm0.4$ are shown in Figure 2, upper panels.

The θ's capture the statistics of triplets of checks arranged in an $\llcorner$-shaped configuration. Since there are four possible orientations for an $\llcorner$-shaped configuration within a 2×2 windowpane, there are four θ-statistics, $\theta_\lrcorner$, $\theta_\llcorner$, $\theta_\ulcorner$, and $\theta_\urcorner$. Each of them measures the third-order correlation within the corresponding $\llcorner$-shaped region by comparing the probability that the region contains an even number of white checks versus an odd number of white checks. If $\theta = 1$, only an odd number of white checks (one or three) are present; $\theta = -1$ means the opposite. For example, $\theta_\lrcorner$ means that only the configurations $\begin{pmatrix} & 1 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} & 0 \\ 1 & 0 \end{pmatrix}$, or $\begin{pmatrix} & 0 \\ 0 & 1 \end{pmatrix}$ are present (the fourth element of the 2×2 region is unconstrained); such images will have prominent white triangular-shaped regions pointing downward and to the right. If $\theta_\lrcorner = -1$, only the configurations $\begin{pmatrix} & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} & 0 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} & 1 \\ 0 & 1 \end{pmatrix}$, or $\begin{pmatrix} & 1 \\ 1 & 0 \end{pmatrix}$ are present; such images will have prominent black triangular-shaped regions. Image patches with $\theta = \pm0.72$ (for $\theta_\lrcorner$, $\theta_\llcorner$, and $\theta_\ulcorner$) are shown in Figure 2, lower panels.

The final coordinate, α, captures the statistics of quadruplets of checks in a 2×2 block: $\alpha = 1$ means that an even number of them are white, and $\alpha = -1$ means that an odd number are white. This gamut has been studied extensively (Julesz et al., 1978; Victor, Chubb, & Conte, 2005; Victor & Conte, 1989, 1991, 1996, 2004), and is not our focus here.

In sum, of the 10 coordinates $\{\gamma, \beta_-, \beta_|, \beta_\backslash, \beta_/, \theta_\lrcorner, \theta_\llcorner, \theta_\ulcorner, \theta_\urcorner, \alpha\}$, eight of them carry orientation information. They consist of the four second-order statistics $\{\beta_-, \beta_|, \beta_\backslash, \beta_/\}$ and the four third-order statistics $\{\theta_\lrcorner, \theta_\llcorner, \bar{\theta}_\ulcorner, \theta_\urcorner\}$.

Stimuli

To determine psychophysical sensitivity to image statistics and their combinations, we used the texture segmentation paradigm (Figure 1) first developed by Chubb and coworkers for the study of textures in which each check's luminance is independently chosen from the same distribution (Chubb et al., 2004), and later adopted for correlated textures (Victor et al., 2005; Victor & Conte, 2012). The basic stimulus consisted of a 64×64 array of checks, in which a target region (a 16×64 rectangle) was positioned eight checks from one of the four edges of the array. This target region was distinguished from the remainder of the array by its statistics. To ensure that the subject responded on
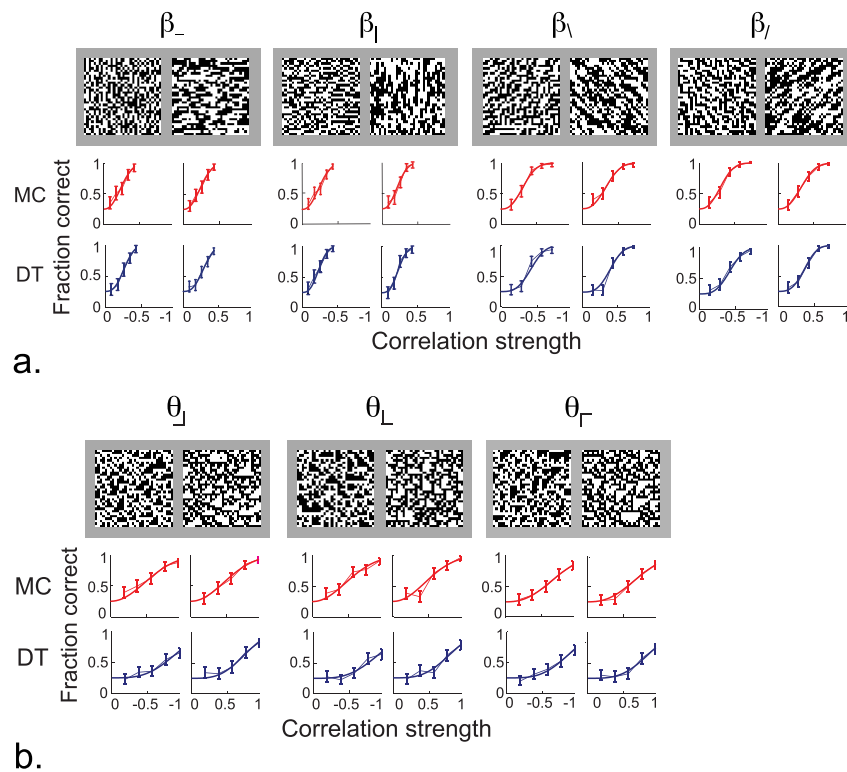
Figure 2. Psychometric functions for individual image statistics that carry orientation information. (A) Second-order correlations in cardinal (β_-, $\beta_|$) and oblique directions ($\beta_\backslash$, $\beta_/$) directions. For each statistic, psychometric functions are shown for negative excursions (left element of each pair) and positive excursions (right element of each pair). Chance performance is 0.25. The patches above the psychometric functions show typical 32×32 samples of images defined by the corresponding second-order statistic, constructed with $\beta = \pm 0.4$. For second-order statistics, positive values correspond to correlation in one direction; negative values correspond to anticorrelation in that direction. (B) Third-order correlations in each of three directions ($\theta_\lrcorner$, $\theta_\llcorner$, $\theta_\ulcorner$). Psychometric functions as in (A). The image samples are constructed with $\theta = \pm 0.72$. For third-order statistics, positive values correspond to an excess of white triangular patches in a particular orientation; negative values correspond to an excess of dark triangular patches in the same orientation. Data from two representative subjects, MC and DT, out of $N = 6$.

the basis of segmentation (rather than, say, a texture gradient), we randomly intermixed trials of two types: (a) trials in which the background was random, and the target had a nonzero value of one or more image statistic, and (b) trials in which the background had the nonzero values, and the target was random (see Figure 1B).

We explored the stimulus space in a radial fashion, i.e., by choosing points along rays in different directions. Three kinds of directions were used: (a) Along a coordinate axis: A single coordinate (one of the β's or one of the θ's) was set at a nonzero value, either positive or negative. In each direction, four or five equally-spaced values were chosen to span the range from below threshold to well above threshold based on pilot experiments. For β_- and $\beta_|$, the maximum (absolute value) was 0.45, for $\beta_\backslash$ and $\beta_/$, the maximum was 0.75, and for the θs, the maximum was 1.0. (b) In a coordinate plane: A pair of coordinates was set at a nonzero value. This was done in all

quadrants (i.e., in all sign combinations: both coordinates positive, both negative, and coordinates that were opposite in sign). The ratio of the coordinate values was fixed and chosen in approximate proportion to the above maximum values. Two values along each direction were studied. (c) Combinations of four coordinates of the same order (all four β's or all four θ's). All four coordinates had the same absolute value, and their signs were chosen either to match or to alternate as a function of orientation (see Figure 7). Four equally-spaced values were chosen, 0.075, 0.125, 0.175, and 0.225. (0.225 is 90% of the maximum possible value.) As described in detail in Victor and Conte (2012; see its table 2), the unspecified coordinates were assigned by first setting the values of all lower-order coordinates to zero, and then setting the remaining coordinates to values that maximized the entropy of the resulting images. In most cases, these other coordinate values were zero; in the cases in which the value was nonzero, it was below the

perceptual threshold. For example, for a $(\beta_-, \beta_|)$ combination, the maximum-entropy value of α is approximately $\beta_-^2 + \beta_|^2$. The thresholds are <0.2 for this combination of β's, and the corresponding $\alpha = 0.08$ is far below its threshold, which is >~0.5. Full details for the construction of the on-axis and coordinate-plane stimulus are provided in (Victor & Conte, 2012).

For stimuli specified by four nonzero coordinates, we used the "donut algorithm" of Victor and Conte (2012) to mix two in-plane textures. For the β's, we mixed a stimulus specified by nonzero values of β_- and $\beta_\backslash$ with a stimulus specified by nonzero values of $\bar{\beta}_|$ and $\beta_/$. For the θ's, we mixed a stimulus specified by nonzero values of $\theta_\lrcorner$ and $\theta_\llcorner$ with one specified by $\theta_\vdash$ and $\theta_\neg$ (when θ's had the same signs), and we mixed a stimulus specified by $\theta_\lrcorner$ and $\theta_\vdash$ with one specified by $\theta_\llcorner$ and $\theta_\neg$ (when θ's had the alternating signs). These strategies ensured that the values of the unspecified coordinates were exactly zero, with the exception that for stimuli constructed as a mixture of four β's, the value of α could be as high as 0.36 (for all β's set at 0.225, the largest value used). Since this value was not negligible, we assessed its effect on threshold in two subjects (MC and DT), by determining sensitivities to α and its pairwise interactions with the β's.

Stimuli were presented on a mean-gray background, followed by a random mask. The display size was $15° \times 15°$ (check size, 14 min), contrast was 1.0, and viewing distance was 1 m. Studies were carried out on an LCD monitor with a mean luminance of 23 cd/m^2, a refresh rate of 100 Hz, and a presentation duration of 120 ms, driven by a Cambridge Research ViSaGe system (Cambridge Research Systems, Ltd., Rochester, Kent, UK).

