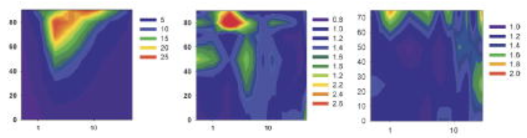
ELSEVIER

VISION
RESEARCH

An International Journal for Functional Aspects of Vision



Biochemistry & Cell Biology • Molecular Biology & Genetics
Anatomy, Physiology, Pathology & Pharmacology • Optics, Accommodation & Refractive Error
Circuitry & Pathways • Psychophysics • Perception • Attention & Cognition
Computational Vision • Eye Movements & Visuomotor Control

ISSN 0042-6989 | Volume 46 | Number 24 | November 2006

Available online at www.sciencedirect.com
ScienceDirect
Vision Research 46 (2006) 4152–4162
Vision Research
www.elsevier.com/locate/visres

Encoding and stability of image statistics in working memory

Jonathan D. Victor *, Mary M. Conte

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

Received 16 April 2006; received in revised form 20 July 2006

Abstract

Visual working memory contains a representation of certain image statistics (Victor & Conte, 2004), in addition to a pixel-by-pixel representation. Here, we show that the representation of statistics is more stable in time (up to 3000 ms) than the pixel-by-pixel representation, especially for changes in luminance and local high-order statistics, and is not affected by visual masking. Bilaterally symmetric arrays and arrays with local correlations are more readily encoded than random ones, but a change in the presence of bilateral symmetry, *per se*, contributes only modestly to the ability to detect that an array has changed.

Keywords: Symmetry; Isodipole; Working memory; Image statistics

1. Introduction

Images can be represented in visual working memory not only on a pixel-by-pixel basis, but also in a more abstract way, i.e., in terms of their statistical structure. In a paradigm based on random checkerboard arrays, Cornelissen and Greenlee (2000) showed that the pixel-by-pixel representation is indeed available to working memory, but that its capacity is quite limited. On intuitive grounds, one suspects that representation of statistical structure is more important than a pixel-by-pixel representation for certain tasks of mid-level vision, such as texture identification and surface analysis. Previously (Victor & Conte, 2004) we modified the Cornelissen and Greenlee (2000) paradigm by introducing controlled and graded statistical correlations (luminance, even–odd isodipole correlations, and bilateral symmetry) among the pixels. This allowed us to identify the role of these statistical elements in visual working memory, and to demonstrate that certain image statistics, namely, luminance and local high-order correlations, are represented. This representation is a graded one, and augments the pixel-by-pixel representation to improve performance on a visual working memory task.

Here, we further examine the role of image statistics in visual working memory, by studying the temporal stability of these two kinds of representations and dissociating the effects of image statistics on encoding and on representation. Images of objects have a complex statistical structure: well-characterized correlations at low orders (Field, 1987); and important, but less readily characterized, correlations at high orders (Franz & Scholkopf, 2005; Olshausen & Field, 1996; Zetzsche & Nuding, 2005); and symmetry, especially bilateral symmetry (Olivers & van der Helm, 1998; Tyler, 1995). A main motivation for the use of block arrays to study general mechanisms of visual working memory is that appropriate manipulation of such arrays can dissociate various aspects of the image to a much greater extent than may be possible with images of natural objects. This in turn can provide for a more fine-grained analysis of the contribution of image attributes to encoding and representation in visual working memory.

2. Methods

The experimental paradigm and visual stimuli were identical to those used in a previous study (Victor & Conte, 2004), and will therefore be described only briefly here. That paper also discusses our rationale for

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

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.07.023

choosing to study these particular image statistics (Fig. 1A), among the large range of other possibilities (Cho, Yang, & Hallett, 2000; Harvey & Gervais, 1981) that might be surveyed.

2.1. Subjects

Studies were conducted in 9 normal subjects (3 male, 6 female), ages 22–57. Other than author MC, subjects were naïve to the purpose of the experiments. Prior to data collection in Experiment IA (see below), subjects practiced the task until performance became stable. The four subjects who were experienced psychophysical observers in the previous study (Victor & Conte, 2004) or in a related task involving targets in the same positions relative to fixation (Victor & Conte, 2001) practiced for 0.75 h. The subject (CFC) who was an experienced psychophysical observer in unrelated studies practiced for 2.75 h. The four naïve subjects practiced for 4–6 h each. All had visual acuities (corrected if necessary) of 20/20 or better.

2.2. Stimuli

The stimulus frame S1 (Fig. 1B) consists of four arrays of checks on a mean gray background. The arrays were positioned along the cardinal axes, with centers 200 arc min from fixation. In most experiments, each array subtended 160 arc min and contained 64 (8×8) contiguous checks, each of which was either black or white and subtended 20 arc min. The stimulus frame S2 also consisted of four arrays, three of which were identical to those in the S1 frame of the trial. The target array, determined at random, differed from the corresponding array in S1 by a contrast inversion of 16 of the 64 checks. For each experiment, a particular kind of statistical structure was introduced into the arrays: luminance bias, high-order statistical structure (the "isodipole" textures), and symmetry. As detailed in Victor and Conte (2004), this was done in a manner that allowed the same number of contrast inversions to either induce a change in an image statistic ("change" trials), or to leave the image statistic unchanged ("no change" trials).

For each kind of image statistic, the strength of the statistical structure was parameterized by a quantity c. $c = 0$ denotes a maximally random assignment, and $c = 1$ (or $c = -1$) denotes a maximally structured assignment. For the luminance trials, c assumed the values of ± 0.25. $c = 0.25$ corresponded to an array that had 40 of its 64 checks white, and the remaining 24 checks black; for $c = -0.25$, the luminance bias was reversed. For the isodipole trials, arrays either had $c = 1$, corresponding to a maximally "even" texture (Julesz, Gilbert, & Victor, 1978), or $c = -1$, corresponding to a maximally "odd" texture. (Note that isodipole stimuli have long-range correlations induced by the short-range correlation rule, and for $c = \pm 1$, short- and long-range correlations are equally strong. However, as reviewed in Section 4.3, previous work indicates that the local correlations drive their visual salience.) In the symmetry trials, arrays either had $c = 1$, corresponding to a texture in which all pairs of checks that were related by the vertical symmetry axis were matched in luminance, or $c = 0$, corresponding to a texture in which half of such pairs were opposite in luminance.

