VR 4183
18 November 2004 Disk Used
ARTICLE IN PRESS
No. of Pages 11, DTD = 5.0.1

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research xxx (2004) xxx–xxx

Vision Research

www.elsevier.com/locate/visres

# Local processes and spatial pooling in texture and symmetry detection

Jonathan D. Victor *, Mary M. Conte

*Department of Neurology and Neuroscience, Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021, United States*

Received 21 April 2004; received in revised form 11 September 2004

**Abstract**

We examined the ability of human observers to detect three kinds of statistical structure in binary arrays: first-order statistics (luminance), local fourth-order statistics (isodipole textures), and long-range statistics (bilateral symmetry). Performance was closest to ideal on the luminance task and furthest from ideal on the symmetry task. For each kind of statistic, the dependence of performance on the degree of structure was well described by a model consisting of an initial stage of multiple independent detectors, followed by a pooling stage. For the luminance task and the isodipole task, performance was well-modeled by local processing followed by extensive spatial pooling. For the symmetry task, limitations at the local detection stage *and* a near-absence of spatial pooling were needed to model for performance.


*Keywords:* Symmetry; Isodipole

## 1. Introduction

A wide range of phenomena related to discrimination and detection of visual textures can be accounted for by two-stage models, consisting of a stage of local processing, followed by a stage of pooling of the local signals over a wider area (Bergen & Adelson, 1988; Chubb & Landy, 1991; Graham, 1989; Graham, Beck, & Sutler, 1992; Grossberg & Mingolla, 1985; Malik & Perona, 1990; Victor & Conte, 1991; Wilson, 1993; Zhu, Wu, & Mumford, 1998). These architectures have in common at least one locally acting nonlinearity in addition to the decision process. This intervening nonlinearity ensures that signals from different kinds of local features (or features of opposite polarities) cannot cancel, and that the model is formally distinguishable from a single linear filter.

Not all kinds of local statistical structure are readily detected by human observers. Observers are insensitive to most local correlations beyond second-order (Julesz, 1981; Julesz, Gilbert, Shepp, & Frisch, 1973; Maddess & Nagai, 2001; Maddess, Nagai, James, & Ankiewcz, 2004; Victor & Conte, 1991). Some fourth-order statistics (e.g., the one that distinguishes the "even" and "odd" isodipole textures) are visually salient (Julesz, Gilbert, & Victor, 1978). But even for discriminations based on this statistic, subjects' performance is substantially worse than that of an ideal observer (Joseph, Victor, & Optican, 1997). Our first goal in this study is to use the two-stage model structure to determine to the contributions of the initial stage of local detection and the second stage of spatial pooling to the less-than-ideal behavior. In some variants of the model framework, more than one local nonlinearity is required to account for the details of dependence of performance on contrast (Graham et al., 1992), or, for the pattern of sensitivity to higher-order correlations (Victor & Conte,

* Corresponding author. Tel.: +1 212 746 2343; fax: +1 212 746 8984.
*E-mail address:* jdvicto@med.cornell.edu (J.D. Victor).

0042-6989/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.visres.2004.10.012

1991). One might therefore hypothesize that the relatively low efficiency for detection of higher-order local correlations is related to the greater complexity of local analysis, and thus, is restricted to the initial stage of the model.

The immense variety of visual textures (Maddess et al., 2004) makes it impossible to study all texture types; instead, as in previous work (Victor & Conte, 2004), we focus on a comparison of luminance-based textures (Fig. 1A) and isodipole textures (Fig. 1B). This choice is motivated by physiological principles and previous work with texture discrimination and segmentation: luminance variations and variations of second-order correlation structure result in changes in spatial contrast, and thus, are anticipated to lead to differences in the overall firing rate of retinal ganglion cells. Conversely, isodipole textures (Julesz et al., 1978), which share a common power spectrum (and luminance), are likely to be detected only after cortical analysis of local variation in firing patterns. Sensitivity of cortical (Purpura, Victor, & Katz, 1994) but not lateral geniculate neurons (Victor, 1986) to these statistics has been demonstrated experimentally.

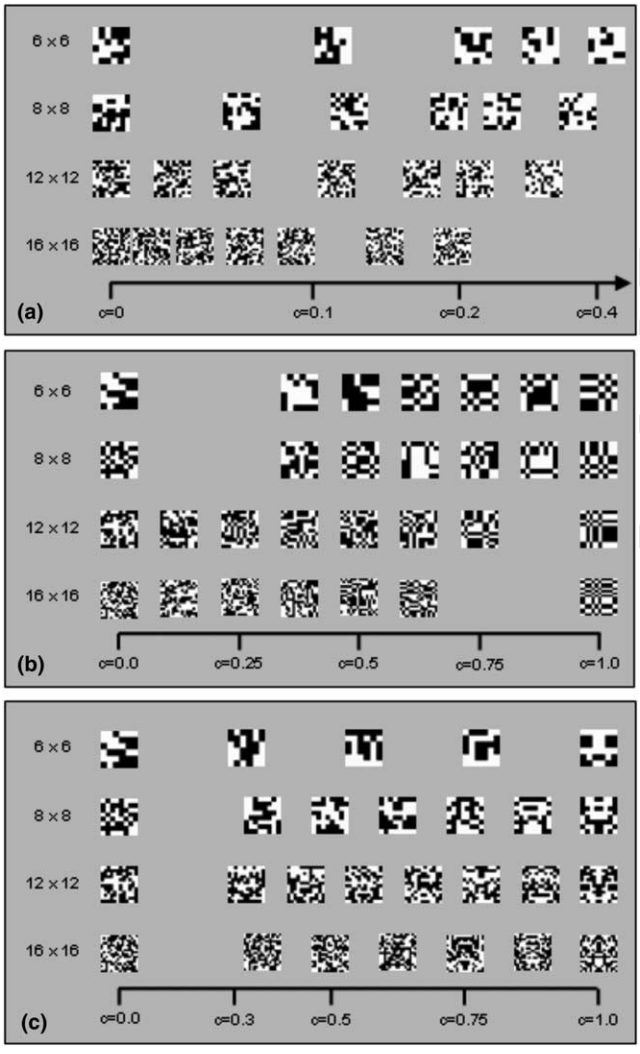


Fig. 1. The three kinds of statistical manipulations used in Experiment I. (A) Luminance statistics. (B) Isodipole statistics. (C) Vertical symmetry.