Subjects

Studies were conducted in seven normal subjects (three male, four female), ages 21 to 54. Six subjects (MC, DT, JD, DF, KP, and TT) participated in extensive experiments to assess on-axis sensitivities and pairwise combinations. Four of these subjects (MC, DT, JD, and DF) viewed all pairwise combinations; KP and TT did not view stimuli consisting of combinations containing one β and one θ. Three of the seven subjects (MC, DT, and DC) participated in experiments to assess four-component combinations. On-axis sensitivities from DC were also measured, but to a much more limited extent. Of the seven subjects, MC is an experienced psychophysical observer, DC had no observing experience prior to the current study, and the other subjects had modest viewing experience (10 to 100 hours). All subjects other than MC and DT were naive to the purposes of the experiment. All subjects had visual acuities (corrected if necessary) of 20/20 or better.

Procedure

The subject's task was to identify the position of the target, in a four-alternative forced choice (4-AFC) texture segregation task (Figure 1A). Subjects were told that the target was equally likely to appear in any of four locations (top, right, bottom, or left), and they were shown examples of stimuli of both types: target structured/background random and target random/background structured. They were instructed to maintain central fixation, rather than to attempt to scan the stimulus. Auditory feedback for incorrect responses was given during training trials but not during data collection. After performance stabilized (approximately 2 hr for a new subject), blocks of trials (with trials presented in randomized order) were presented. Block order was counterbalanced across subjects.

Experiments were organized into two kinds of blocks. In the first kind of block, stimuli were presented in the positive and negative directions along two axes (four directions, five strengths), and in oblique directions in the plane that these axes determined (four directions, two strengths). Each coordinate value was used for an equal number of stimuli in each target location, and an equal number of the two stimulus types (target structured/background random or target random/background structured, see Figure 1B). This resulted in 288 trials per block (160 on-axis stimuli, 128 pairwise combinations). We collected 15 such blocks per subject (4,320 trials), grouped into three experimental sessions, yielding 120 to 240 responses for each set of coordinate values.

In the second, kind of block, on-axis, and four-component stimuli were presented. On-axis stimuli were presented along three $(\beta_-, \beta_|, \beta_\backslash)$ or two $(\theta_\lrcorner, \theta_\llcorner)$ axes (four positive and four negative strengths), and in each of three combination directions (four strengths each). This resulted in 192 (β's) or 128 (θ's) on-axis stimuli and 96 combination stimuli, resulting in 288 (β's) or 224 (θ's) trials per block. We collected 15 such blocks per subject (4,320 β's or 3,360 θ's trials), grouped into five experimental sessions, yielding 120 responses for each set of coordinate values.

Analysis

Determination of thresholds from psychophysical data

As in Victor et al. (2005), measured values of the fraction correct (FC) are fit to Weibull functions via maximum likelihood:

$$FC(x) = \frac{1}{4} + \frac{3}{4}\left(1 - 2^{-(x/a_r)^{b_r}}\right) \tag{1}$$

For the analysis of the blocks consisting of on-axis and pairwise combinations, this fitting procedure is initially carried out separately along each ray r. For rays along the coordinate axes, x is the coordinate value; for the rays in the oblique directions, x is the distance from the origin. In most cases, the Weibull shape parameter (the exponent b_r) was in the range 2.2 to 2.7 for each ray, or had confidence limits that included this range. Therefore, we fit the entire dataset in each coordinate plane by a set of Weibull functions constrained to share a common exponent b, but allowing the position parameter a_r to vary across rays. (Note, however, that the exponent b was allowed to vary *between* planes; see Table 1, below). Next a_r was taken as a measure of threshold, as $x = a_r$ yields performance halfway between floor and ceiling (here, FC = 0.625). The 95%-confidence limits for a_r were determined via 1,000-sample bootstraps. When performance was sufficiently close to chance for the entire ray, the upper confidence limit of these bootstraps was large (e.g., $>10^5$); in these cases, the threshold was taken to be infinity. Unless otherwise noted, averages across subjects were calculated as harmonic means. (We used harmonic means to avoid divergences that would have resulted from averaging immeasurably large thresholds. The harmonic mean of thresholds is equivalent to the arithmetic mean of the sensitivities.)

For the analysis of the blocks consisting of on-axis and four-component combinations, a similar procedure was used: Each dataset was fit by a set of Weibull functions constrained to share a common exponent b, but allowing for different values of a_r for each combination. Here, for the combination directions, we used the convention that x in Equation 1 is the common coordinate value of each component, since this facilitates the key comparisons below. (This convention is not the same as expressing thresholds in terms of distance from the origin. Since all four coordinates had the same absolute value, the Euclidean distance of a typical point ($\pm x, \pm x, \pm x, \pm x$) from the origin is $2x$. Thus, to convert thresholds expressed as single coordinate values into thresholds expressed as distance from the origin, the numerical value should be doubled.)

Modeling of four-component thresholds

To compare the thresholds obtained with four-component stimuli with the thresholds obtained from the on-axis (one-component) and in-plane (two-component) stimuli, we used a simple descriptive model: Individual image statistics are combined to form a decision variable, and when the decision variable reaches threshold, the subject is able to perform the segmentation task.

For the combination rule that takes the individual statistics to the decision variable, we used a general quadratic form, $\sum_{i,j} Q_{i,j} c_i c_j$, where c_i represents an individual image statistic (for mixtures of second-order statistics, $c_1 = \beta_-$, $c_2 = \beta_|$, $c_3 = \beta_\backslash$, and $c_4 = \beta_/$; for mixtures of third-order statistics, $c_1 = \theta_\lrcorner$, $c_2 = \theta_\llcorner$, $c_3 = \theta_\ulcorner$, and $c_4 = \theta_\urcorner$) and the quantities $Q_{i,j}$ to describe how the image statistics c_i and c_j combine and interact. Without loss of generality, the threshold is set to 1 (since alternative values could be absorbed into the $Q_{i,j}$). Thus, the model states that threshold is reached when

$$\sum_{i,j} Q_{i,j} c_i c_j = 1. \tag{2}$$

This is the equation of a generic ellipsoid, whose shape is determined by the parameters $Q_{i,j}$.

We chose a quadratic combination rule because it has proven effective for cue combination in other settings (Macadam, 1942; Poirson, Wandell, Varner, & Brainard, 1990; Saarela & Landy, 2012) and because its intersection with any coordinate plane is an ellipse, generally consistent with the isodiscrimination contours measured experimentally (see Figures 3, 4, and 5, below). (We do not intend to rule out the possibility that the combination rule could also be characterized by a higher exponent [Quick, 1974; Shephard, 1964; To, Baddeley, Troscianko, & Tolhurst, 2011], perhaps with slightly greater accuracy.) As has been previously noted (Poirson et al., 1990), Equation 2 has two interpretations: individual channels with sensitivity to more than one image statistic or channels with exclusive sensitivity to one image statistic that interact at a later stage. Since both interpretations have identical mathematical formulations, we do not attempt to distinguish between them.

For individual subjects, we determined the values of the parameters $Q_{i,j}$ from the thresholds T_r measured along all rays r (including the on-axis and in-plane rays). That is, we adjusted the $Q_{i,j}$ so that along each ray r, $\sum_{i,j} Q_{i,j} c_i(T_r) c_j(T_r)$ was as close as possible to 1, where $c_i(T_r)$ is the value of the texture coordinate i when threshold is reached in direction r. The adjustment of the $Q_{i,j}$ was accomplished by minimizing

$$F = \sum_r \left(\left(\sum_{i,j} Q_{i,j} c_i(T_r) c_j(T_r)\right) - 1\right)^2, \tag{3}$$

which is a linear least-squares fitting procedure for the $Q_{i,j}$. Note that $F = 0$ only if the threshold T_r in each direction r is exactly predicted by Equation 2, namely, $\sum_{i,j} Q_{i,j} c_i(T_r) c_j(T_r) = 1$. Once the parameters $Q_{i,j}$ are determined by minimizing Equation 3, Equation 2 provides a prediction of thresholds along any ray,

including rays in directions that correspond to the four-component mixtures: it is the value T_{pred} for which

$$\sum_{i,j} Q_{i,j} c_i(T_{pred}) c_j(T_{pred}) = 1. \qquad (4)$$

The above procedure was modified slightly for the second-order statistics because, as mentioned above, these stimuli also contain fourth-order correlations. (The fourth-order correlations are described by the parameter α, which indicates the fraction of 2×2 blocks that contain an even number of white checks. Nonzero values of α arise because the correlations among each pair of checks in a 2×2 block induces correlations among the quadruple of checks. For example, if all nearest-neighbor pairs tend to match, then the number of white checks within a 2×2 block is more likely to be even. For further details, see Victor and Conte [2012]). To take the fourth-order correlations into account, we include a fifth coordinate $c_5 = \alpha$ in Equation 2 and Equation 3, along with the four coordinates corresponding to the second-order statistics ($c_1 = \beta_{\_}$, $c_2 = \beta_{|}$, $c_3 = \beta_{\backslash}$, and $c_4 = \beta_{/}$). In two of the subjects (MC and DT) in which we measured responses to the four-component mixtures, we also determined the isodiscrimination contours in the planes spanned by α and the βs (Victor & Conte, 2012), and we used those results here to fit the five-coordinate version of Equation 3, and to predict thresholds via Equation 4. The remaining subject (DC) only participated in experiments involving the individual image statistics and the four-component mixtures. For this subject, we determined the model parameters by using his measured thresholds along the coordinate axes and the orientations of the ellipses obtained from MC or DT. This amounts to stretching the best-fitting ellipses from subject MC or DT along each coordinate axis to match DC's single-statistic thresholds (i.e., rescaling each $Q_{i,j}$ by a factor g_i), and rescaling the interaction terms $Q_{i,j}$ by the geometric mean of these scaling factors $\sqrt{g_i g_j}$.