For the isodipole and symmetry trials, there was no luminance bias. For luminance and even/odd statistics, the chosen values of c (in ratio 1:4) provide similar levels of salience, based on our previous studies(Victor, Chubb, & Conte, 2005; Victor & Conte, 2005). For bilateral symmetry, $c = 1$ is less readily detectable than $c = 1$ for even/odd statistics (Victor & Conte, 2005); however, a proportionately lower value of c for the luminance task (e.g., $c = 0.125$) would lead to a bias in the number of checks that would not significantly deviate from random coloring of an 8×8 array. Possible confounds related to the reduced salience of bilateral symmetry are considered in Section 4.

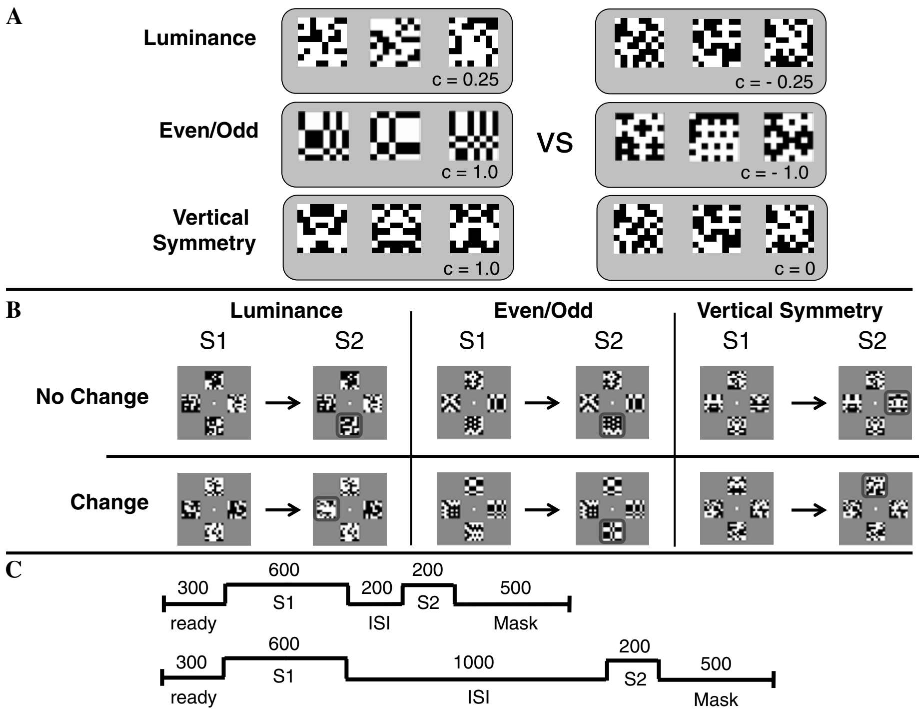


Fig. 1. (A) Examples of the three kinds of statistical manipulations used: luminance statistics, isodipole statistics, and vertical symmetry versus absence of symmetry. (B) Stimulus samples of typical trials. The subject's task is to determine which of the four arrays in S1 has changed in S2. For each kind of statistical structure, there were an equal number of trials in which the target did not change along the statistical axis of interest ("no change" trials), and trials in which it did change ("change" trials). For example, in the "change" luminance trial illustrated (lower left), the target was dark in S1 and bright in S2, and of the three distractors, two were bright and one was dark. Across trials, the target and distractor arrays were independently assigned to the statistical classes, so that all 16 possible assignments of the four arrays to bright and dark were present equally often for both "no change" and "change" conditions. (C) Trial timecourse.

2.3. Apparatus

The above visual stimuli were produced on a Sony Multiscan 17seII (17 in. diagonal) monitor, with signals driven by a PC-controlled Cambridge Research VSG2/5 graphics processor programmed in Delphi II to display bitmaps precomputed in Matlab. The resulting 768×1024 pixel display had a mean luminance of 47 cd/m^2, a refresh rate of 100 Hz and subtended 11×15 deg (approximately 1 arc min/pixel) at the viewing distance of 114 cm. The intensity versus 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

The design is identical to that of Victor and Conte (2004), which is a modification of the Cornelissen and Greenlee (2000) visual working memory task. Experiments consisted of a sequence of 4-alternative forced choice trials (Fig. 1B and C). 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 (S1, described in detail above) appeared, consisting of four arrays of checks, surrounding a central "X" subtending approximately 30 arc min. After presentation of S1 for 600 ms, the display returned to mean luminance for a variable retention interval (200, 1000, or 3000 ms), following which a second stimulus S2 (described above) appeared, containing a "target" that differed from the corresponding array in S1. After presentation of S2 for 200 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 S1 and S2. The subject's task was to identify the target array via a button-press on a response box with 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 S2) were collected via the Delphi II display software. Trials in which the subject responded before the onset of S2, or after 8000 ms, were discarded and repeated. In Experiment II, this procedure was modified by introducing a random checkerboard mask during the retention interval.

In Experiment IA, responses for each of the three kinds of statistical structure were collected in a single 4-block session, containing two interleaved 128-trial blocks for each retention interval (200 and 1000 ms). In Experiment IB, responses were collected in 6-block sessions, containing one block for each of two classes of statistical structure and the three retention intervals (200, 1000, and 3000 ms). The experiment consisted of 3 such sessions; in each session, two of the three classes of statistical structure were tested. This provided two 128-trial blocks for each condition, as in Experiment IA. Prior to each block, the subject completed 30 (Experiment IA) or 10 (Experiment IB) practice trials of the texture type and the retention interval of the block that would follow. Experiment IA consisted of a total of 13824 trials (128 trials per block $\times$ 4 blocks $\times$ 3 classes of structure $\times$ 9 subjects). Experiment IB consisted of a total of 9216 trials (128 trials per block $\times$ 6 blocks $\times$ 3 classes of structure $\times$ 4 subjects).

In Experiment II (masking), each class of statistical structure was tested in a separate session. There were four blocks of 256 trials: the baseline condition (identical to the 1000 ms retention interval of Experiment I), and three masking conditions. The mask consisted of a random checkerboard identical to the mask described above that followed S2. The masking conditions were: (i) mask presented in the final 800 ms of the retention interval, (ii) mask presented in the first 800 ms of the retention interval, and (iii) mask presented for the entire retention interval. In all cases, a second 500 ms mask followed S2, as in Experiment I. Experiment II consisted of a total of 12288 trials (256 trials per block $\times$ 4 blocks $\times$ 3 classes of structure $\times$ 4 subjects).

In all experiments, order of sessions, and of blocks within sessions, was randomized and counterbalanced across subjects. Experiment IA was analyzed as a function of block for evidence of learning, and none was found: there was an overall improvement in fraction correct of 0.011, and no significant difference for any of the six (three kinds of statistical classes, two interstimulus intervals) kinds of blocks ($p > 0.2$, one-tailed paired t-test).