Our second goal is to compare the characteristics of detection of texture with that of bilateral symmetry (Attneave, 1954; Tyler, 1995; Wenderoth, 1994). We are concerned here with detection of statistical symmetry, rather than perfect symmetry—that is, detection of correlations between the two halves of an image, and not necessarily an exact correspondence. Detection of such statistical symmetry is arguably more relevant to natural vision, in which perfect symmetry is rare, but approximate bilateral symmetry is common (e.g., faces).

Many authors have pointed out that symmetry detection has special characteristics, in terms of spatial pooling, eccentricity dependence, and processing dynamics (Conte, Purpura, & Victor, 2002; Conte & Victor, 2003; Rainville & Kingdom, 1999, 2002; Tyler, 1995, 1999, 2001; Victor & Conte, 2004), though some of these studies (Rainville & Kingdom, 1999, 2002; Tyler, 1995, 1999, 2001) did not make a direct within-subject comparison based on texture stimuli with comparable characteristics. However, at a formal level, the two-stage model framework for texture processing described above can also be applied to detection of statistical symmetry. As in texture processing, independent analyzers can determine whether corresponding parts of an image are correlated, and then the statistical evidence of these correlations can be pooled over a wide area. However, in contrast to texture analysis, the independent analyzers need not be local, and the regions to be compared depend on the location of the symmetry axis. Thus, one might hypothesize that either of the two stages of computation will have different characteristics for symmetry processing than for processing of spatially homogeneous textures. The results reported here, which are based on direct within-subject comparisons and stimuli of comparable spatiotemporal characteristics, show that the processing of symmetry indeed differs from that of spatially homogeneous textures at both levels.

## 2. Methods

### 2.1. Subjects

Studies were conducted in eight normal subjects (3 male, 5 female), ages 21–54. Other than author MC, the remaining subjects were naive to the purpose of the experiments. Subjects were practiced psychophysical observers in a related task involving targets in the same positions relative to fixation (Victor & Conte, 2001), and had visual acuities (corrected if necessary) of 20/20 or better.

### 2.2. Stimuli

The stimulus consists of four arrays of checks on a mean gray background. The arrays were positioned along the cardinal axes, with centers 200 min from fixation (Fig. 2). In Experiment I, each array subtended 160 min and was subdivided into $6 \times 6$, $8 \times 8$, $12 \times 12$, or $16 \times 16$ equally sized contiguous checks, each of which was either black or white (Fig. 2A shows the $8 \times 8$ case, and Fig. 2B shows the $16 \times 16$ case). In three of the four arrays, the luminance values (black or white) were assigned to the checks at random. In the fourth array ("the target"), a particular kind of statistical structure was introduced into the arrays. The subject's task was to identify the target, whose position in the stimulus was chosen at random from the four alternatives.

We used three kinds of statistical structure: luminance, higher-order statistical structure (the "isodipole" textures), and bilateral symmetry. In each case, the strength of the statistical structure was parameterized by a quantity $c$, where $c = 0$ denotes a maximally random assignment, and $c = 1$ denotes a maximally structured assignment. For luminance statistics, an array corresponding to a value $c$ had $\frac{1+c}{2}$ of its checks white, and $\frac{1-c}{2}$ of its checks black. For higher-order statistics, $c = 1$ corresponded to a maximally "even" isodipole texture. For symmetry, $c = 1$ corresponded to a texture in which all pairs of checks that were related by a vertical mirror symmetry axis were matched in luminance. For details on construction of these textures (see Victor & Conte, 2004). As pointed out in that paper, all of these textures are maximum-entropy textures given the above constraints, and that the luminance and isodipole textures are Markov random fields (Pollack, 1971; Zhu et al., 1998).

### 2.3. Apparatus

The above visual stimuli were produced on a Sony Multiscan 17seII (17″ diagonal) monitor, with signals

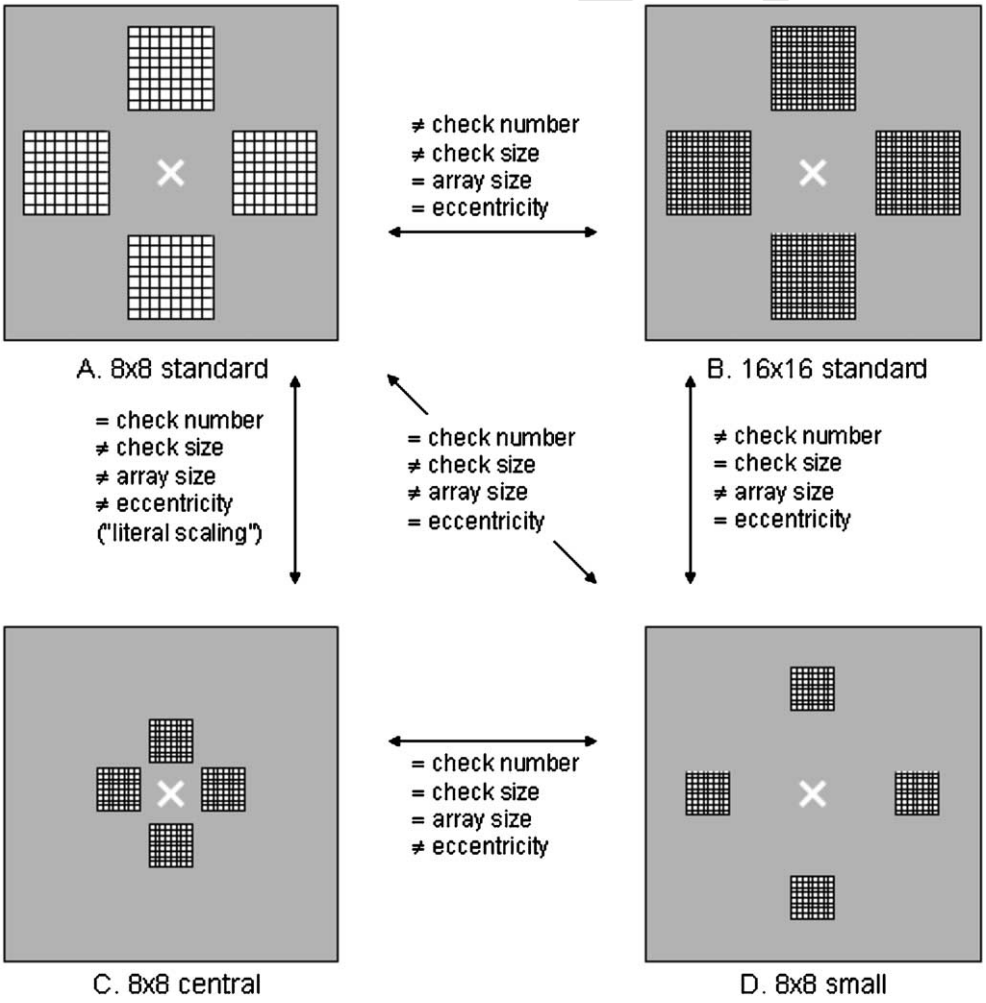


Fig. 2. Scaled diagrams of the layout of the stimulus frame. (A) and (B) $8 \times 8$ and $16 \times 16$ standard arrangement, used in Experiments I and II. (C) $8 \times 8$ arrangement with checks moved centrally, used in Experiment II. (D) $8 \times 8$ arrangement with checks reduced in size, used in Experiment II.