Complete pooling and independent processing

The model described above contains two limiting cases, which we denote by "complete pooling" and "independent processing." For the complete pooling case, image statistics are combined by simple summation within a single channel, so the decision variable is $\sum_k q_k c_k$ (where q_k is the sensitivity of the channel to the statistic c_k). The above model reduces to this case when the off-diagonal parameters $Q_{i,j}$ are chosen according to $Q_{i,j} = \sqrt{Q_{i,i} Q_{j,j}}$. In this case, $\sum_{i,j} Q_{i,j} c_i c_j = (\sum_k q_k c_k)^2$ and Equation 2 is equivalent to

$$\left| \sum_k q_k c_k \right| = 1. \qquad (5)$$

For the "independent processing" case, interactions between image statistics are assumed to be 0; this corresponds to setting $Q_{i,j} = 0$ for $i \neq j$ in the above, so Equation 2 becomes

$$\sum_k Q_{k,k} c_k^2 = 1. \qquad (6)$$

The parameter values q_k and $Q_{k,k}$ for the above models were determined from the on-axis thresholds.

Results

We analyzed human sensitivity to orientation signals via a 4-AFC segmentation task, in which the difference between the target texture and that of the background was determined by one or more image statistics. Four of the image statistics were second-order ($\{\beta_{\_}, \beta_{|}, \beta_{\backslash}, \beta_{/}\}$, which captured pairwise correlations in each of four orientations); four were third-order $\{\theta_{\lrcorner}, \theta_{\llcorner}, \theta_{\ulcorner}, \theta_{\square}\}$, which captured correlations among three checks, in each of four orientations. We first describe sensitivity to individual statistics, then examine their combinations.

Individual oriented image statistics

Figure 2A shows, for each of the four second-order statistics, psychometric functions that quantify the subjects' ability to segment a target by discriminating various pairwise correlation strengths from random noise. Thresholds were lower for the two cardinal directions ($\beta_{\_}$, $\beta_{|}$) than for the two oblique directions ($\beta_{\backslash}$, $\beta_{/}$), but there was little difference between the two directions within each category or between positive and negative excursions. These findings held across the $N = 6$ subjects. The (geometric) mean threshold for cardinal directions was 0.286 (0.258 to 0.316, 95% confidence limits via t test on log thresholds); for oblique directions it was 0.402 (0.362 to 0.446). This difference was highly significant ($p < 0.001$, two-tailed paired t test on log thresholds). Note that the cardinal and oblique statistics do not differ merely in orientation: $\beta_{\_}$ and $\beta_{|}$ describe the correlation between checks that share an edge, while $\beta_{\backslash}$ and $\beta_{/}$ describe the correlation between checks that share a corner. For either kind of image statistic, thresholds across subjects varied by only 10% (standard deviation from the mean on a log scale).

There were no significant differences between the two cardinal directions or between the two oblique directions (either for positive or negative excursions), or between positive and negative excursions in any direction ($p > 0.05$). There were no systematic differences in performance in the conditions in which

the target was random and the background was structured versus conditions in which the target was structured and the background was random. There also was no consistent difference in performance for any of the four target positions.

Figure 2B shows psychometric functions for the third-order statistics obtained by these subjects. Thresholds for the three examples tested ($\theta_\lrcorner$, $\theta_\llcorner$, $\theta_\ulcorner$) were not significantly different ($p > 0.05$ for pairwise comparisons of any two of the θ's, in either positive or negative excursions); $\theta_\neg$ was only tested in pilot fashion. However, there was a consistent difference between positive excursions (images with white triangular regions) and negative excursions (images with black triangular regions): Positive excursions had a slightly lower threshold ($p \approx 0.05$ for $\theta_\lrcorner$, $p < 0.005$ for $\theta_\llcorner$, $p < 0.1$ for $\theta_\ulcorner$, and $p < 0.01$ when pooled). Across subjects, the threshold for positive excursions was 0.767 (0.663 to 0.887) and for negative excursions, 0.847 (0.721 to 0.995). This corresponds to a variation between subjects of 15% (standard deviation on a log scale).

In sum, subjects are sensitive to individual image statistics that carry second- and third-order orientation information. Across subjects, thresholds were highly consistent, varying by no more than 15% about the mean. Sets of statistics that were equivalent under rotational transformations had indistinguishable thresholds $\{\beta_-, \beta_|\}$, $\{\beta_\backslash, \beta_/\}$, and $\{\theta_\lrcorner, \theta_\llcorner, \theta_\ulcorner\}$. For second-order statistics, thresholds for positive excursions (an increase in the number of neighbor-pairs that match) were indistinguishable from thresholds for negative excursions (an increase in the number of neighbor pairs that do not match). For third-order statistics, thresholds for positive excursions (bright triangles) were about 10% lower than thresholds for negative excursions (dark triangles).

Detection based on emergent features?

Because images with high values of one of the third-order statistics tend to contain large triangular blobs, one might wonder whether the detection of visual structure is based on these emergent features. Furthermore, the observation that thresholds for the third-order statistics are two- to three-fold higher than for second-order statistics might suggest that this is the case. Together, these considerations raise the possibility that detection of visual structure for third-order statistics is not based on the local statistics per se, but rather on some mechanism that detects larger features that emerge when image patches in several local regions interact in a cooperative fashion.

To test this possibility, we first quantify this cooperativity via a simple calculation. As described above, the individual θ-statistic, by definition, indicates the extent to which the texture is enriched with uniform three-check triangular regions (containing two checks on a side). The next-smallest triangular region, a six-check region (three checks on a side), contains within it three overlapping examples of the three-check triangle. Because the six-check triangle contains three instances of the three-check triangle (and therefore, three independent instances in which the θ-bias is applied), the extent to which there is an enrichment of six-check triangles is given by θ^3. Similarly, a 10-check triangle (four checks on a side) contains six instances of the three-check trial, and therefore six independent applications of the θ-bias, and is enriched by θ^6. These accelerating functions quantify the presence of triangles: For values of θ near zero, they are no more frequent than chance; as θ approaches -1 or $+1$, their frequency increases rapidly. Thus, if detection is based on these features, the transition from subthreshold to suprathreshold is expected to be more rapid than if detection is based on a feature whose presence is merely proportional to θ. The rapidity of this transition is quantified by the Weibull exponent in Equation 1: If detection is based on a feature whose presence grows like θ^p rather than θ, then the Weibull exponent (for fraction correct as a function of $x = |\theta|$ in Equation 1) is expected to be p times higher:

$$
\begin{aligned}
FC(\theta) &= \frac{1}{4} + \frac{3}{4}\left(1 - 2^{-(|\theta|^p/a)^b}\right)\\
&= \frac{1}{4} + \frac{3}{4}\left(1 - 2^{-(|\theta|/a')^{bp}}\right)
\end{aligned} \tag{7}
$$

Thus, the Weibull exponent provides an index of co-operativity, and we can use it to determine the extent to which there is evidence for cooperative processing of the third-order statistics.

Results of this analysis are shown in Table 1. We use the Weibull exponents for second-order statistics as a baseline, since it is well-established that local second-order statistics are readily detected (e.g., Graham, 1989; Graham, Beck, & Sutter, 1992). Interestingly, there is subject-to-subject variability in the Weibull exponent, but within subject, the average Weibull exponent for second- and third-order statistics do not differ from each other by more than 30%. Averaged across subjects, the exponents for third-order statistics are higher, but only marginally so (11% higher, $p \approx 0.05$, one-tailed paired t test, $N = 6$). This is in contrast with the several-fold change in the exponent that would be expected if detection of structure was based on cooperative interactions of local statistics.

Pairs of oriented image statistics

The above results show that the statistical structure in textures defined by single second- and third-order

Subject	Second-order statistics				Third-order statistics		
	$(\beta_-, \beta_|)$	$(\beta_-, \beta_\backslash)$	$(\beta_\backslash, \beta_/)$	Geometric mean	$(\theta_\lrcorner, \theta_\llcorner)$	$(\theta_\lrcorner, \theta_\ulcorner)$	Geometric mean
MC	2.35	2.82	2.54	2.56	2.09	2.54	2.31
DT	2.78	2.96	3.14	2.95	3.28	3.49	3.38
JD	3.00	3.65	3.50	3.37	4.55	4.21	4.37
DF	2.83	2.81	2.71	2.78	3.33	3.13	3.23
KP	2.83	3.42	3.23	3.15	3.06	3.27	3.16
TT	2.92	2.89	2.72	2.84	3.42	3.41	3.41
Geometric mean	2.78	3.08	2.95	2.93	3.21	3.30	3.25

Table 1. Comparison of Weibull exponents for second- and third-order statistics. *Notes*: For each subject, the overall Weibull exponents (see Methods) obtained from experiments based on pairs of parameters drawn from a single order are combined (geometric mean) to determine an overall exponent for each order. The Weibull exponents for third-order statistics are slightly (~11%) higher than for second-order statistics ($p \approx 0.05$, one-tailed paired t test, $N = 6$).

image statistics is readily detected, but does not provide insight into whether the detection is mediated by oriented or nonoriented (Motoyoshi, Nishida, Sharan, & Adelson, 2007) mechanisms. In principle, nonoriented mechanisms could detect structure in a texture characterized by the third-order statistics considered here, since when filtered at a scale comparable to the triangular "blobs," these textures yield luminance distributions that are skewed. However, these two possibilities—detection by nonoriented versus oriented mechanisms—yield contrasting predictions for textures defined by *pairs* of statistics. This motivates the analysis below, where we examine how threshold depends on the relative orientation and sign of two image statistics.