Initial data analysis was performed in EXCEL. Analysis of variance was performed in SPSS.

3. Results

3.1. Experiment I: Time course of representation of image statistics

In Experiment I, we determined how the subject's ability to identify a change in the elements of an array depended on whether this change was associated with a change in image statistics, and how performance depended on retention interval. Three kinds of statistics were studied: the overall number of black and white checks (first-order, or luminance statistics), their local correlation (even/odd isodipole statistics), or the presence or absence of bilateral symmetry.

In Experiment IA, nine subjects were studied at retention intervals of 200 and 1000 ms (Fig. 2A and B). Experiment IB, carried out 9–16 months later in a subset of 4 subjects, examined retention intervals of 200, 1000, and 3000 ms (Fig. 3A and B).

3.1.1. Analysis of variance

Analysis of variance (Table 1) was carried out separately for each experiment, with fraction correct (FC) and reaction time (RT) as dependent variables. The analysis was organized along four factors: two levels of CHANGE ("no change" vs. "change" trials), two levels (Experiment 1A) or three levels (Experiment 1B) of ISI (the retention intervals), three levels of CLASS (luminance vs. isodipole vs. symmetry), and subject.

For all analyses, most of the total FC variance could be ascribed to within-subject differences, while most of the total RT variance could be ascribed to between-subject differences (top section of Table 1). Overall RT's ranged from a mean of 507 ms (subject CC) to 1348 ms (subject SD), while overall FC's varied ranged from a mean of 0.461 (subject JR) to 0.752 (subject CC). The intrasubject variability of RT did not appear to be related to levels of performance: the cross-subject correlation between mean FC and mean RT, though negative, was not significant ($r = -0.44$, $p > 0.2$ via Fisher z-transform with $N = 9$ in Experiment IA).

We focus on findings that were consistent across subject, and later summarize the interaction of these factors with SUBJECT. For the two dependent variables (FC and RT), there is a similar apportionment of the within-subject variance attributable to these factors and their interactions (the within-subject analysis in the lower portion of Table 1). This similarity, along with the lack of correlation of RT and FC across subjects, is consistent with the notion that RT has variability related to its motor component that is not influenced by the perceptual aspects of the task.

All within-subject main effects (CHANGE, ISI, CLASS) were highly significant ($p < 0.001$) for both dependent

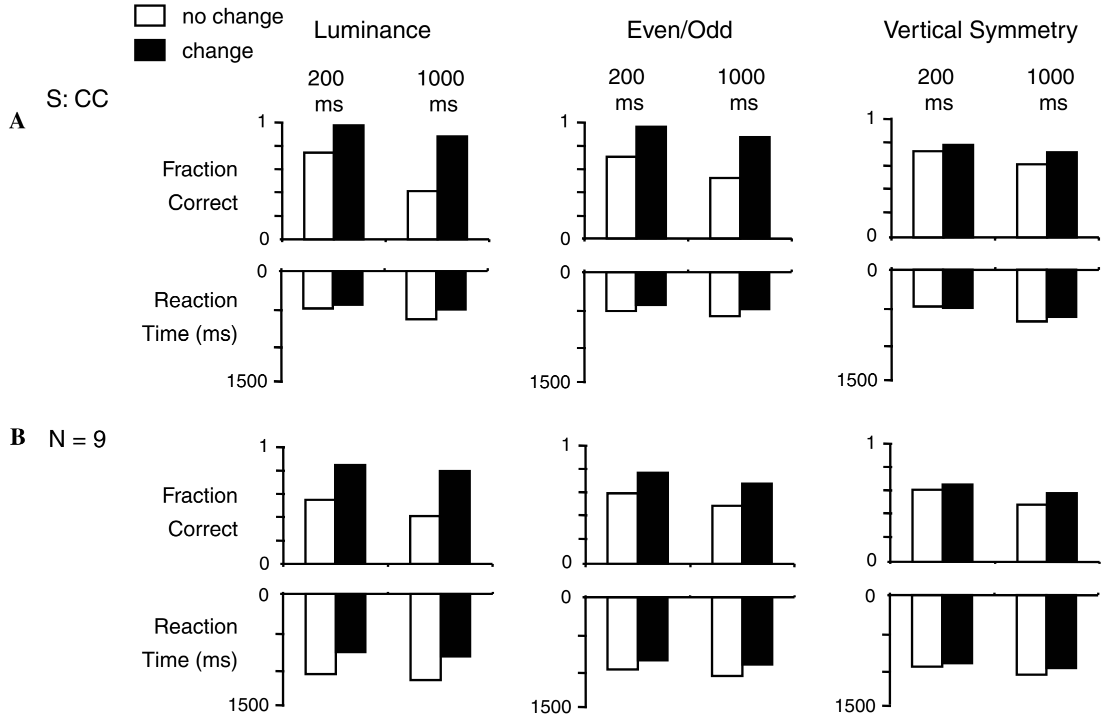


Fig. 2. The effect of statistical change on performance in a visual working memory task. (A) Results from a single subject (CC) in experiment IA (200 and 1000 ms retention intervals). (B) Average performance of 9 subjects in experiment IA. Note that in all cases, fraction correct for trials in which there was a change in statistics (solid) was higher than for trials in which there was no such change (open), and reaction times were shorter.