driven by a PC-controlled Cambridge Research VSG2/3 graphics processor programmed in Delphi II to display precomputed maps (generated in Matlab) for specified periods of time. The resulting $768 \times 1024$ pixel display had a mean luminance of 47 cd/m$^2$, a refresh rate of 100 Hz and subtended $11° \times 15°$ (approximately 1 min/pixel) at the viewing distance of 100 cm. The intensity vs. voltage behavior of the monitor was linearized by photometry and lookup table adjustments provided by VSG software. Stimulus contrast was 1.0.

### 2.4. Procedure

Experiments were organized as a sequence of 4-alternative forced choice trials, whose common features are as follows. After binocular fixation on a uniform gray background, the subject initiated a trial via a button-press on a Cambridge Research CT3 response box. Three hundred milliseconds later, a stimulus (S, described in detail above) appeared, consisting of four arrays of checks, surrounding a central "X" subtending approximately 30 min. After presentation of S for 100 ms, a mask was presented for 500 ms, consisting of a full-field random checkerboard whose checks were half as large (linear dimension) as those in S. The mask was followed by a uniform field at the mean luminance. The subject's task was to identify the target array via a button-press on the response box, which had four buttons positioned corresponding to the stimulus arrays. Subjects were instructed to maintain central fixation and to respond as quickly as possible, but not to compromise accuracy for speed. Responses and reaction times (measured with respect to the onset of S) were collected via the Delphi II display software. Trials in which the subject responded before the onset of S, or after 4000 ms, were discarded and repeated.

An experimental session consisted of blocks of 216–224 trials in which an equal number of targets were constructed from a range of 4–7 $c$-values and one kind of statistical structure (luminance, isodipole, and symmetry) and were placed in each of the four possible locations ($4 \times 4 \times 14 = 224$, $5 \times 4 \times 11 = 220$, $6 \times 4 \times 9 = 216$, $7 \times 4 \times 8 = 224$). Typically, four blocks of a single kind of statistical structure but varied spatial parameters (e.g., check size and check number) were combined into a single experimental session, lasting 1–2 h. To accumulate a sufficient number of trials (800–1000) for each condition, data from sessions on separate days were combined. Sessions corresponding to the three statistical structure types were presented in interleaved order, randomized across subjects.

Prior to data collection, subjects received practice and training in the task, as follows. First, they were shown paper exemplars of the targets with the highest $c$-value used, and asked to identify "which array is different". Then, they were asked to perform the above computer-controlled task, but with timing requirements relaxed, beginning with a 500–600 ms duration for S. As performance stabilized, the presentation duration for S was gradually decreased (over 2–8 h of practice, distributed over several sessions) to 100 ms. Once data collection began, subjects were shown paper exemplars of trials with low and high $c$-values at the beginning of each session.

## 3. Results

### 3.1. Psychophysical data

In Experiment I, we measured performance for $6 \times 6$, $8 \times 8$, $12 \times 12$, or $16 \times 16$ arrays. Check number and check size varied reciprocally so that the array size remained fixed at 160 min, and arrays were positioned as in Fig. 2A. For each kind of statistical structure and array size, pilot experiments determined a range of $c$-values that spanned the range from near-threshold to near-ceiling performance (or, to $c = 1$ if near-ceiling performance could not be obtained), as illustrated in Fig. 1. Six subjects were tested with all three kinds of stimuli; one subject was tested with luminance and symmetry stimuli only and one subject was tested with isodipole stimuli only.

Results for two typical subjects are shown in Fig. 3 (subject CC) and Fig. 4 (subject EC). For luminance imbalance, performance at a fraction correct of 0.5 required progressively lower amounts of luminance imbalance as the number of checks increased: $c \approx 0.1$ for $6 \times 6$ arrays, but $c \approx 0.06$ for $16 \times 16$ arrays. The slope of the psychophysical curve also appears to increase as the number of checks increases. For isodipole textures, performance changed more dramatically as the number of checks increased: performance at a fraction correct of 0.5 required maximal levels of statistical structure ($c = 1$) for $6 \times 6$ arrays, but was typically achieved for $c < 0.5$ for $16 \times 16$ arrays. For symmetry, maximal levels of statistical structure ($c = 1$) led to performance that ranged from just barely above chance (Fig. 3) to a fraction correct of 0.5 (Fig. 4), and showed little if any change as the number of checks increased.

In Experiment II (Fig. 5), we manipulated combinations of check size, check number, and eccentricity, as shown in Fig. 2. This experiment was carried out for a subset of four of the seven subjects who participated in Experiment I. There was only a very minor and inconsistent effect of these manipulations. A fourfold reduction in the area of the checks, with the arrays centered in the same locations ("$8 \times 8$ standard" vs. "$8 \times 8$ small") reduced performance slightly in three of four subjects (EC, MH, KS) for the luminance statistics, in one of the subjects (KS) for isodipole statistics (with subject MC showing the opposite trend), and produced