To see why these alternatives make contrasting predictions, consider an image characterized by a pair of statistics that differ by orientation and also in sign, such as $\theta_\lrcorner = 0.4$ and $\theta_\llcorner = -0.4$. If these statistics are processed in a completely pooled fashion (as would be the case if the underlying mechanisms had no orientation tuning), then the image would be difficult to distinguish from a random one. This is because the two image statistics have magnitudes that are equal but are opposite in sign. (For further discussion specific to the Motoyoshi et al., [2007] model, see Discussion). On the other hand, if the two image statistics are processed independently (as would be the case if there are multiple mechanisms, each with its own orientation tuning), then the image would be readily discriminable from a random image, since cancellation would not occur. Thus, testing thresholds for a combination of image statistics that differ in orientation and sign will determine if the image statistics are processed in an orientation-specific way: If processing is pooled across orientations, there will be cancellation; if processing is orientation-specific, cancellation need not occur. A useful control for this analysis is the companion image in which the same two statistics have the same sign: This will not result in cancellation even if the statistics are processed in a pooled fashion.

Based on this rationale, we examined segmentation thresholds for image patches defined by a pair of statistics, with specific attention to comparisons between images in which the texture statistics had the same sign versus the opposite sign. We use this to calculate a "pooling index" I_{pool} for any two image statistics, which compares the same-sign thresholds ($h_{+,+}$ and $h_{-,-}$) to the opposite-sign thresholds ($h_{+,-}$ and $h_{-,+}$):

$$I_{pool} = \frac{h_{+,-} + h_{-,+}}{h_{+,+} + h_{-,-}}. \tag{8}$$

$I_{pool} = 1$ means that the combination thresholds are the same, whether the statistics are present with the same sign or opposite sign, indicating independent processing (a lack of cancellation). $I_{pool} > 1$ means that some cancellation has occurred. $I_{pool} = \infty$ means that cancellation is complete (i.e., opposite-sign thresholds are infinite, and any potential orientation information is lost).

Second-order statistics

Results for pairwise combinations of the second-order orientation statistics are shown in Figure 3. For the combination of statistics representing pairwise correlation in the two cardinal directions (β_- and $\beta_|$), the results are particularly simple (Figure 3, first row): Thresholds are nearly identical whether they are combined with opposite sign or with the same sign.

Correspondingly, I_{pool} is very nearly 1 (I_{pool} ranges from 0.97 to 1.21 for individual subjects, $I_{pool} = 1.09$ from harmonic means, $N = 6$, Figure 6, below).

For the pair consisting of one cardinal and one oblique second-order statistic (β_- and $\beta_\backslash$, second row of Figure 3), the individual thresholds are different (as expected from Figure 2A), but again thresholds for the combination are independent of whether the signs are
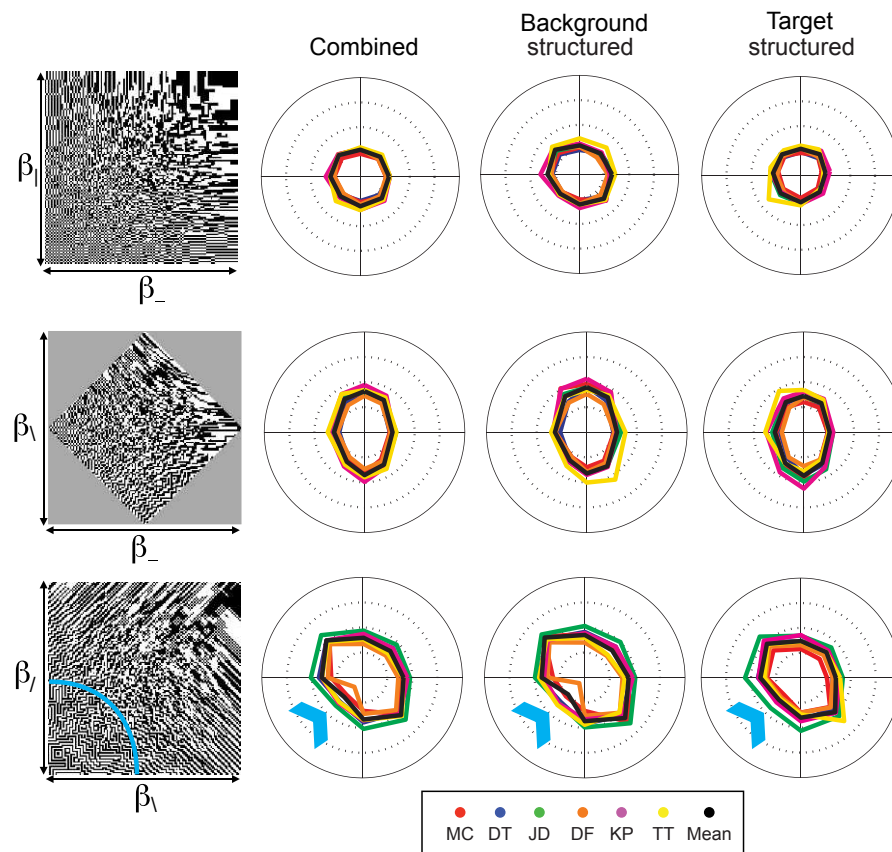
Figure 3. Isodiscrimination contours for pairs of second-order statistics. The image at the left of each row shows the gamut of images that can be constructed with each pair of statistics. The random texture (the origin) is at the center of each gamut; image statistics increase with as distance from the origin grows and the ends of the axes correspond to values of ± 1. The polar plots are isodiscrimination contours, i.e., the location in the gamut at which criterion performance (fraction correct $= 0.625$) was obtained, in each of eight directions (along each axis and along diagonals). Each colored trace corresponds to a separate subject; the black trace is the harmonic mean across subjects. The first column of polar plots was determined from all trials; the second column was determined from the half of the trials in which the background was structured according to the image statistics and the target was a random patch; the third column was obtained from the other half of the trials, in which the target was structured according to the image statistics and the background was a random patch. The isodiscrimination contours are generally circular or elliptical, but there is a consistent flattening for the ($\beta_\backslash$, $\beta_/$)-pair, when both statistics are negative (blue arrow, lower row). This corresponds to the region of the gamut in which corners are present (rounded square, gamut). For all pairs of second-order statistics, there is little dependence on whether the background or the target was structured. $N = 6$.

opposite or the same (I_{pool} ranges 1.03 to 1.11, mean $=$ 1.06, Figure 6, below).

The final pairing of second-order statistics, $\beta_\backslash$ and $\beta_/$, elicits a different behavior. In contrast to the above two cases, the isodiscrimination contours are not aligned with the coordinate axes—instead, they are tilted, with the long axes extending into the quadrants in which $\beta_\backslash$ and $\beta_/$ have opposite signs (Figure 3, third row). The tilt means that thresholds for the combinations with opposite sign are higher than for combinations with the same sign (I_{pool} ranges from 1.34 to 2.10, mean $=$ 1.62, $N = 6$, Figure 6, below).

The isodiscrimination contours for the pairing of $\beta_\backslash$ and $\beta_/$ show a small but consistent deviation from the elliptical shape: Thresholds when both statistics are

negative are lower than when both statistics are positive (blue arrow in Figure 3, bottom row). This deviation was seen, with varying degrees of prominence, in all six subjects. Interestingly, the images defined by $\beta_\backslash < 0$ and $\beta_/ < 0$ have the appearance of a maze, with many corners oriented along the cardinal axes (see region inside the blue arc in Figure 3, bottom row). But even though this quadrant of the space ($\beta_\backslash < 0$ and $\beta_/ < 0$) has "corners," its degree of statistical structure is identical to what is present in the other quadrants of the space (i.e., $\beta_\backslash > 0$ or $\beta_/ > 0$ or both (Victor & Conte, 2012). Thus, the lower thresholds indicate that pairings of second-order statistics that produce corners are processed more efficiently than other second-order pairings.
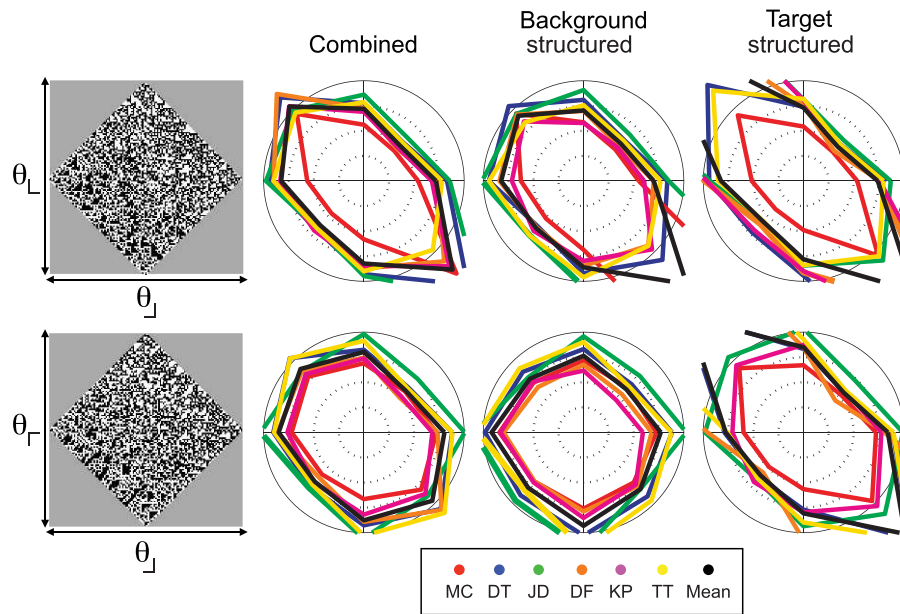
Figure 4. Isodiscrimination contours for pairs of third-order statistics. As in Figure 3, the image at the left of each row shows the gamut of images for each pair of image statistics, and the polar plots are isodiscrimination contours, i.e., the location in the gamut at which criterion performance was reached. For other plotting conventions, see Figure 3. Note that for the pair ($\theta_\sqsubset$, $\theta_\lrcorner$), there is a strong dependence on whether the structured component of the stimulus was the background (near-circular contours, second column) versus the target (elongated contours, third column). $N = 6$.