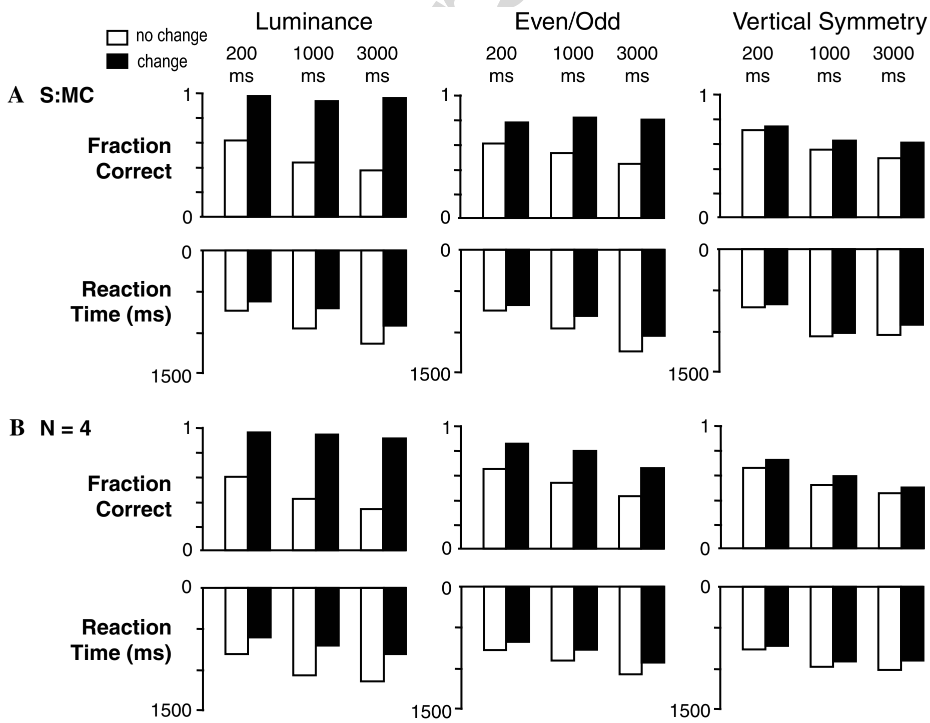


Fig. 3. (A) Results from a single subject (MC) in experiment IB (200, 1000, and 3000 ms retention intervals). (B) Average performance of 4 subjects in Experiment IB. The difference between the "change" and "no change" conditions increased with ISI; most notably for a change in luminance statistics.

variables. The greatest amount of variance was attributable to presence or absence of statistical change, then retention interval, then statistical class.

The largest two-way interaction was between CLASS and CHANGE, indicating that for some of the classes, statistics contribute significantly ($p < 0.001$) more than others to

Table 1
Summary of ANOVA for Experiment I

		Expt. 1A				Expt. 1B			
		All classes		Lum & even/odd		All classes		Lum & even/odd	
		FC	RT	FC	RT	FC	RT	FC	RT
Fraction of total variance									
All $p < 0.001$	Between subject	0.424	0.875	0.378	0.844	0.092	0.602	0.071	0.554
	Within subject	0.576	0.125	0.622	0.156	0.908	0.398	0.929	0.446
Fraction of within-subject variance									
	Main effects								
* $p < 0.05$	Change	0.598‡	0.605‡	0.800‡	0.792‡	0.491‡	0.315‡	0.738‡	0.423‡
† $p < 0.01$	ISI	0.147‡	0.129‡	0.110‡	0.085‡	0.183‡	0.517‡	0.128‡	0.465‡
‡ $p < 0.001$	Class	0.049‡	0.048‡	0.002	0.000	0.077‡	0.008‡	0.010‡	0.000
	Two-way interactions								
	Class × Change	0.191‡	0.210‡	0.075‡	0.118‡	0.222‡	0.106‡	0.094‡	0.074‡
	ISI × Change	0.011‡	0.007*	0.008†	0.005	0.008†	0.017‡	0.016	0.020‡
	Class × ISI	0.000	0.002	0.000	0.000	0.006*	0.029‡	0.005*	0.011‡
	Three-way interaction								
	Class × ISI × Change	0.004	0.000	0.005*	0.000	0.013‡	0.009†	0.009†	0.008†

The dependent variables, fraction correct (FC) and reaction time (RT), were analyzed separately. In Experiment IA, there were two levels of CHANGE ("no change" vs. "change" trials), two levels of ISI (the retention intervals, 200 and 1000 ms), three levels of CLASS (luminance vs. isodipole vs. symmetry), and nine subjects. In Experiment IB, there were three levels of ISI (200, 1000, and 3000 ms), and four subjects. The upper portion of the table indicates the total fraction of variance of each dependent variable ascribed to all between-subject effects and all within-subject effects. The lower portion of the table subdivides the within-subject variance. The fraction of the total variance attributable to each of these effects can be determined by multiplying the tabulated value by the corresponding "within-subjects" entry. Significance levels were determined by an *F*-test.

performance. It appears from Fig. 2 and 3 that most of the main effect of CLASS and the CLASS x CHANGE interaction is due to the bilateral symmetry condition. An ANOVA restricted to luminance and even/odd classes demonstrated that this was the case (Table 1). For this restricted ANOVA, the main effect of CLASS was not significant for either FC or RT in Experiment IA. In Experiment IB, it did not contribute significantly to RT, and contributed a markedly reduced, though still significant, fraction of the variance to FC ($p < 0.001$). However, the CLASS × CHANGE interaction remains significant ($p < 0.001$) in the two-class analysis for both dependent variables, indicating that luminance contributes more to task performance than even/odd statistics, which in turn contributes more than symmetry (Figs. 2 and 3).

There was also a significant interaction between retention interval (ISI) and CHANGE. For two retention intervals (200 and 1000 ms, Experiment IA) this interaction is seen primarily for FC ($p < 0.001$ for the two-class analysis, $p < 0.01$ for the three-class analysis), but when retention intervals up to 3000 ms are analyzed (Experiment IB), similar-size effects are seen for RT as well. As seen in Figs. 2 and 3, the direction of this interaction is that a change in image statistics confers a progressively greater benefit on performance for longer ISI's than for shorter ISI's. That is, the representation of image statistics is more stable than that of the raw pixels.

The interaction between CLASS and ISI is not significant in Experiment IA, and contributes only a small portion of the variance ($p < 0.05$ for FC, $p < 0.001$ for RT) in Experiment IB. This small interaction reflects a greater falloff in performance for the even/odd stimuli than for the luminance stimuli at 3000 ms, independent of whether there is a change in statistics.

The three-way interaction, CLASS × ISI × CHANGE is significant at $p < 0.01$ or better for FC and RT in Experiment IB. As seen from the performance data of Fig. 3, this reflects the greater influence of a change in luminance statistics than even/odd statistics on performance, and the increasing magnitude of this difference as ISI increases from 200 to 3000 ms. When only an ISI of 200 and 1000 ms are considered (Experiment IA, Fig. 2), a trend in the same direction is seen, but this is only minimally significant ($p < 0.05$ in the two-CLASS analysis only).

In sum, performance (as measured by FC and RT) in this visual memory task is enhanced in trials in which the luminance statistics or the local high-order local statistics change, compared to trials in which the same number of checks change in luminance but do not result in an overall change in statistics. Moreover, performance falls off more rapidly in time for trials in which there is no change in statistics, than for trials in which the target's identity is also cued by a statistical change. This suggests that these image statistics are represented in visual working memory, and the representation of these statistics is more stable than the representation of images as individual pixels.

While there was a decline in performance between 1000 and 3000 ms (somewhat more prominent for the group mean than for the individual shown, Fig. 3), performance remained well above chance (25% correct), and the decline was less than the decline in performance from 200 to 1000 ms. Thus, local target statistics are represented in visual memory for at least 3 s, and, over this time period, the representation is more stable than expected from an exponential decline.