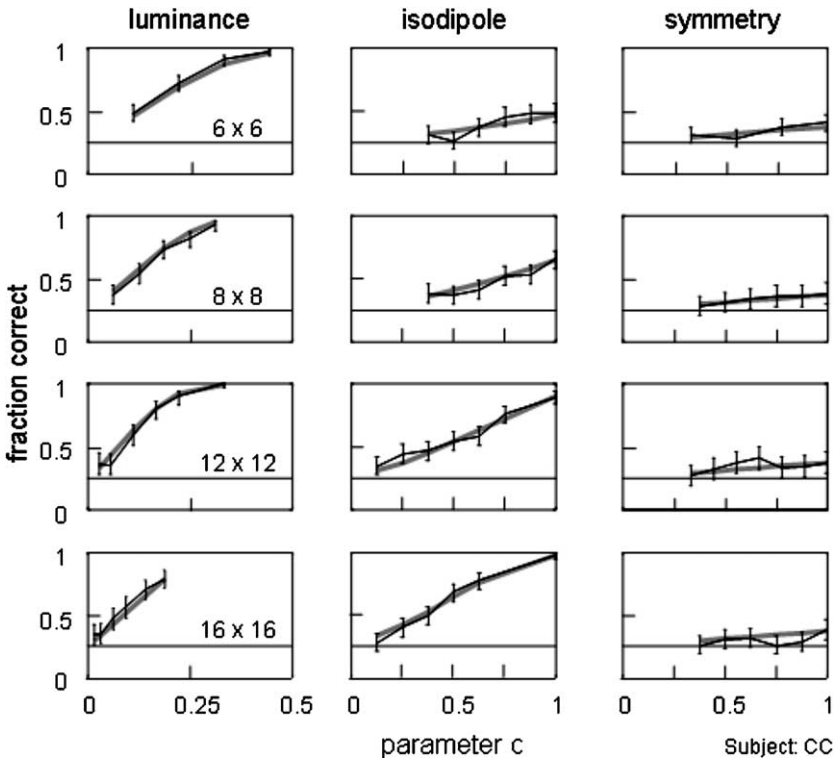


Fig. 3. Psychophysical performance across a range of check numbers, with check size varying inversely to maintain a constant array size, for subject CC (Experiment I). Example stimuli corresponding to each of the abscissa values are shown in Fig. 1. Error bars represent 95% confidence limits for the measured fraction correct, as determined by binomial statistics. The horizontal line at a fraction correct of 0.25 indicates chance performance. The gray trace indicates the fit to the data provided by the model described in the text.

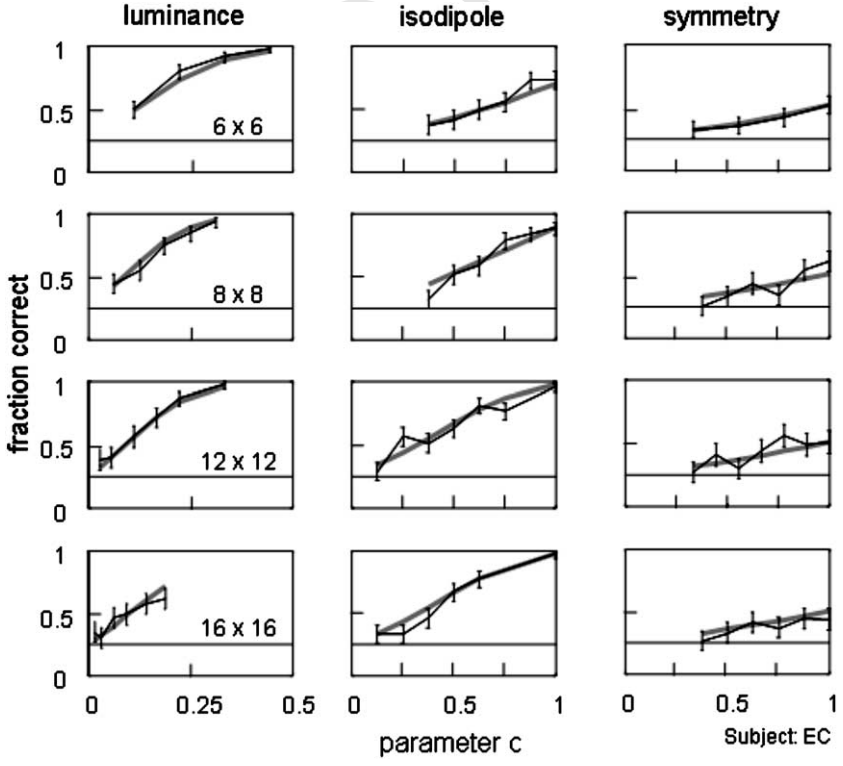


Fig. 4. Psychophysical performance across a range of check numbers, with check size varying inversely to maintain a constant array size, for subject EC (Experiment I). Details as in Fig. 3.

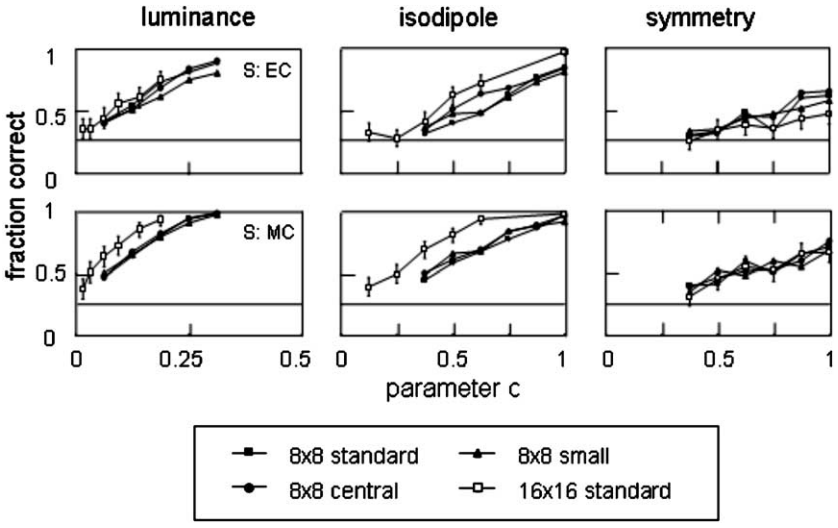


Fig. 5. Dependence of psychometric functions on check size and eccentricity, for subjects EC and MC (Experiment II). The four spatial conditions are shown in Fig. 2. Error bars for the three curves with dark symbols are similar in size to the error bars displayed for the open symbols, and are suppressed for clarity. Plotting details otherwise as in Fig. 3.

no significant change in any subject for the symmetry stimuli. Literal scaling (a twofold reduction in eccentricity, corresponding with a fourfold reduction in check area, and constant check number ("$8 \times 8$ standard" vs. "$8 \times 8$ central")) produced a slight improvement in performance for one subject (EC) for the isodipole statistics, and no significant change for any of the four subjects for the luminance and symmetry stimuli. In contrast, and replicating the findings of Figs. 3 and 4 (in independently measured psychophysical curves), increasing the number of checks in the array fourfold while reciprocally decreasing their area ("$16 \times 16$ standard" vs. "$8 \times 8$ standard") improved performance for the luminance and isodipole stimuli in all four subjects, and produced no change in performance for the symmetry stimuli in three subjects (and a slight worsening in subject EC).