For all of the pairwise combinations of the second-order statistics, there was no consistent difference between conditions in which the target was random and the background was structured versus conditions in which the target was structured and the background was random (second and third columns of Figure 3).

Third-order statistics

Figure 4 shows a parallel set of results for pairings of third-order statistics. Since all four third-order statistics are aligned with the cardinal axes, there are two cases to consider: statistics that differ by 90° from each other (top row: $\theta_\lrcorner$ and $\theta_\llcorner$), and statistics that differ by 180° from each other (bottom row: $\theta_\lrcorner$ and $\theta_\sqsubset$). In contrast to the findings for pairwise interactions of second-order statistics, thresholds are markedly elevated when the two statistics have opposite signs versus when they are the same. Correspondingly, I_{pool} ranged from 1.60 to 2.55 for individual subjects (mean = 2.05, $N = 6$) for statistics that differ by 90°, and from 1.15 to 1.74 (mean = 1.45, $N = 6$) for statistics that differ by 180° (Figure 6, below).

For these image statistics, threshold depended on which component of the stimulus contained the statistical structure—i.e., whether the background was structured and the target was random (second column of Figure 4), or, alternatively, the target was structured and the background was random (third column of

Figure 4). Cancellation occurred primarily in the latter case. This is clearest for the statistics that differ by 180° ($\theta_\lrcorner$, $\theta_\sqsubset$): I_{pool} ranged from 1.54 to ∞ for individual subjects (mean = 2.60) when the target was structured, but was 0.90 to 1.27 (mean = 1.12) when the background was structured. Note that for two subjects, $I_{pool} = \infty$ in the structured-target condition. That is, for these individuals, thresholds were too high to measure reliably when image statistics had equal and opposite signs, implying nearly complete cancellation. No such cancellation occurred when the statistical structure was in the background, or when image statistics had the same sign. Across subjects, I_{pool} for the two kinds of conditions (structured target vs. structured background) differed by approximately a factor of two for third-order statistics; for the second-order statistics, the difference was less than 30% (Figure 6).

In sum, pairwise interactions of third-order statistics differ in two ways from pairwise interactions of second-order statistics: Cancellation for statistics that differ in orientation and have opposite sign is more prominent, and thresholds depend on whether the statistically structured region is the target versus the background.

Mixed orders

The survey of interactions between pairs of oriented image statistics is completed in Figure 5, which shows the interactions between oriented image statistics of
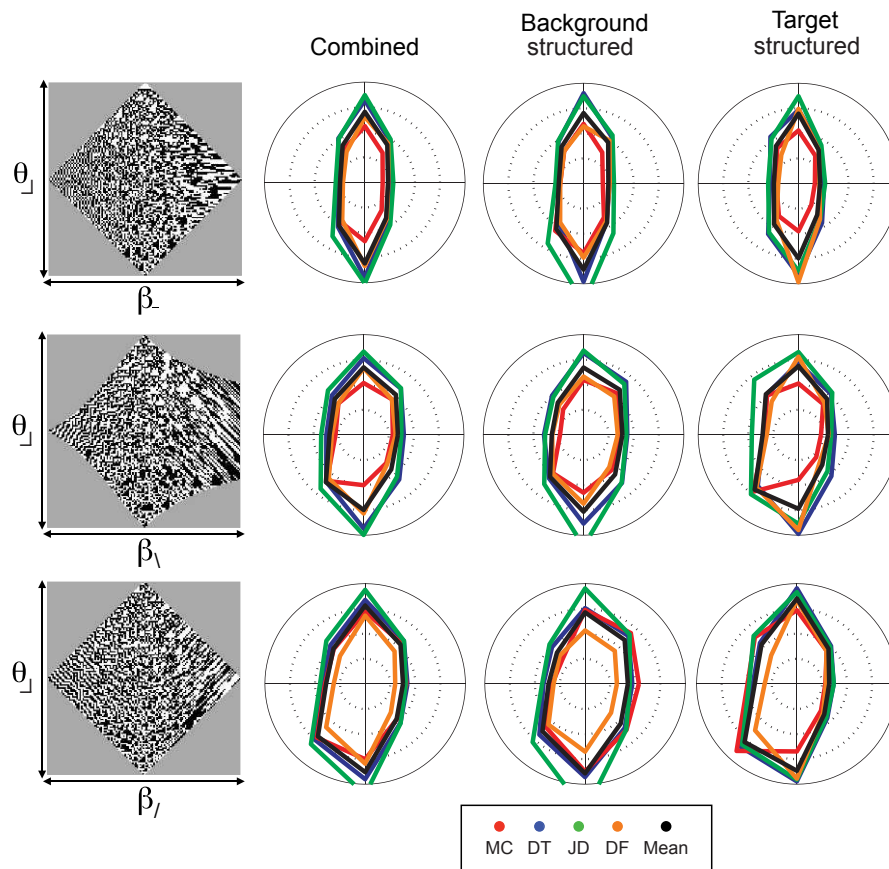
Figure 5. Isodiscrimination contours for mixtures of second- and third-order statistics. As in Figure 3, the image at the left of each row shows the gamut of images for each pair of image statistics, and the polar plots are isodiscrimination contours, i.e., the location in the gamut at which criterion performance was reached. For other plotting conventions, see Figure 3. Isodiscrimination contours are approximately elliptical, with the short axis along the β-axis, and the long axis along the θ-axis, corresponding to the lower thresholds for β than for θ (see Figure 2). There is little dependence on whether the background or the target was structured. $N = 4$.

different orders. There are three cases to consider. When the second-order statistic is oriented in a cardinal direction, it is necessarily aligned with one of the "arms" of the third-order statistic, no matter which third-order statistic is chosen. So all of the eight possible pairings of one cardinal β (β_- or $\beta_|$) with one of the θ's are identical, other than a rotation or a reflection. Since we found no significant differences between the individual statistics that differ solely by a rotation, we focused on one example, (β_-, $\theta_\lrcorner$) (first row of Figure 5).

The other two cases to consider involve a pairing of a second-order statistic oriented in an oblique direction with one of the θ's (say $\theta_\lrcorner$). The two cases differ according to whether the second-order statistic involves the vertex of the $\lrcorner$ (i.e., $\beta_\backslash$, $\theta_\lrcorner$), or alternatively, it spans across the vertex (i.e., $\beta_/$, $\theta_\lrcorner$). These cases are shown in the last two rows of Figure 5.

The three cases have several features in common. First, threshold does not depend on whether the structured region is the background versus the target. Second, the isodiscrimination contours are largely

aligned with the coordinate axes, indicating that the interactions between the statistics is small. But interestingly, these small interactions are consistent: All isodiscrimination contours are slightly tilted into the quadrant in which the image statistics share the same sign. Correspondingly, the value of I_{pool} is slightly less than 1: 0.90 for (β_-, $\theta_\lrcorner$), 0.81 for ($\beta_\backslash$, $\theta_\lrcorner$), and 0.77 for ($\beta_/$, $\theta_\lrcorner$) (means across $N = 4$ subjects, range 0.73 to 0.96). In qualitative terms, the tilt of the isodiscrimination contour translates into the statement that detection of pairwise correlation ($\beta > 0$) is slightly enhanced in the context of large dark regions ($\theta < 0$), and detection of pairwise *anti*correlation ($\beta < 0$) is slightly enhanced in the context of large bright regions ($\theta > 0$), independent of their relative orientations.

Combinations of four oriented image statistics

To determine whether the interactions between pairs of oriented image statistics can account for the way that orientation statistics are processed in a more general
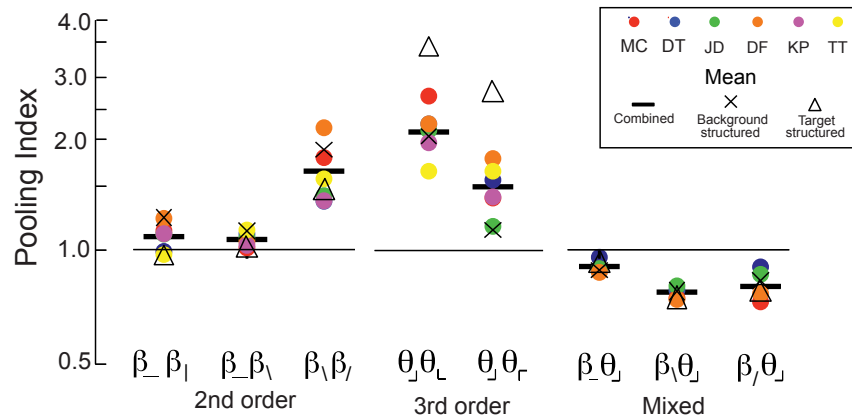
Figure 6. Pooling indices. The pooling index (Equation 8) describes how a pair of image statistics interact: It is the ratio of the thresholds when they have opposite signs, to when they have the same sign; values larger than one indicate cancellation in the opposite-sign condition. For individual subjects, pooling index is calculated from all trials. For the group means, the pooling index is calculated from all trials (solid bar), and for the subsets of trials separated according to whether the background was structured or the target was structured. $N = 6$ for pairings within order (second- or third-order), $N = 4$ for pairings between orders.

context, we measured thresholds for selected combinations of four image statistics. In view of the contrasting way that second- and third-order statistics interacted amongst themselves, we considered images that were specified by multiple statistics of the same order. To emphasize the distinction between independent and pooled processing, we focused on images in which all four values of the β's, or all four values of the θ's, had the same magnitude but might differ in sign. For independent processing, thresholds would be independent of sign. For pooled processing, thresholds would be markedly increased when statistics had opposite signs, because their influences would cancel. We first consider the experimentally-determined thresholds, and then the predictions of models based on pooled processing, independent processing, and an intermediate scenario.