3.1.2. Intersubject variability

As seen in Table 1, most of the variance for RT, and a substantial portion of the variance for FC, was attributable to intersubject variability (a main effect of SUBJECT, and its interactions with the other three factors). The main effect of SUBJECT was significant ($p < 0.001$) in all analyses, as was its interaction with the factors that had significant main effects in Table 1. SUBJECT also interacted significantly ($p < 0.01$ or $p < 0.001$) with the more significant pairwise interactions of Table 1. These interactions reflected intersubject variability in the degree of the interactions described above, not in their direction. The four-way interaction CLASS × ISI × CHANGE × SUBJECT was not significant.

3.1.3. Influence of statistical structure during encoding

As described above, a change in luminance statistics had a greater and more persistent influence on performance than a change in even/odd statistics, and we found little if any effect of a change in bilateral symmetry (presence vs. absence). This latter finding is consistent with previous studies (Victor & Conte, 2004) at the 200 ms retention interval. It suggests that this aspect of an image is not used during encoding during the initial presentation, or if used for encoding, is not retained in visual working memory.

To make this distinction, we asked whether the presence of statistical structure during encoding (rather than a change between a random and a structured texture between encoding and recall) influenced performance. That is, we partitioned the "no change" conditions according to the presence of statistical structure of the target (which was necessarily the same in S1 and S2). In the symmetry blocks, an ANOVA revealed a highly significant main effect of the presence of bilateral symmetry: fraction correct of 0.51 for random stimuli, fraction correct of 0.58 for symmetric stimuli ($p < 0.002$). This effect was present for all three retention intervals; there was no interaction with ISI. Similarly, reaction time was shorter (1105 ms vs. 1153 ms, $p \approx 0.07$) for symmetric than for random targets, and there was no interaction of this with ISI.

Thus, even though a change in statistical structure from "symmetric" to "random" (or vice versa) did not influence performance (the lack of a main effect of CHANGE in Table 1), the analysis of the "no change" trials indicates that some symmetry-specific processing must have taken place, since changes between S1 and S2 were more readily detected in symmetric targets.

It might be argued that improved performance on symmetric targets is to be expected, merely because only half of the stimulus needs to be remembered. This hypothesis can be tested by examination of performance on the other textures. For the isodipole textures, only one row and one column of the textures needs to be remembered, since the interior of the texture is determined by propagation of a parity rule. That is, for these textures, it suffices to remember only one-quarter of the checks. This is equally true for the "even" and for the "odd" textures. However, performance for isodipole textures in the "no change" trials depended strongly on the spatial organization of the correlation—i.e., the even vs. odd condition—and not just the number of checks that needed to be remembered (which was the same for even and odd stimuli). Fraction correct was 0.60 for even textures, and 0.48 for odd textures ($p < 0.001$), and reaction time was 1101 ms for even textures, and 1171 ms for odd textures ($p \approx 0.01$), with no interaction with ISI. For luminance textures, there was no significant difference between the bright ($c = 0.25$) and the dark ($c = -0.25$) checks; fraction correct was 0.48 and mean reaction time was 1154 ms. This is comparable to performance on the "random" trials of the symmetry condition and also to performance on the "odd" trials of the isodipole condition—even though the latter stimuli contained only 15 non-redundant checks.

In sum, changes in some kinds of targets (symmetric targets and targets with "even" structure) are more readily identified than changes in other kinds of targets (asymmetric targets, and targets with independently distributed pixel values, with either an unbiased or modestly biased distribution). These differences must relate to the spatial organization of the targets, rather than the set size or the number of non-redundant checks. Moreover, although changes in symmetric targets are demonstrably easier to recognize than changes in random ones, a change in whether symmetry is present does not seem to be an available cue for the working memory task.

3.2. Experiment II: Masking

Experiment I showed that the memory trace of the individual pixel values decayed faster than that of the statistical category. We therefore attempted to use a masking paradigm to analyze this difference in time course in more detail for the 1000 ms ISI condition of Experiment I. The mask was a random checkerboard and was presented during the final 800 ms of the ISI ("late mask"), during the initial 800 ms of the ISI ("early mask"), during the entire 1000 ms ISI ("full mask"), or omitted as in Experiment I.

As in Experiment I, most of the total FC variance could be ascribed to within-subject differences, while most of the total RT variance could be ascribed to between-subject differences. For both dependent variables (Table 2), there were large and highly significant main effects of statistical change and stimulus class, and their interaction ($p < 0.001$). Additionally, the presence of a mask significantly ($p < 0.001$) reduced the fraction correct (Table 2 and Fig. 4). However, there was no interaction between masking and the presence or absence of statistical change in terms of FC ($p > 0.05$), and only a minimal interaction in terms of RT ($p < 0.05$). Thus, this masking paradigm failed to identify a difference between the timecourse of processing or retrieval of a pixel-based memory and one based on image statistics.

There was no significant difference between the effect of early or late masking on FC (last two columns of Table 2) but a highly significant effect of this manipulation on RT.

Table 2
Summary of ANOVA for Experiment II

		Expt. II					
		All masking conditions		No mask vs. Full mask		Early vs. Late mask	
		FC	RT	FC	RT	FC	RT
Fraction of total variance							
All $p < 0.001$	Between subject	0.111	0.743	0.101	0.723	0.122	0.710
	Within subject	0.889	0.257	0.899	0.277	0.878	0.290
Fraction of within-subject variance							
	Main effects						
$^{*}p < 0.05$	Change	$0.627^{‡}$	$0.663^{‡}$	$0.526^{‡}$	$0.633^{‡}$	$0.673^{‡}$	$0.535^{‡}$
$^{†}p < 0.01$	Mask	$0.019^{‡}$	0.016‡	$0.031^{‡}$	$0.095^{‡}$	0.001	$0.174^{‡}$
$^{‡}p < 0.001$	Class	$0.087^{‡}$	$0.083^{‡}$	$0.093^{‡}$	$0.049^{‡}$	$0.072^{‡}$	$0.101^{†}$
	Two-way interactions						
	Class × Change	$0.261^{‡}$	$0.207^{‡}$	$0.246^{‡}$	$0.197^{‡}$	$0.250^{‡}$	$0.167^{‡}$
	Mask × Change	0.000	0.010^{*}	0.000	0.006^{*}	0.000	0.010^{*}
	Class × Mask	0.004	$0.017^{†}$	0.002	$0.016^{†}$	0.005	0.010^{*}
	Three-way interaction						
	Class × Mask × Change	0.001	0.004	0.001	0.004	0.000	0.004

There were two levels of CHANGE ("no change" vs. "change" trials), four levels of MASK (no mask, late mask, early mask, full mask), three levels of CLASS (luminance vs. isodipole vs. symmetry), and four subjects. Other details as in Table 1.