### 3.2. Model

We next attempted to account for the above results in terms of a simple psychophysical model. Our goal was not to construct a detailed correspondence between computational stages and neural elements, but rather, to frame a model that could distinguish local and global limitations on psychophysical performance. The strategy was also guided by the above observation that the number of check elements appeared to be the main determinant of performance, rather than their absolute size or retinal position. This approach is also consistent with the scaling properties of texture detection (Sutter, Sperling, & Chubb, 1995; Victor & Conte, 1989) and symmetry (Rainville & Kingdom, 1999, 2002).

Thus, we chose to formalize a model of the sort used by many authors (Chubb & Landy, 1991; Graham et al., 1992) in which a local detection, which may be nonlinear, is followed by a spatial pooling stage. In view of the above scaling properties, we expressed the spatial aspects of the "back pocket" model in terms of checks (i.e., scaled to the texture), not in terms of visual angle. Local limitations on processing are modeled by imperfect detection of elements of statistical structure, and global limitations are modeled by limits on the extent of spatial pooling of these detected elements. As described below (Fig. 6), a similar model framework could be used for the three kinds of statistical structure, thus allowing a direct comparison of performance limitations. In the first stage of the model, the term "local" is justified because the postulated computational element has the smallest possible area of any computational element that could detect the relevant statistical structure.

For each kind of statistical structure, we postulate that detection of a statistical element is characterized by two probabilities: $p_h$, the "hit rate", and $p_f$, the "false-alarm rate". $p_h$ is the probability that a detector is activated when a statistical element is present, and $p_f$ is the probability that a detector is spuriously activated when the statistical element is not present. For the luminance stimuli, $p_h$ is merely the probability that a bright check is signaled as bright, while $p_f$ is the probability that a dark check is signaled as bright. For the isodipole stimuli, $p_h$ is the probability that a $2 \times 2$ quadruple of checks containing an even number of bright checks (i.e., consistent with the isodipole constraint) is signaled as such, while $p_f$ is the probability that a quadruple that

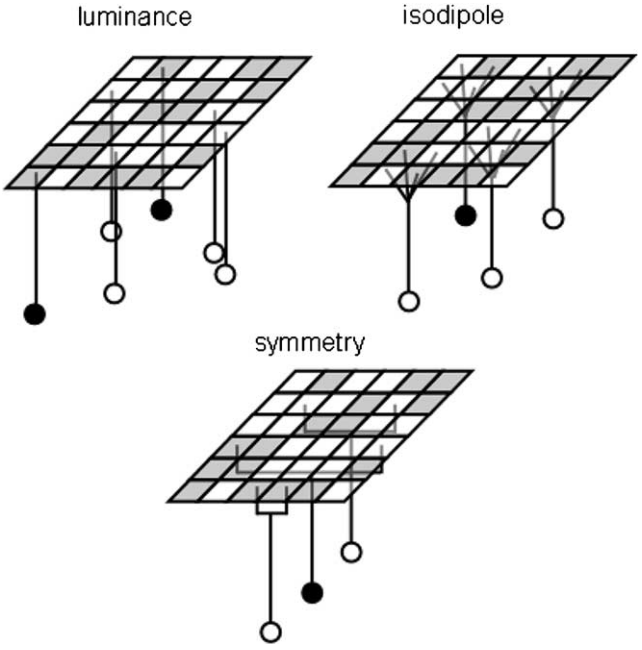


Fig. 6. A diagram of the stochastic back pocket model. At the first stage of the model, independent detectors either detect (state 1, diagrammed as an open symbol) or fail to detect (state 0, diagrammed as a filled symbol) a statistical element, such as a bright check, an even number of bright checks within a $2 \times 2$ array, or a pair of matching checks related by the mirror symmetry axis. The probability of correct detections and false alarms are the model parameters $p_h$ and $p_f$, respectively. Subsequently, the state of up to $N_{max}$ of these detectors, selected at random, is then pooled.

contains an odd number of bright checks is spuriously signaled as being even.

For the mirror-symmetry stimuli, $p_h$ is the probability that a pair of checks related by the mirror axis which match in brightness is signaled as such, while $p_f$ is the probability that a mismatched pair is spuriously signaled as matching. In all cases, we postulate that these local detectors operate independently. Thus, $p_h$ and $p_f$ describe the detection of the smallest statistical elements contained in each stimulus class.

We then postulate that, within each array, a subset of up to $N_{max}$ detectors is chosen at random, and the state of these detectors is pooled by summation (1: detection, 0: no detection). (For an $n \times n$ array of checks, there are $n^2$ luminance elements, $(n-1)^2$ isodipole elements, and $n^2/2$ mirror symmetry elements.) The decision rule is that the subject then selects the array for which the total number of detections is largest (ignoring of course whether the detection events were true or false positives). With this decision rule, a set of parameters $\{p_h, p_f, N_{max}\}$ determines a model psychophysical curve. Ideal behavior corresponds to $p_h = 1$, $p_f = 0$, and $N_{max}$ no less than the total number of statistical elements. Deviations of $p_h$ and $p_f$ from these values characterize limitations in performance due to detection of the statistical elements, while reductions of $N_{max}$, the effective summation area, characterize limitations in performance due to global pooling.

To calculate the model fraction correct $f_{model}(n,c)$, we first used the binomial distribution to determine the exact distribution of number of hits and false alarms for subsets of $N_{max}$ elements of arrays constructed with structure parameter $c$, and for $c = 0$ (the distractor arrays). We then determined the exact probability that the number of detected stimulus elements (total of hits and false alarms) was greatest for the target. This probability was taken to be the fraction correct $f_{model}(n,c)$.

Model parameters were determined independently for each dataset (one subject, one kind of statistical structure, all array sizes, all values of $c$). Parameter values for $\{p_h, p_f, N_{max}\}$ were estimated by minimizing the total squared difference between the observed and modeled fraction correct,

$$R = \sum_{n,c} (f_{obs}(n,c) - f_{model}(n,c))^2, \quad (1)$$

where the summation is over all array sizes $n$ and values of $c$, $f_{obs}(n,c)$ is the observed fraction correct calculated as above, and $f_{model}(n,c)$ is the model fraction correct. Matlab's nonlinear minimization routine `fminsearch` acting on $p_h$, $p_f$, and $1/N_{max}$ was used for this minimization. Confidence limits on the fitted parameters and derived quantities (Fig. 7) were determined via numerical calculation of the Hessian at the minimum point. Goodness of fit was determined by a $\chi^2$-statistic:

$$\chi^2 = \sum_{n,c} \frac{(m(n,c)(f_{obs}(n,c) - f_{model}(n,c)))^2}{m(n,c) f_{model}(n,c)(1 - f_{model}(n,c))}, \quad (2)$$

where $m(n,c)$ is the number of trials performed and the other quantities were defined above. The denominator is the expected variance associated with each count, determined by Bernoulli statistics for the model fraction correct $f_{model}(n,c)$ and the number of trials $m(n,c)$. The number of degrees of freedom was taken to be 3 less than the number of conditions.