Results for the second-order statistics are shown in Figure 7A, for two experienced subjects and one novice (DC). We found that the thresholds for stimuli with opposite-sign β's (filled circles) differed from the thresholds with same-sign β's, but the difference was modest. When the cardinal β's (β_-, $\beta_|$) were positive and the oblique β's ($\beta_\backslash$, $\beta_/$) were negative, the thresholds ranged from being the same as that of the same-sign β's, to somewhat lower (range of ratios, 0.76 to 0.99). When the cardinal β's (β_-, $\beta_|$) were negative and the oblique β's ($\beta_\backslash$, $\beta_/$) were positive, the thresholds ranged from being similar to that of the same-sign β's, to moderately higher (range, 1.00 to 1.74). In contrast, for the third-order statistics, thresholds with opposite-sign θ's were almost threefold higher than for same-sign thresholds (range, 2.62 to 2.95). There was virtually no difference between thresholds with all-positive θ's and all-negative θ's (ratio range, 0.94 to 1.03).

Thus, in qualitative terms, the findings for the pairwise combinations extend to more complex combinations: For combinations of third-order statistics, thresholds depend markedly on whether the statistics have the same sign or opposing signs; for combinations of second-order statistics, this dependence is more modest.

To determine whether the findings for pairwise combinations extend to the four-parameter combinations in a quantitative way, we consider three models: two models formalize the extremes of sign-dependent and sign-independent interactions, and a third model formalizes an intermediate scenario. In each model, the image statistics combine to form a decision variable, and threshold is reached when the decision variable reaches a criterion magnitude (see Methods).

Complete pooling model

To model strongly sign-dependent interactions, we consider a scenario in which image statistics are completely pooled—that is, a model in which the decision variable is simply the linear sum of image statistics. Using c_k to represent individual image statistics (for mixtures of second-order statistics, $c_1 = \beta_-$, $c_2 = \beta_|$, $c_3 = \beta_\backslash$, and $c_4 = \beta_/$; for mixtures of third-order statistics, $c_1 = \theta_\lrcorner$, $c_2 = \theta_\llcorner$, $c_3 = \theta_\ulcorner$, and $c_4 = \theta_\urcorner$) and q_k to represent the sensitivity to the statistic c_k, this model states that threshold is reached when

$$\left| \sum_k q_k c_k \right| = 1. \tag{9}$$

For the image statistics considered here, the model can be simplified based on the experimental findings presented above. For second-order statistics, the
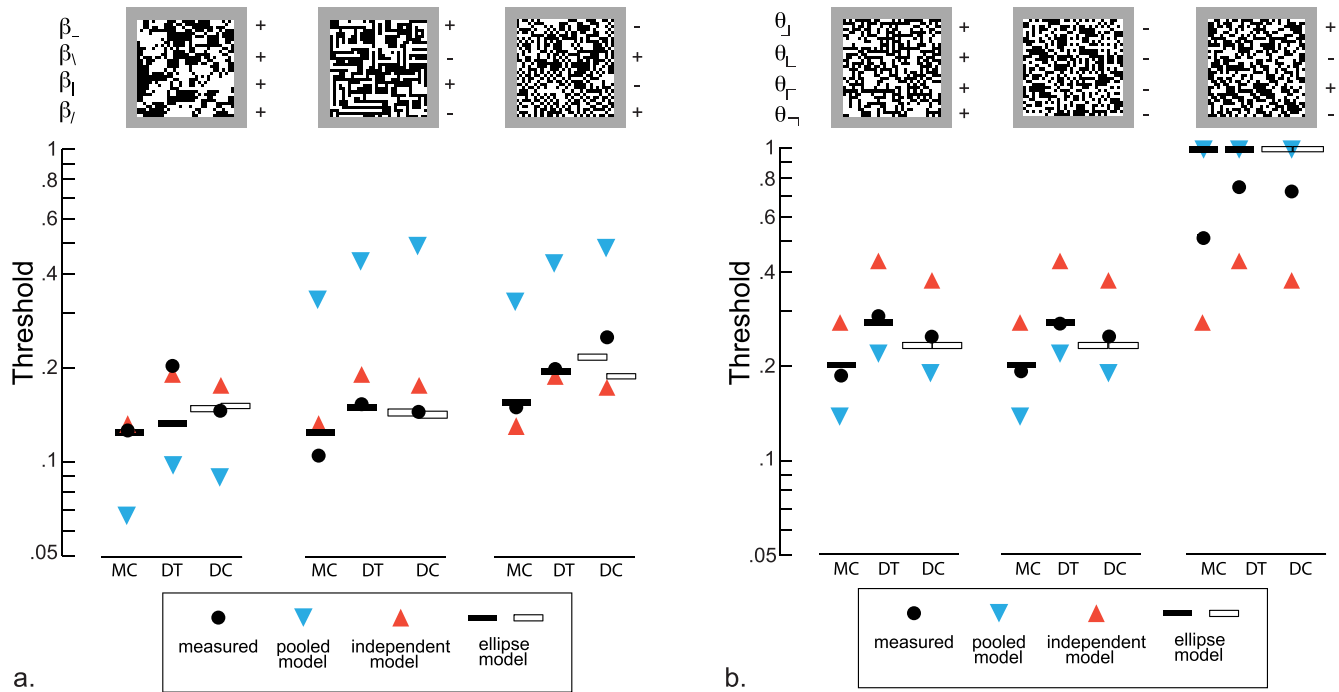
Figure 7. Thresholds for four-component mixtures of second-order statistics (A) and third-order statistics (B), and comparison with model predictions. Measured thresholds (solid symbols) are shown for each kind of mixture; these are compared with predictions based on independent processing of each statistic (upward triangle), complete pooling (downward triangle) and the ellipse model (bars). Predicted thresholds exceeding 1 (which is outside of the gamut) are plotted at 1. Since subject DC did not participate in experiments in which thresholds for pairs of image statistics were measured, the interaction parameters were determined either from subject MC (left open bar) or DT (right open bar); see text for further details. Example images, shown above each set of measured thresholds, have 32 × 32 checks, and are constructed with all image statistics set to ±0.225.

thresholds for β_- and $\beta_|$ were statistically indistinguishable, so they can be set to the same common value, q_{card}. Similarly, the sensitivities for $\beta_\backslash$ and $\beta_/$ were indistinguishable, and we set them to the common value, q_{obl}. With these substitutions, Equation 9 simplifies to

$$\left| q_{card}(\beta_- + \beta_|) + q_{obl}(\beta_\backslash + \beta_/) \right| = 1. \quad (10)$$

For third-order statistics, the sensitivities for the three measured θ's ($\theta_\lrcorner$, $\theta_\llcorner$, $\theta_\ulcorner$) are indistinguishable, and we assume that the threshold for $\theta_\urcorner$ shares the same value, which we designate by q_θ. So for third-order statistics, Equation 9 simplifies to

$$q_\theta|\theta_\lrcorner + \theta_\llcorner + \theta_\ulcorner + \theta_\urcorner| = 1. \quad (11)$$

Equations 10 and 11 make predictions for the thresholds for the stimuli shown in Figure 7, by assigning the β's or the θ's to common multiples of ±1. For example, to determine the model's predicted threshold T_{++++} for a mixture of all positive β's, we set $\beta_- = \beta_| = \beta_\backslash = \beta_/ = T_{++++}$ in Equation 10, to find that

$$T_{++++} = \frac{1}{2(q_{card} + q_{obl})}. \quad (12)$$

Similarly, to determine the predicted threshold T_{++-} for a mixture of positive cardinal βs and negative oblique β's, we set $\beta_- = \beta_| = T_{++-}$ and $\beta_\backslash = \beta_/ = -T_{++-}$ in Equation 10, to find that

$$T_{+-+-} = \frac{1}{2(q_{card} - q_{obl})}, \quad (13)$$

which is much larger than T_{++++}. These values are plotted in Figure 7A (downward triangles). As is shown, they are at odds with the measured thresholds: The measured same-sign thresholds are larger than predicted by the pooled model, and the opposite-sign thresholds are much smaller than predicted. For the third-order statistics, on the other hand, the model is at least qualitatively consistent with experimental findings (Figure 7B). For same-sign θ's, the predicted threshold (from Equation 11, by taking $\theta_\lrcorner = \theta_\llcorner = \theta_\ulcorner = \theta_\urcorner = T_{++++}$) is ($1/4q_\theta$). This modestly underpredicts the experimental threshold. For opposite-sign θ's, the predicted threshold is infinite, since when the θ's have opposite signs in pairs, $\theta_\lrcorner + \theta_\llcorner + \theta_\ulcorner + \theta_\urcorner = 0$. Measured thresholds are in fact infinite in some subjects, and in others, it is close to the maximal value that could be measured (Figure 4). In sum, the complete pooling model fails dramatically to predict the measured thresholds for combinations of

second-order statistics (it predicts a large dependence on relative sign, when the data show only modest sign-dependence). For third-order predictions, the failure is only a quantitative one, as the predicted large sign-dependence is in fact observed.

Independent processing model

To model sign-independent interactions, we posit that there are separate channels for each image statistic, and that the decision variable corresponds to the total energy across channels. Again using c_k to represent individual image statistics and q_k to represent the sensitivity to the statistic c_k, this model states that threshold is reached when

$$\sum_k (q_k c_k)^2 = 1 \tag{14}$$

(see also Methods, Equation 6). Because of the similarity of the thresholds for the two cardinal β's, the two oblique β's, and the θ's, this equation reduces to

$$q_{card}^2 = (\beta_-^2 + \beta_|^2) + q_{obl}^2(\beta_\backslash^2 + \beta_/^2) = 1 \tag{15}$$

for mixtures of second-order statistics and

$$q_\theta^2 \left(\theta_\lrcorner^2 + \theta_\llcorner^2 + \theta_\ulcorner^2 + \theta_\urcorner^2 \right) = 1 \tag{16}$$

for mixtures of third-order statistics. As expected, the form of Equations 15 and 16 shows that the independent-processing model predicts thresholds for mixtures that are independent of whether the statistics have the same sign, or opposite signs.