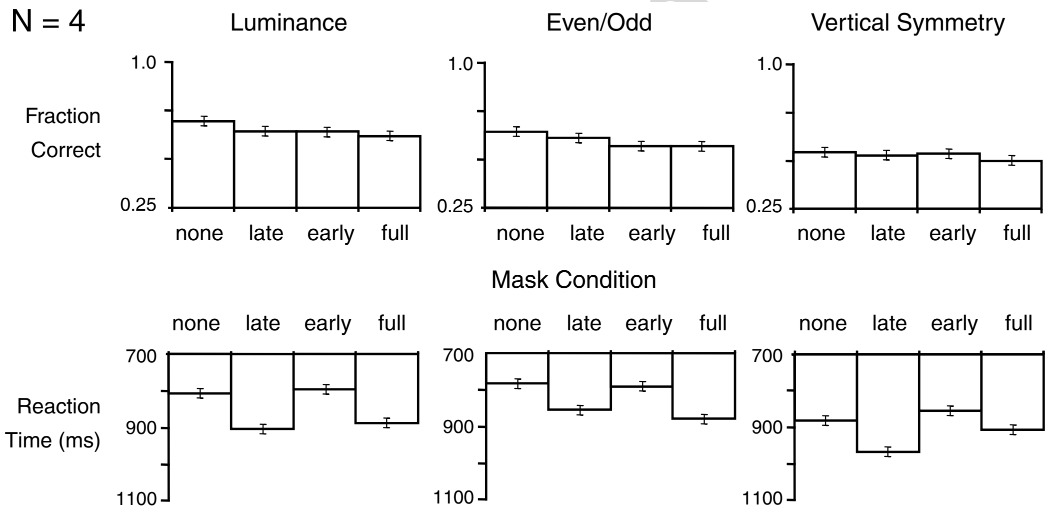


Fig. 4. Effect of masking on performance in a visual working memory task, averaged across four subjects (Experiment II). Masking conditions during the 1000 ms ISI were: no mask, late mask (last 800 ms), early mask (first 800 ms), and full mask. Results are pooled across the presence or absence of statistical change (which did not interact with the presence of the mask, see Table 2), to highlight the effect of the mask.

Most likely this is an effect of the mask on the execution of the motor response, not on visual processing.

4. Discussion

4.1. Analysis of the task

As reviewed by Cornelissen and Greenlee (2000), identification of a change in an array of checks is an often used paradigm for the study of visual working memory (Avons & Phillips, 1987; Inui, 1988; Irwin, 1991; Phillips, 1974). To put our study into context, it is helpful to first consider in some detail the requirements of this task. The main point of the analysis that follows is that performance depends on not only memory capacity, but also the manner in which the stimuli are encoded. Secondarily, the analysis will show that reliance on the behavior of an "ideal observer" limited by memory capacity (but not encoding) leads to seeming paradoxes—from which we conclude that limitations in encoding play a crucial role in this kind of task.

Clearly, if an observer knew that the mean luminance of the target array would change, then storage of the mean luminance of each array would suffice for perfect performance on the memory task. An ideal observer could therefore achieve perfect performance with a memory that represented only the mean luminance of each array. This representation would require only a few bits, depending on the *a priori* distribution of mean luminances. Thus, general knowledge about a task can reduce the memory capacity required to perform it.

Somewhat less obviously, encoding schemes of a different sort can also be used to dramatically reduce the memory requirements in the change-detection task, even for random arrays when no statistical cue (such as a change in mean luminance) is present. Consider a task in which k random arrays, each with N checks, are presented in S1, and only one check (in only one array) changes in S2. The observer is asked to detect which array was changed between S1 and S2. A naïve view of this task is that perfect performance would require representation of all kN checks in working memory (i.e., a working memory capacity of kN bits, if each check has an equal chance of being black or white). A slightly more clever observer could also achieve perfect performance by representing all but one of the arrays. If no change was detected in the attended arrays, then the observer could reliably conclude that a change occurred in the unattended array.

The observer's response (one of k equally likely choices) carries only $\log_2 k$ bits, much less than the $(k-1)N$ bits required to represent the checks in all but one of the arrays. This suggests that with a different encoding strategy, the task could be accomplished by holding fewer than $(k-1)N$ bits in memory. For example, the observer need not represent each array, but only whether each array contained an even or an odd number of bright checks. The encoding strategy is complex, but only k bits (one bit per array) need to be held in memory. Since any one-check change would change the parity of its array, this strategy would suffice for perfect performance. Extensions of this strategy (encoding the total parity across several arrays) reduce the memory requirement from k bits to $\log_2 k$ bits—the strategy of an ideal observer with limited memory capacity. The parity strategy also will support perfect performance on a related task, in which an observer is required to determine whether a single array does, or does not, have a single-pixel change. These encoding strategies are highly implausible, but provide simple illustrations that encoding strategies can drastically reduce memory requirements. A similar conclusion holds for more plausible encoding strategies (e.g., representing whether chunks of the image are uniform in contrast), but the analysis of these "simpler" strategies is substantially more complex.

Now consider a related task in which two or more checks change. Intuitively and in practice, this makes the task easier. However, it does not make the task easier for the "ideal observer" strategy described above: the parity strategy will fail if an even number of checks are changed. The resolution of this seeming paradox is straightforward—a human observer's performance is so far from that of an ideal observer (because of the latter's available encoding strategies) that comparison to the ideal observer performance is not helpful in understanding task performance. Conversely, this demonstrates that even in an observer with a tiny memory capacity (e.g., 2 bits, for the present $k=4$), the encoding process may be the limiting factor—since, in the absence of limitations on encoding strategies, 2 bits would suffice to detect which of four arrays had a single altered pixel. Encoding limitations are especially important considering that in our task, we give observers only a limited time (600 ms) to view the S1 component of each trial.