For all subjects and all three kinds of statistical structure, the goodness of fit was acceptable ($p > 0.5$). The fitted parameter values are summarized in Fig. 7A–C and Table 1. Although there is considerable scatter across subjects, two trends are clear. First, the effective summation area $N_{max}$ for luminance and isodipole statistics was substantially higher (geometric mean of 85 checks for luminance, 148 for isodipole; $N_{max} > 75$ checks in all but one subject, KS) than for symmetry (geometric mean of 21 checks, $N_{max} < 25$ checks in all but one subject, MH). These differences were highly significant (for luminance vs. symmetry: $p < 0.02$ for two-tailed paired $t$-test comparisons among the six subjects who participated in all experiments, $p < 0.001$ for unpaired comparisons among all seven subjects; for isodipole vs. symmetry: $p < 0.002$ and $p < 0.001$; comparisons based

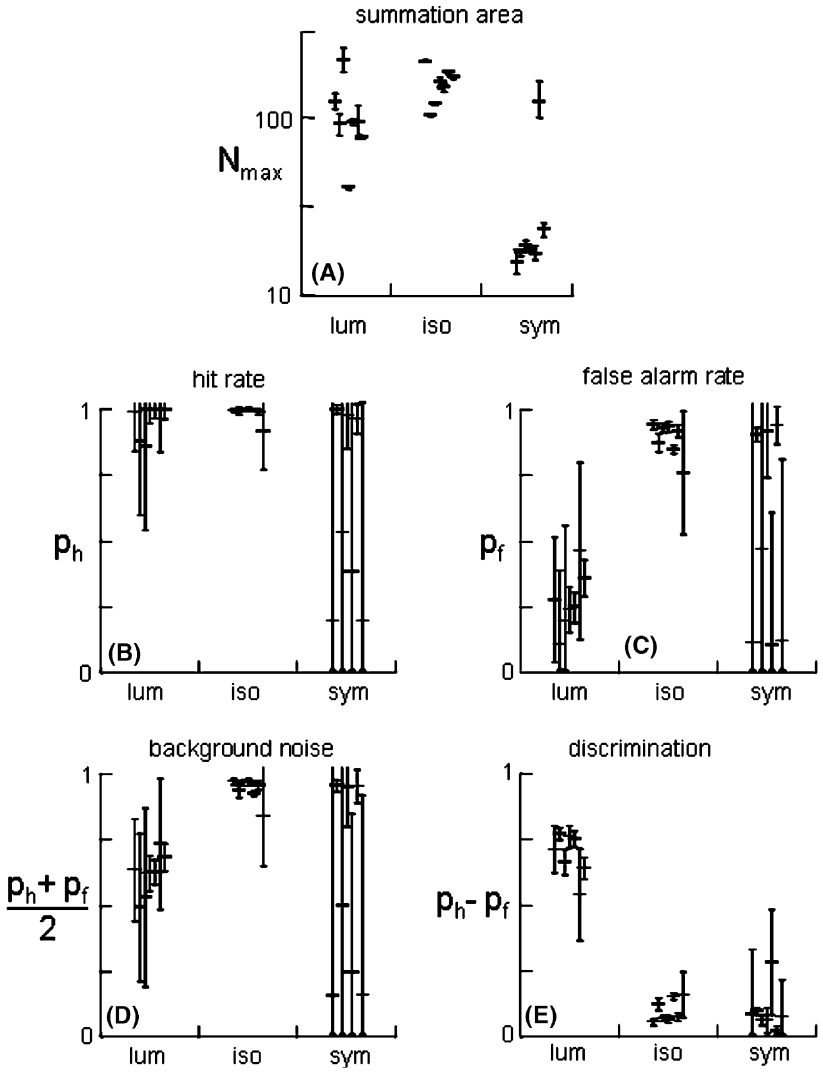


Fig. 7. Summary of model parameters for Experiment I. 95% confidence limits determined as described in text.

on $\log(N_{max})$). The somewhat greater effective summation area for isodipole statistics, compared to luminance statistics, was borderline significant ($p \approx 0.08$ for paired comparisons, $p < 0.02$ for unpaired comparisons).

There also appear to be some differences in the parameters that described local processing: the false alarm rate $p_f$ for luminance is significantly less than for isodipole textures. However, the saturation of the hit rate parameter $p_h$ for luminance and isodipole textures, and the error bars and inter-individual variability for symmetry, make it difficult to discern other differences. The situation is clarified by considering instead two derived parameters: $\frac{(p_h+p_f)}{2}$ and $p_h - p_f$. The former parameter can be regarded as a measure of background noise: it is the average probability that a local processing unit signals the presence of structure when confronted with a fully random ($c = 0$) texture. The latter parameter can be regarded as a measure of local detection. In terms of these parameters (Fig. 7D and E, and Table 1), other differences are apparent. Processing of luminance and isodipole statistics differs both in terms of background noise (significantly lower for luminance than for isodipole textures) and intrinsic local detection (significantly higher for luminance than for isodipole statistics). Local processing of symmetry differs from both of the other two kinds of statistics. In comparison to processing of luminance statistics, background noise is comparable, but local detection is poorer. In comparison to processing of isodipole statistics, background noise is lower, and local detection is poorer as well. (For all of these comparisons, $p < 0.05$ by two-tailed paired and unpaired $t$-statistics.)

In sum, the superior performance on the luminance task compared to the isodipole task reflects more efficient local processing (lower background noise $\frac{(p_h+p_f)}{2}$ and better local detection $p_h - p_f$). The inferior performance on the symmetry task compared to both the luminance and the isodipole task reflects less-efficient spatial combination (lower $N_{max}$) and poorer local detection ($p_h - p_f$).