These predictions are compared with the measured thresholds in Figure 7 (upwards triangles). For second-order statistics, the predictions of Equation 15 are reasonably close to the measured thresholds (within 20% for subjects MC and DT, but approximately a 50% deviation for one of the combinations for subject DC.) For third-order statistics, the deviations are larger: up to a factor of two between the predictions of Equation 16 and measured values (Figure 7B). In both cases, the reason for the mismatch between the independent model and the data is that the independent model predicts that thresholds do not depend on whether the image statistics have the same sign. For the second-order statistics, the measured sign-dependence is, in fact, small (Figure 3), but for third-order statistics (Figure 4) thresholds are substantially higher when signs are opposite than when signs are the same.

An intermediate scenario

The previous two sections showed that both the complete-pooling model and the independent-processing model fail to account for the observed thresholds

for four-component mixtures, but they fail in very different ways: The complete-pooling model fails severely for second-order statistics, while the independent model fails severely for third-order statistics. Since these two models are at opposite ends of the spectrum, these failure modes suggest that a model that is intermediate between complete pooling and independent processing may capture the main features of the threshold data for the four-component mixtures.

The intermediate model posits a quadratic combination rule for the individual image statistics, but allows for interactions between them. In geometric terms, this is equivalent to assuming that the threshold isodiscrimination surface is an ellipsoid (analogous to the Macadam ellipsoids used to quantify color discriminations; Macadam, 1942; Poirson et al., 1990). The intersection of this ellipsoid with each coordinate plane is, necessarily, an ellipse. Finding the parameters of the ellipsoid amounts to adjustment of the in-plane ellipses so that they match the isodiscrimination contours determined in the pairwise-interaction experiments (Figures 3, 4, and 5). In this geometric view, a four-component mixture corresponds to a ray whose direction is determined by the proportions of the four image statistics. The model's prediction for the threshold for this mixture is the point at which this ray pierces the 4-dimensional isodiscrimination ellipsoid.

Thresholds predicted by these models are indicated by the horizontal bars in Figure 7. Most predicted thresholds are within 20% of their measured values, and many are within 10%. For both second-order (panel A) and third-order (panel B) mixtures, the ellipsoidal model captures the qualitative features of the measured thresholds. For the second-order mixtures, the model correctly predicts that the threshold for the all-positive β-condition is intermediate between the predictions of the pooled model and the independent model. For the condition in which the cardinal β's are positive and the oblique β's are negative, it correctly predicts thresholds that are lower than both the pooled and independent model. For the condition in which the cardinal β's are negative and the oblique β's are positive, it correctly predicts thresholds that are close to, but somewhat higher than, the thresholds predicted by the independent model. With the exception of the all-positive β condition for subject DT, where there is a 50% deviation, all thresholds are correctly predicted within 20%, and many within 10%. A similar level of agreement is seen for the mixtures of third-order statistics (panel B): For same-sign θ's, the threshold values are correctly predicted at values intermediate between the predictions of the pooled and independent models. For opposite-sign θ's, the ellipsoid model predicts that thresholds are too high to measure; the experimental data show that they are much higher than for the same-sign θ's, but close to the limits that can be measured.

Discussion

The broad motivation for this study is to understand how local orientation information is analyzed. The importance of extracting local orientation for early visual processing is obvious and well-recognized, but the problem is also a complex one. The reason is that there are many kinds of cues to orientation: Orientation can be signaled not only by positive pairwise correlations (as is the case for lines and gratings), but also by negative pairwise correlations, and by multi-point correlations of any order. Even at a single location, different kinds of cues can coexist, and more than one orientation can be present. Thus, to understand how local orientation is processed, it is necessary to determine the sensitivity of the visual system to these different kinds of cues, and how they interact.

To begin to approach this problem, we developed a space of binary texture stimuli that contain two different kinds of orientation cues—a "second-order" cue, akin to what is present in gratings, and a "third-order" cue that is not present in visual stimuli that are routinely studied. (As mentioned above, we use the term "order" in its mathematical sense—the number of image points that must be simultaneously inspected to detect the cue.) Each of these kinds of cues can be introduced with either positive or negative sign, and with graded strength, independently and at any of four orientations (Victor & Conte, 2012). The stimulus space is "calibrated": The statistical strength of each cue is equally strong and independent of sign, and each of the cues is independent. That is, for the ideal observer, the isodiscrimination contours are spheres centered at the origin, and their intersection with each of the coordinate planes are identical circles. This setup enables measurement of visual sensitivity to positive and negative variations of each statistic and to their combinations and comparison of these sensitivities on an equal footing (namely, with reference to that of the ideal observer).

For human subjects, a relatively simple picture emerges, but one that differs in several respects from that of an ideal observer. These differences were present both with regard to sensitivities to individual statistics, and their interactions.

For individual statistics (Figure 2), human sensitivity is selective: For second-order cues, it is two to three times higher than for third-order cues. Sensitivity to positive and negative deviations of image statistics was not significantly different for second-order statistics. This is noteworthy, as positive deviations correspond to positive correlations in a particular direction (a standard orientation cue), while negative deviations correspond to *anti*correlations in a particular direction. At third-order, there were subtle differences in sensitivity for positive and negative correlations, with sensitivity to positive correlations (corresponding to white oriented regions) about 10% greater than sensitivity to negative correlations (corresponding to black oriented regions). We emphasize that the positive and negative quantities considered here are the *correlations* of image pixels, not the contrast polarity of the image tokens themselves (Motoyoshi & Kingdom, 2007).

With regard to how image statistics interact, we found a difference between second- and third-order statistics: For third-order statistics, opposite-sign combinations of statistics at different orientations resulted in higher thresholds than same-sign combinations (Figure 4); cancellation was much less prominent for second-order statistics (Figure 3). This indicates that third-order statistics are processed in a manner that is more pooled across orientations than second-order statistics (Figure 6). The gamut from pooling to independence can be formalized in terms of a quadratic model of cue combination; this model predicted thresholds to combinations of four orientated image statistics, usually within 20% (Figure 7). The analysis was simplified by the finding (Figure 5) that there was little interaction between oriented statistics of different order. Finally, since we used a figure-ground task to assay sensitivity, we were able to determine whether image statistics were processed differently when they defined the target object versus the background. No such difference was found for second-order statistics (Figure 3), but for third-order statistics (Figure 4), pooling across orientation was more prominent within a target than within the background.

It is worth emphasizing that even though there is greater pooling across orientations for third-order cues than for second-order ones, this pooling is incomplete: Positive orientation signals at one orientation only partially cancel negative orientation signals at another. The fact that cancellation is only partial has an implication for mechanism: Specifically, it rules out the nonoriented mechanism that Motoyoshi et al. (2007) proposed to account for perception of some high-order aspects of texture related to surface qualities. The Motoyoshi et al. (2007) mechanism consists of an ON and an OFF pathway, each containing linear circularly-symmetric center-surround filters. The outputs of these filters are processed by a strong rectifying nonlinearity, and the spatially pooled outputs of these nonlinearities are compared to yield an estimate of the skewness of the luminance distribution of the original image, or of a linearly-filtered transformation of it. However, this kind of mechanism cannot account for detection of structure in images based on two opposite-signed θ-statistics, because—as we show below—such stimuli must generate an equal and symmetric distribution of signals in the ON and OFF pathways.

To see that this is the case, we apply a symmetry argument to the texture T in which $\theta_\lrcorner$ and $\theta_\llcorner$ (a pair that differ by 90°) are present with equal magnitudes

but opposite signs. Since the ON filters of Motoyoshi et al. (2007) are circularly symmetric, they must yield the same distribution of responses x from this texture, and from the left-right mirror reflection of it, i.e., $p_T^{ON}(x) = p_{mirror(T)}^{ON}(x)$. Since $\theta_\lrcorner$ and $\theta_\llcorner$ carry opposite signs, this left-right reflection—which interchanges one statistic with the other—inverts each of their values. Sign-inversion of the θs is equivalent to exchanging black for white (contrast inversion). Since the filters are assumed linear, contrast inversion of the image (consequent to exchanging T for its mirror) results in an inversion of the distributions of signals that emerge from the filters, so $p_{mirror(T)}^{ON}(x) = p_T^{ON}(-x)$. Combining $p_T^{ON}(x) = p_{mirror(T)}^{ON}(x)$ with $p_{mirror(T)}^{ON} = p_T^{ON}(-x)$ yields $p_T^{ON}(x) = p_T^{ON}(-x)$. Similarly, for the OFF pathway, $p_T^{OFF}(x) = p_T^{OFF}(-x)$. Thus, prior to the nonlinearity, the signals on both ON and OFF pathways are even-symmetric (and consequently, have no skewness). Moreover, the distributions of signals on the ON and OFF pathways are identical, since linearity requires that $p_T^{ON}(-x) = p_T^{OFF}(x)$. This means that opponent processing must lead to a null signal, regardless of an intervening nonlinearity prior to pooling.

For a pair of θ-statistics that differ by 180° (such as $\theta_\lrcorner$ and $\theta_\ulcorner$), a similar argument holds, based on a 180° rotation of the texture, rather than a mirror flip.

Possible physiologic basis

To address the physiologic basis of the psychophysical findings reported here, the first consideration is that for the ideal observer, each image statistic is independent and equally informative. Since all the stimulus checks are all readily visible (14 min checks, 1.0 contrast), the selective sensitivity we observe must reflect the properties (i.e., the limitations) of neural processing.