Here and in a previous study (Victor & Conte, 2004), we modified the Cornelissen and Greenlee paradigm so that checks were colored black or white not according to an independent, equal-probability rule, but according to rules that introduced bias at each check independently ("luminance") or correlation between checks ("isodipole" and "symmetry"). Our main goal was to determine whether these statistical aspects were represented in working memory—i.e., whether a change in the luminance bias, or a change in correlation structure, provided a cue that an array had changed. But clearly, the statistical structure we introduce may also affect the encoding process (Inui, 1988). To dissociate the influence of statistical structure on encoding from its possible role in working memory, we used a counterbalancing scheme: in half of the trials, statistical structure of the target was changed between S1 and S2 and in half of the trials, it did not—but this factor was independent of the kind of structure that was present in S1. Thus, encoding demands were the same for the trials in which there was a change in statistical structure, and trials in which there was no such change. Moreover, our finding of interactions between a change in statistical structure and retention interval (ISI × CHANGE, Table 1) is difficult to attribute to the encoding process.

4.2. A distinct representation of image statistics

It is well known that fundamental visual processes, such as segmentation (Julesz, 1981; Julesz et al., 1978) and motion extraction (Chubb & Sperling, 1988), can be driven not only by differences in luminance, but also by differences in local correlations. However, the manner in which the visual system uses correlations is as yet unclear (Tyler, 2004b). Correlations, by definition, are a statistical entity. One possibility is that visual inferences are only based on the individual images (Yellott, 1993), although this inference may make use of a statistical analysis within the individual texture sample (Tyler, 2004b). Alternatively, visual inferences may at least in part make use of learned or intrinsic statistical descriptors that apply to ensembles, rather than individual textures (Victor, 1994), and can thus be harnessed in a visual memory task.

Our approach to addressing this issue builds on the approach of Cornelissen and Greenlee (2000). These authors asked subjects to determine whether two checkerboard arrays, presented sequentially, were different. Arrays differed in a small number (adjusted to be close to threshold) of checks, and each array was spatially random. Here, however, the arrays had spatial correlations. The pixel-by-pixel changes could then be made in a manner that either preserved, or altered, the degree of spatial correlation (Fig. 1). If working memory only contained a pixel-by-pixel representation, then the presence or absence of a change in spatial correlation would not influence performance– since exactly the same number of checks changed state on each trial. Instead, we found (Figs. 2 and 3) that a change in overall correlation structure was an important cue, both increasing fraction correct and decreasing reaction time. Moreover, especially for the luminance cue, the advantage conferred by the statistical change increased at the longer retention intervals (Figs. 2 and 3, and significant interaction between ISI and CHANGE, Table 1). Thus, we conclude that image statistics are indeed encoded and represented in working memory, and this representation is more stable than that of the individual pixels.

Phillips and others identified a high-capacity "sensory storage", distinct from visual working memory, that is highly susceptible to masking and tied to retinal position (Irwin, 1991; Phillips, 1974). This sensory storage appears to have a duration of much less than 300 ms. The pixel-by-pixel and statistical representations considered here last substantially more than 300 ms, and the effect of masking on both representations was minor. These considerations indicate that the pixel-by-pixel representation, as well as the representation of image statistics, are components of working memory, not sensory storage.

We analyze our results in terms of a comparison between trials with a statistical change and trials without a statistical change, rather than via comparison with an "ideal observer", because of the major qualitative differences between our findings and those expected from an ideal observer with unlimited encoding capacity. Our results also cannot be accounted for simply by pixel-by-pixel encoding and a specific capacity limitation. Verghese and Pelli (1992) estimate visual working memory capacity to be approximately 30 bits. If these bits are devoted to representing ten checks in each of three arrays, predicted performance on our task (based on detection of any of the 16 changed checks in the attended arrays) would be greater than 96% correct, far in excess of what we observe. If more severe capacity limitations in a pixel-by-pixel representation (Inui, 1988) were the primary determinant of performance, performance would be best in the even/odd (isodipole) trials, next-best in the symmetry trials, and worst in the luminance trials, based on the number of checks required to specify a stimulus. This "set size" effect is very large: if capacity limitations accounted for performance in the even/odd trials, then performance in the other kinds of trials should be near chance, since there are 2^{64} unique stimuli in the luminance trials, 2^{32} unique symmetric stimuli, but only 2^{16} unique stimuli in the isodipole trials. These predictions are in contrast to our observations of generally similar performance levels, above chance but not near ceiling, in all kinds of trials. Finally, discrimination of luminance and isodipole statistics by the human observer has absolute efficiencies of approximately 4% and 0.3%, respectively (Victor et al., 2005). These factors combine to indicate that task performance reflects limitations both on encoding and on capacity.

4.3. Local and long-range correlations

We use the terms "short-range" (or equivalently, "local") vs. "long-range" to describe the relationship of the length scale of correlations to the size of the largest units that can resolve individual texture elements. Note that this is a relative, rather than an absolute, length scale. This usage is in accord with observations that texture processing depends on this relative scale, and is approximately independent of the absolute size of the texture element, over several octaves' range (Joseph, Victor, & Optican, 1997; Sutter, Sperling, & Chubb, 1995; Victor & Conte, 1989).

For the textures used in the luminance trials, correlations are entirely within checks: there are no correlations between any two checks. The even/odd textures, however, have both short- and long-range correlations. They are defined (Julesz et al., 1978) by specifying the probability p that 2×2 blocks have an even number of bright checks (a local rule), but iteration of this rule across the texture induces long-range correlations. For a given local correlation strength $c = 2p - 1$, long-range correlations among four checks at the corners of an $(I+1) \times (J+1)$ rectangle are given by c^{IJ} (Victor & Conte, 1989). Thus, for the maximally structured textures ($c = \pm 1$) used here, the long-range correlations are as strong as the short-range correlations. However, for partially correlated textures ($|c| < 1$), short- and long-range correlations can be dissociated. Additionally, short- and long-range correlations can be dissociated using a variety of other techniques (Tyler, 2004a).

For partially correlated even/odd textures, $c = 0.25$ is above threshold for detection of structure (Victor & Conte, 2005). A change in statistics between $c = -0.5$ and $c = +0.5$ improves performance on the present working memory task (Victor & Conte, 2004). These findings, along with other VEP and psychophysical evidence derived from partially correlated even/odd textures (Joseph et al., 1997; Victor & Conte, 1989, 1991, 2005), indicates that the short-range correlations within the even/odd textures drives their perceptual salience and influences performance on visual working memory, even though long-range correlations are present.