Table 1
Summary of the model parameters across subjects

| | Luminance | | | Isodipole | | | Symmetry | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | c.l. | | Mean | c.l. | | Mean | c.l. | |
| $N_{max}$ | 85.1 | 59.9 | 121.0 | 147.6 | 122.9 | 177.3 | 21.4 | 12.8 | 35.9 |
| $p_h$ | 0.78 | 0.62 | 0.93 | 0.94 | 0.86 | 1.00 | 0.38 | 0.03 | 0.73 |
| $p_f$ | 0.10 | 0.02 | 0.17 | 0.81 | 0.66 | 0.95 | 0.33 | 0.02 | 0.64 |
| $(p_h + p_f)/2$ | 0.40 | 0.25 | 0.55 | 0.87 | 0.76 | 0.99 | 0.35 | 0.02 | 0.69 |
| $p_h - p_f$ | 0.61 | 0.53 | 0.70 | 0.06 | 0.04 | 0.09 | 0.03 | 0.01 | 0.06 |
| Comparisons | | | | | | | | | |
| | Paired | Unpaired | | Paired | Unpaired | | Paired | Unpaired | |
| $N_{max}$ | 0.083 | 0.017 | | 0.016 | <0.001 | | 0.001 | <0.001 | |
| $p_h$ | 0.037 | 0.085 | | 0.213 | 0.058 | | 0.039 | 0.015 | |
| $p_f$ | <0.001 | <0.001 | | 0.131 | 0.169 | | 0.044 | 0.022 | |
| $(p_h + p_f)/2$ | <0.001 | <0.001 | | 0.835 | 0.808 | | 0.041 | 0.018 | |
| $p_h - p_f$ | <0.001 | <0.001 | | <0.001 | <0.001 | | 0.020 | 0.011 | |

Confidence limits are ±2 s.e.m. as determined via $t$-statistics. Significance of comparisons are determined via 2-tailed paired $t$-statistics for the six subjects who participated in all studies, and via 2-tailed unpaired $t$-statistics for the full population of eight subjects. For $N_{max}$, statistics are determined after logarithmic transformation.

### 3.3. Symmetry and the vertical axis

Bilateral symmetry is most readily detected when the target is on the vertical axis, as shown in many previous studies (Dakin & Herbert, 1998; Rainville & Kingdom, 1999; Rainville & Kingdom, 2000) and also for targets of the kind used here (Conte & Victor, 2004). Since our stimuli consisted of four targets, with two on the vertical axis and two on the horizontal axis, it is possible that mixing "hard" and "easy" targets might have confounded the above analysis. With this in mind, we subdivided the symmetry data into trials in which the target was on the vertical axis, and trials in which the target was on the horizontal axis. Fig. 8 shows data from a representative subject with a large bias towards detection of symmetry targets on the vertical axis. The positional bias precludes a meaningful fit of the above model to these subsets of the data, since responses to targets on the horizontal axis are very nearly at chance. Nevertheless, it is clear from the similarity of performance for $6 \times 6$ and $16 \times 16$ arrays that there is virtually no pooling of statistics across arrays with larger numbers of checks. Thus, although anisotropies are readily detectable, they do not alter our basic conclusion that there is little if any pooling of statistics in the symmetry task.

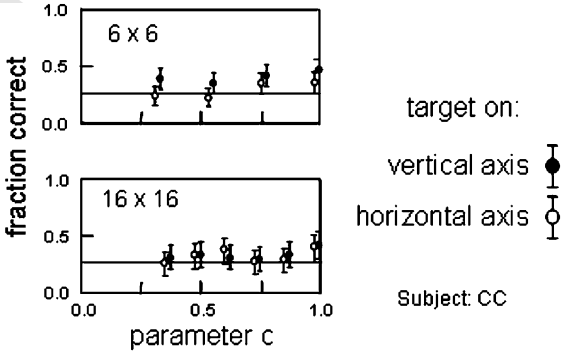


Fig. 8. Dependence of psychometric function for bilateral symmetry detection on position of the target, for two check sizes, in subject CC. Filled symbols: targets on the vertical axis. Open symbols: targets on the horizontal axis. Data for to horizontal-axis targets are displaced slightly along the absicssa to eliminate overlap. Plotting details otherwise as in Fig. 3.

## 4. Discussion

We used a stimulus set that allowed us to compare human subjects' ability to detect three kinds of deviation from randomness: local first-order statistics (luminance), local high-order statistics (an isodipole texture), and symmetry, which formally consists of second-order but nonlocal correlations. For each kind of statistical

deviation, a common parameter $c$ quantified the deviation of test stimuli from randomness. For all subjects, detection of first-order statistics occurred at the lowest values of $c$, followed by detection of local fourth-order statistics, followed by detection of symmetry, consistent with previous studies (Victor & Conte, 2004).

A common model framework, consisting of an initial stags of multiple independent detectors followed by a stage of spatial summation, provides a reasonable account of human subjects' performance on all three tasks. Such a model structure (Bergen & Adelson, 1988; Chubb & Landy, 1991; Graham, 1989; Graham et al., 1992; Grossberg & Mingolla, 1985; Malik & Perona, 1990; Victor & Conte, 1991; Wilson, 1993; Zhu et al., 1998) has been frequently used to interpret studies of image statistics, but is not typically applied to detection of approximate symmetry. For symmetry detection (see Fig. 6), the initial detectors of our model are independent and analyze only a restricted portion of the image, but, in contrast to the detectors of the typical "back pocket" model, they need not be adjacent.

Our intent is not to imply that symmetry is detected in by a parallel computation. Indeed, other work (Conte et al., 2002; Conte & Victor, 2003) strongly suggests that this is not the case, and psychophysical performance does not reach ceiling, so this range of performance is unavailable to test the model. Rather, the goal is to determine the extent to which two factors limit performance: the ability to identify individual statistical elements and the ability to pool statistical evidence over a wide area. We found that both factors contributed to the poorer performance on the symmetry task; local detection was modestly worse than for the isodipole task and substantially worse than for the luminance task; spatial pooling of symmetry signals was substantially worse than for both luminance and isodipole tasks. In contrast, spatial pooling for the luminance and isodipole tasks were comparable; differences in local detection were responsible for the performance differences between these tasks.

It is not surprising that detection of even vs. odd isodipole elements is less efficient than detection of white vs. black checks, since the latter is a nonlinear process that depends on the former. However, the finding that the spatial pooling of these signals is very similar could not have been predicted from such considerations: for both texture classes, each local element, once detected, represents an independent clue as to whether the image deviated from randomness. For isodipole textures, detection of statistical structure is less efficient (Joseph et al., 1997; Victor & Conte, 1989, 1991) for $64 \times 64$ than for $16 \times 16$ arrays. This supports the present inference that for isodipole statistics, efficient pooling is limited to 100–200 checks.