The second-order statistics are readily detected by linear filters, since pairwise correlations directly affect the spatial-frequency content of the stimulus. Thus, oriented linear filters—a fundamental component of standard models of V1 neurons (Rust & Movshon, 2005)—suffice to extract second-order cues. However, third-order statistics do not influence pairwise correlations, and therefore do not influence the average responses of such linear filters. Simple rectification (Victor & Conte, 1991) does not account for their perceptual salience, and since power is not influenced by third-order statistics, overall gain controls cannot have a significant effect. However, it is likely that actual V1 neurons (in contrast to models of them) can extract this kind of orientation information: A fraction of V1 neurons demonstrate orientation-selective responses to images containing third-order orientation cues (Victor, Yu, Schmid, Hu, & Mechler, 2011). This meshes with the inference (see Table 1) that the psychophysical findings reported here result from

local processing, rather than identification of emergent or large-scale structure. Not surprisingly, and in line with the results reported here, neural sensitivity to oriented third-order statistics was found to be substantially less than to second-order statistics.

The neural mechanisms that may underlie sensitivity to combinations of orientation cues are less-well studied and are restricted to combinations of simple cues. Neurophysiological studies suggest that interactions between orientation cues are not prominent, and when present, can be accounted for by a gain control or suppression by non-preferred orientations. In macaque V2, a systematic study of neuronal responses to pairs of line tokens (Ito & Komatsu, 2004) showed that approximately a quarter of neurons respond selectively to orientation pairs. In these cells, the interactions between orientations (Ito & Goda, 2005) could be accounted for by linear summation of signals from each of the tokens, along with a suppressive signal from nonpreferred orientations. No corresponding study appears to have been carried out in V1, but the presence of strongly-tuned surround suppression (Das & Gilbert, 1999) in some neurons may lead to a phenomenological interaction between orientations. Moreover, computational modeling of how multiple orientation cues are combined and represented are largely unexplored, as studies have focused on how a single orientation is determined from the activity of a population of neurons (Deneve, Latham, & Pouget, 1999; Series, Latham, & Pouget, 2004), and these neurons are each considered to represent only the second-order kind of cue.

Relationship to previous work

Several previous studies have examined sensitivity to orientation cues beyond those that are present in ordinary gratings. Most work (Baker & Mareschal, 2001; Landy & Henry, 2007; Landy & Oruc, 2002; Larsson et al., 2006) makes use of stimuli in which the presence of a local feature (such as contrast or noise) is modulated by a grating carrier. These stimuli (traditionally called "second-order") correspond to *fourth-order* stimuli in the current terminology, since in general, four points are needed to extract the orientation. As such, while those studies do not make use of the same kinds of stimuli used here, they demonstrated that high-order orientation cues are salient and identified the crucial stimulus characteristics (Landy & Henry, 2007; Landy & Oruc, 2002). However, these studies stopped short of determining selectivity compared to that of the ideal observer, or how cues combine—since it is not obvious how to use those stimuli to build a "calibrated" space of the sort used here. It is also worth noting that these higher-order cues are extracted by V1 neurons (Baker & Mareschal, 2001), and that simple models (consisting of a

nonlinear subunit followed by an oriented filter) suffice to account for this. Conversely, the third-order stimuli used here are well-known (Julesz et al., 1978), but their capacity to carry orientation information, and their interaction with simple (i.e., second-order) orientation cues has not previously been investigated.

Conclusion

The set of statistical features that can carry orientation information is large and complex and includes correlations of different signs and orders—which may be present alone and in combination. To understand how these cues are processed, we introduced a stimulus set in which multiple oriented image statistics could be independently manipulated. A relatively simple picture of human visual sensitivity emerged: Sensitivity to positive and negative correlations are approximately similar; statistics of different orders are processed largely independently; and within orders, a quadratic combination rule (with greater cross-orientation pooling of third-order statistics than of second-order statistics) accounts for the bulk of the interactions.

Keywords: high-order statistics, visual textures, perceptual metric, corner

References

Baker, C. L., Jr., & Mareschal, I. (2001). Processing of second-order stimuli in the visual cortex. *Progress in Brain Research*, *134*, 171–191.

Ben-Shahar, O., & Zucker, S. W. (2004). Sensitivity to curvatures in orientation-based texture segmentation. *Vision Research*, *44*, 257–277.

Caelli, T., & Julesz, B. (1978). On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics*, *28*(3), 167–175.

Caelli, T., Julesz, B., & Gilbert, E. (1978). On perceptual analyzers underlying visual texture discrimination: Part II. *Biological Cybernetics*, *29*(4), 201–214.

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

Das, A., & Gilbert, C. D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*, *399*(6737), 655–661.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, *2*(8), 740–745.

Graham, N. (1989). *Visual pattern analyzers*. Oxford: Clarendon Press.

Graham, N., Beck, J., & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research, 32*(4), 719–743.

Hu, Q., & Victor, J. D. (2010). A set of high-order spatiotemporal stimuli that elicit motion and reverse-phi percepts. *Journal of Vision*, *10*(3):9, 1–16, http://www.journalofvision.org/content/10/3/9, doi:10.1167/10.3.9. [PubMed] [Article]

Ito, M., & Goda, N. (2005). Mechanisms underlying the representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *European Journal of Neuroscience*, *33*, 130–142.

Ito, M., & Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *Journal of Neuroscience*, *24*, 3313–3324.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, *290*(5802), 91–97.

Julesz, B., Gilbert, E. N., & Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, *31*(3), 137–140.

Landy, M. S., & Henry, C. A. (2007). Critical-band masking estimation of 2nd-order filter properties. *Perception, 36 ECVP Abstract Supplement*.

Landy, M. S., & Oruc, I. (2002). Properties of second-order spatial frequency channels. *Vision Research*, *42*, 2311–2329.

Larsson, J., Landy, M. S., & Heeger, D. J. (2006). Orientation-selective adaptation to first- and sec-

ond-order patterns in human visual cortex. *Journal of Neurophysiology*, *95*(2), 862–881.

Lu, Z. L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, *35*(19), 2697–2722.

Macadam, D. L. (1942). Visual sensitivities to color vision in daylight. *Journal of the Optical Society of America*, *32*, 247–273.

Maddess, T., Nagai, Y., Victor, J. D., & Taylor, R. R. (2007). Multilevel isotrigon textures. *Journal of the Optical Society of America A: Optics, Image Science, & Vision*, *24*(2), 278–293.

Motoyoshi, I., & Kingdom, F. A. A. (2003). Orientation opponency in human vision revealed by energy-frequency analysis. *Vision Research*, *43*, 2197–2205.

Motoyoshi, I., & Kingdom, F. A. A. (2007). Differential roles of contrast polarity reveal two streams of second-order visual processing. *Vision Research*, *2007*, 2047–2054.

Motoyoshi, I., & Kingdom, F. A. A. (2010). The role of co-circularity of local elements in texture perception. *Journal of Vision, 10*(1):3, 1–8, http://www.journalofvision.org/content/10/1/3, doi:10.1167/10.1.3. [PubMed] [Article]

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

Movshon, J., Adelson, E., Gizzi, M., & Newsome, W. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern Recognition Mechanisms, Experimental Brain Research* (Suppl. 11; pp. 117–151). Berlin: Springer-Verlag.

Poirson, A., Wandell, B., Varner, D., & Brainard, D. H. (1990). Surface characterizations of color thresholds. *Journal of the Optical Society of America*, *A7*, 783–789.

Quick, R. F. (1974). A vector magnitude model of contrast detection. *Kybernetik*, *16*, 65–67.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience*, *8*(12), 1647–1650.

Saarela, T., & Landy, M. S. (2012). Combination of texture and color cues in visual segmentation. *Vision Research*, *58*, 59–67.

Sagi, D. (1988). The combination of spatial frequency and orientation is effortlessly perceived. *Perception & Psychophysics*, *43*, 601–603.

Schofield, A. J., Rock, P. B., Sun, P., Jiang, X., & Georgeson, M. A. (2010). What is second-order vision for? Discriminating illumination vs. material changes. *Journal of Vision*, *10*(9):2, 1–18, http://www.journalofvision.org/content/10/9/2, doi:10.1167/10.9.2. [PubMed] [Article]

Series, P., Latham, P., & Pouget, A. (2004). Tuning curve sharpening for orientation selectivity: Coding efficiency and the impact of correlations. *Nature Neuroscience*, *7*, 1129–1135.

Shephard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.

To, M. P. S., Baddeley, R. J., Troscianko, T., & Tolhurst, D. J. (2011). A general rule for sensory cue summation: Evidence from photographic, musical, phonetic, and cross-modal stimuli. *Proceedings of the Royal Society of London B*, *278*, 1365–1372.

Victor, J. D. (1994). Images, statistics, and textures: Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture [comment]. *Journal of the Optical Society of America A*, *11*(5), 1680–1684.

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

Victor, J. D., & Conte, M. M. (1989). Cortical interactions in texture processing: Scale and dynamics. *Vision Neuroscience*, *2*(3), 297–313.

Victor, J. D., & Conte, M. M. (1991). Spatial organization of nonlinear interactions in form perception. *Vision Research*, *31*(9), 1457–1488.

Victor, J. D., & Conte, M. M. (1996). The role of high-order phase correlations in texture processing. *Vision Research*, *36*(11), 1615–1631.

Victor, J. D., & Conte, M. M. (2004). Visual working memory for image statistics. *Vision Research*, *44*(6), 541–556.

Victor, J. D., & Conte, M. M. (2012). Local image statistics: Maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America A*, *29*, 1313–1345.

Victor, J. D., Yu, Y., Schmid, A. M., Hu, Q., & Mechler, F. (2011). Responses of macaque V1 neurons to local image statistics. *Society for Neuroscience* (Washington, DC), Program No. 799.07. 2011. Online.

Wolfson, S. S., & Landy, M. S. (1998). Examining edge- and region-based texture analysis mechanisms. *Vision Research*, *38*, 439–446.