4.4. Bilateral symmetry

Bilateral symmetry is visually salient and ethologically important (Attneave, 1954; Baylis & Driver, 1994; Olivers & van der Helm, 1998; Tyler, 1995; Wenderoth, 1994). Thus, the very modest effect of symmetry, here and in a previous study [200 ms retention intervals only, (Victor & Conte, 2004)] was surprising. One possibility is that symmetry is not fully processed under the conditions used (120–280 arc min from fixation, 600 ms duration). Previous studies have shown that symmetry processing is impaired off the vertical axis (Dakin & Herbert, 1998; Rainville & Kingdom, 1999, 2000, 2002; Tyler, 2001), and that discrimination of symmetry is not well described by a simple parallel feedforward model (Victor & Conte, 2005), especially when multiple targets are present (Olivers & van der Helm, 1998). In a symmetry detection task with the spatial conditions used here (Conte, Purpura, & Victor, 2002), fraction correct in a 4-AFC task at 400 ms is 0.5–0.55; perfect performance would be achieved on that task with unlimited free viewing time. The limitations are likely temporal more than spatial, in that there was no difference between peripheral and central RSVP presentations. Since bilateral symmetry was less readily detectable than the other image statistics, it is possible that stimuli with multiple axes (4- or 8-fold) of symmetry might have yielded results more comparable to luminance and even/odd image statistics.

However, our findings nevertheless imply an impoverished representation of bilateral symmetry in visual memory, rather than that symmetry was not detected, or did not influence encoding. Although a switch from symmetry to lack of symmetry (or vice versa) does not help to signal that an array has changed, observers are more likely to detect a change in a bilaterally symmetric target than a random one (0.58 vs. 0.51, $p < 0.002$). This indicates that bilateral symmetry was detected at some stage of the visual system, because it evidently played a role in encoding. This is not simply a generic set size effect—e.g., due to the fact that for a symmetric target, only half of the image needs to be represented. Were that the case, performance on the isodipole textures would have been much higher (since only the first row and first column would need to be represented, along with one more bit to identify the sign of c), and there would have been no difference between performance on "even" and "odd" stimuli—both counter to our observations. Thus, certain specific kinds of structure (overall fraction of bright checks, bilateral symmetry, and the "even" fourth-order correlation, but not the "odd" fourth-order correlation) facilitate performance, most likely because their presence aids encoding. This facilitation is separate from the representation of statistical structure once it is encoded (which appears to be present for local correlations but absent for bilateral symmetry), and highlights the role of sensory factors in working memory tasks (Pasternak & Greenlee, 2005).

5. Conclusion

We examined visual working memory for binary arrays of checks, and found evidence for a pixel-by-pixel representation and a representation of certain kinds of image statistics, with different properties. The pixel-by-pixel representation is labile, as demonstrated by a reduction in performance over longer retention intervals (1000 and 3000 ms) when no statistical cues are available. A change in luminance or high-order local correlations augments performance, indicating that these local image statistics are represented in working memory. Their role is increasingly important at the longer retention intervals, and thus the representation of these image statistics appears to be more stable than the pixel-by-pixel representation. As in the previous study (Victor & Conte, 2004), a change in overall bilateral symmetry contributes very little to performance. However, one cannot conclude that symmetry is irrelevant to the task. Rather, changes in symmetric targets are more readily detected than changes in asymmetric targets, implying symmetry-specific mechanisms for encoding into working memory—even though a change between symmetry and asymmetry, *per se*, does not contribute to performance.

Acknowledgments

A portion of this work has been presented at the 2003 meeting of the Vision Sciences Society (Conte & Victor, 2003), Sarasota, FL. The work was supported by NIH NEI EY7977. We thank Jeff Tsai for programming assistance and Ana Ashurova for assistance with preparation of the manuscript.

References

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

Avons, S. E., & Phillips, W. A. (1987). Representation of matrix patterns in long- and short-term visual memory. *Acta Psychology (Amsterdam), 65*(3), 227–246.

Baylis, G., & Driver, J. (1994). Parallel computation of symmetry but not repetition within single visual shapes. *Visual Cognition, 1*, 377–400.

Cho, R. Y., Yang, V., & Hallett, P. E. (2000). Reliability and dimensionality of judgments of visually textured materials. *Perception & Psychophysics, 62*(4), 735–752.

Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America [A], 5*(11), 1986–2007.

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

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

Cornelissen, F. W., & Greenlee, M. W. (2000). Visual memory for random block patterns defined by luminance and color contrast. *Vision Research, 40*(3), 287–299.

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

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America [A], 4*(12), 2379–2394.

Franz, M., & Scholkopf, B. (2005). Implicit Wiener series for higher-order image analysis. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems* (vol. 17, pp. 465–472). Cambridge, MA: MIT Press.

Harvey, L. O., Jr., & Gervais, M. J. (1981). Internal representation of visual texture as the basis for the judgment of similarity. *Journal of Experimental Psychology: Human Perception and Performance, 7*(4), 741–753.

Inui, T. (1988). Properties of human visual memory for block patterns. *Biology Cybernetics, 59*(3), 179–187.

Irwin, D. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*, 420–456.

Joseph, J. J., Victor, J. D., & Optican, L. M. (1997). Scaling effects in the perception of higher-order spatial correlations. *Vision Research, 37*(22), 3107–3907.

Julesz, B. (1981). A theory of preattentive texture discrimination based on first-order statistics of texture. *Biological Cybernetics*(41), 131–138.

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

Olivers, C. N., & van der Helm, P. A. (1998). Symmetry and selective attention: a dissociation between effortless perception and serial search. *Perception & Psychophysics, 60*(7), 1101–1116.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*(6583), 607–609.

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews. Neuroscience, 6*(2), 97–107.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283–290.

Rainville, S. J., & Kingdom, F. A. (1999). Spatial-scale contribution to the detection of mirror symmetry in fractal noise. *Journal of the Optical Society of America A Optics, Image Science, 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. (2001). The symmetry magnification function varies with detection task. *Journal of Vision, 1*(2), 137–144.

Tyler, C. W. (2004a). Beyond fourth-order texture discrimination: generation of extreme-order and statistically balanced textures. *Vision Research, 44*(18), 2187–2199.

Tyler, C. W. (2004b). Theory of texture discrimination of based on higher-order perturbations in individual texture samples. *Vision Research, 44*(18), 2179–2186.

Verghese, P., & Pelli, D. G. (1992). The information capacity of visual attention. *Vision Research, 32*(5), 983–995.

Victor, J. (1994). Images, statistics, and textures: a comment on "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture". *Journal of the Optical Society of America, A11*, 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. *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.

Victor, J. D., & Conte, M. M. (2005). Local processes and spatial pooling in texture and symmetry detection. *Vision Research, 45*(8), 1063–1073.

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

Yellott, J. (1993). Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture. *Journal of the Optical Society of America*(A10), 777–793.

Zetzsche, C., & Nuding, U. (2005). Nonlinear and higher-order approaches to the encoding of natural scenes. *Network, 16*(2–3), 191–221.