*Role of an "integration region" for symmetry*? One possible basis for the apparent inefficiency of pooling of symmetry signals could be that there is a specialized region near the vertical axis in which symmetry signals are efficiently combined (e.g., via a special attentional strategy, Wenderoth, 1994), and that outside of this "integration region"(Dakin & Herbert, 1998; Rainville & Kingdom, 2002), symmetry signals are effectively ignored. For unoriented narrowband stimuli presented at the fixation point (for a duration of 250 ms), this effective integration region subtends about 3 cycles along the horizontal axis, and 7 cycles along the vertical axis (Dakin & Herbert, 1998). In a study of anisotropic stimuli presented for 500 m (Rainville & Kingdom, 2000), a similar area was found for the integration region, but it was more elongated when the dominant orientation was vertical and less elongated when the dominant orientation was horizontal. However, these studies do not provide a direct estimate of the effective integration region for the current stimuli, since it is a not spatial frequency per se that sets the scale, but rather, the number of "objects", or micro-elements (Rainville & Kingdom, 2002).

The latter authors, using stimuli presented at fixation for 250 ms, estimated that the integration region subtends about 18 "objects". While this might at first seem to account for our findings that there is little improvement in performance as the array size grows from $6 \times 6$ checks to $16 \times 16$ checks (as in Figs. 3 and 4), there is an important aspect of our study that the "integration region" notion cannot explain. In contrast to the above studies and many others (Dakin & Herbert, 1998; Rainville & Kingdom, 1999, 2000, 2002; Tyler, 2001), the stimuli in this study were not presented at fixation. Indeed, based on the above studies, the estimated size and shape of the integration region (about 18 checks, elongated along the vertical axis) would not cover any of the targets. Thus, while the notion of an integration region whose size is scaled to the stimulus elements accounts for the lack of spatial pooling that we observed, this "integration region" cannot be fixed in location—to account for our data, it must be object-centered, and deployable to multiple targets within the 100 ms presentation period. Moreover (Fig. 8), it appears to be preferentially deployed along the vertical axis.

The distinctive behavior of symmetry supports the notion that symmetry detection does not share a computational substrate with detection of local statistical structure (Tyler, 2001). Identification of an individual symmetry element (match vs. mismatch) is less efficient than identification of an individual isodipole element, even though the former computation is formally simpler (second-order vs. fourth-order). Pooling of signals related to luminance or isodipole statistics can be accomplished over a hundred or more texture elements. In contrast, symmetry signals do not appear to be pooled, even though, for the stimuli used here, the symmetry axis was in a fixed location, so that the locations of

the pairs of checks to be compared remained constant. These differences are consistent with the notion that local image statistics are analyzed by a fixed computational network such as VI, but symmetry detection requires accumulation of evidence in a dynamic process (Conte et al., 2002; Wenderoth, 1994), and not just a parallel feedforward computation within a fixed integration region.

## Acknowledgments

Portions of this work were presented at the 2002 meeting of Vision Sciences Society, Sarasota, FL, and was supported by NIH NEI EY7977. We thank Caitlin Hardy for assistance with data collection and analysis, and Jeff Tsai for programming assistance.

## References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*, 183–193.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature*(333), 363–364.

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

Conte, M. M., Purpura, K. P., & Victor, J. D. (2002). *Processing of image symmetry in an RSVP task, Vol. 28*. Orlando, FL: Society for Neuroscience.

Conte, M. M., & Victor, J. D. (2003). *Temporal stability of image statistics in visual working memory*. Sarasota, FL: Vision Sciences Society.

Conte, M. M., & Victor, J. D. (2004). *Cueing rapidly deploys top-down influences in a mixed symmetry search task*. Sarasota, FL: Vision Sciences Society.

Dakin, S. C., & Herbert, A. M. (1998). The spatial region of integration for visual symmetry detection. *Proceedings of Royal Society of London Series B—Biological Sciences, 265*(1397), 659–664.

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

Graham, N., Beck, J., & Sutler, 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.

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

Joseph, J. S., Victor, J. D., & Optican, L. M. (1997). Scaling effects in the perception of higher-order spatial correlations. *Vision Research, 37*(22), 3097–3107.

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

Julesz, B., Gilbert, E. N., Shepp, L. A., & Frisch, H. L. (1973). Inability of humans to discriminate between visual textures that agree in second-order statistics—revisited. *Perception, 2*(4), 391–405.

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

Maddess, T., & Nagai, Y. (2001). Discriminating isotrigon textures. *Vision Research, 41*(28), 3837–3860.

Maddess, T., Nagai, Y., James, A. C., & Ankiewcz, A. (2004). Binary and ternary textures containing higher-order spatial correlations. *Vision Research, 44*(11), 1093–1113.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, 7*(5), 923–932.

Pollack, I. (1971). Perception of two-dimensional Markov constraints within visual displays. *Perception and Psychophysics, 9*, 461–464.

Purpura, K. P., Victor, J. D., & Katz, E. (1994). Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of National Academy of Sciences of the United States of America, 91*(18), 8482–8486.

Rainville, S. J., & Kingdom, F. A. (1999). Spatial-scale contribution to the detection of mirror symmetry in fractal noise. *Journal of Optical Society of America A—Optics Image Science and Vision, 16*(9), 2112–2123.

Rainville, S. J., & Kingdom, F. A. (2000). The functional role of oriented spatial filters in the perception of mirror symmetry—psychophysics and modeling. *Vision Research, 40*(19), 2621–2644.

Rainville, S. J., & Kingdom, F. A. (2002). Scale invariance is driven by stimulus density. *Vision Research, 42*(3), 351–367.

Sutter, A., Sperling, G., & Chubb, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research, 35*(7), 915–924.

Tyler, C. W. (1995). Empirical aspects of symmetry perception. *Spatial Vision, 9*(1), 1–7.

Tyler, C. W. (1999). Human symmetry detection exhibits reverse eccentricity scaling. *Visual Neuroscience, 16*(5), 919–922.

Tyler, C. W. (2001). The symmetry magnification function varies with detection task. *Journal of Vision, 1*(2), 137–144.

Victor, J. D. (1986). Isolation of components due to intracortical processing in the visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America, 83*(20), 7984–7988.

Victor, J. D., & Conte, M. M. (1989). Cortical interactions in texture processing: scale and dynamics. *Visual 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. (2001). Dynamics of selective spatial attention in a working memory task. *Association for Research in Vision and Ophthalmology, 42*, 863, Ft. Lauderdale.

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

Wenderoth, P. (1994). The salience of vertical symmetry. *Perception, 23*(2), 221–236.

Wilson, H. R. (1993). Nonlinear processes in visual pattern discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 90*(21), 9785–9790.

Zhu, S. C., Wu, Y., & Mumford, D. (1998). Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. *International Journal of Computer Vision, 27*(2), 107–126.