Visual short-term memory of local information in briefly viewed natural scenes: Configural and non-configural factors

Ljiljana Velisavljević

Centre for Vision Research, York University, Toronto, Ontario, Canada

James H. Elder

Centre for Vision Research, York University, Toronto, Ontario, Canada

Typical visual environments contain a rich array of colors, textures, surfaces, and objects, but it is well established that observers do not have access to all of these visual details, even over short intervals (R. A. Rensink, J. K. O'Regan, & J. J. Clark, 1997). Rather, it seems that human vision extracts only partial information from every glance. What is the nature of this selective encoding of the scene? Although there is considerable research on short-term coding of individual objects, much less is known about the representation of a natural scene in visual short-term memory (VSTM). Here, we examine the VSTM of natural scenes using a local recognition task. A major finding is that local recognition performance is better when image segments are viewed in the context of coherent rather than scrambled scenes, suggesting that observers rely on an encoding of a global 'gist' of the scene. Variations on this experiment allow quantification of the role of multiple factors in local recognition. Color statistics and the global configural context are found to be more important than local features of the target, even for a local recognition task.

Keywords: visual short-term memory, natural scenes, color, context effect, configural cues

Citation: Velisavljević, L., & Elder, J. H. (2008). Visual short-term memory of local information in briefly viewed natural scenes: Configural and non-configural factors. *Journal of Vision, 8*(16):8, 1–17, http://journalofvision.org/8/16/8/, doi:10.1167/8.16.8.

Introduction

There are some stunning examples of the limitations of our visual short-term memory (VSTM).[1] For example, Grimes (1996) found that highly significant changes to a photograph made during a saccadic interval were often not recognized. Participants generally did not notice a hat swap between two men, and only 50% noticed a head exchange between two bodies. Some (e.g., O'Regan, 1992) argue that this inability to detect obvious changes across time illustrates that little if any visual information is represented in VSTM, while others (e.g., Rensink, 2002; Simons, 1996) take a more moderate position. Regardless, these illustrations suggest that VSTM codes a limited amount of information. This is often expressed in the context of object recognition, with estimates of coding limits in the range of three to four objects per scene (Luck & Vogel, 1997; Phillips, 1974).

However, sparse object memory cannot be mapped directly to sparse scene memory, as scenes contain both objects and their contexts. In fact, object and context components are dissociable both neurologically (e.g., Epstein, DeYoe, Press, Rosen, & Kanwisher, 2001; Epstein & Kanwisher, 1998) and behaviorally (e.g., Simons, 1996). Thus, an understanding of object VSTM does not necessarily imply an understanding of scene VSTM. Although there is considerable research on object representation (Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001; Liu & Jiang, 2005; Mitroff, Simons, & Levin, 2004; Tatler, Gilchrist, & Land, 2005; VanRullen & Koch, 2003; Varakin & Levin, 2006), much less is known about the holistic representation of a scene (although see Biederman, 1972; Gegenfurtner & Rieger, 2000; Potter, 1976; Potter et al., 2004).

Studies that address natural scene VSTM tend to use match-to-sample designs to examine issues such as consolidation (Potter, 1976) and color representation (Gegenfurtner & Rieger, 2000; Spence, Wong, Rusan, & Rastegar, 2006) or local recognition designs to examine the interaction between a coherent scene and its objects (Antes & Metzger, 1980; Biederman, 1972). Specifically, Biederman (1972) discovered a scene coherence effect: recognition memory is better for objects viewed within coherent rather than scrambled scenes. However, because these local-recognition studies tend to use object probes, it is not always clear if or how these results pertain to more general scene content. Here we examine the coherence effect using a local-recognition paradigm with a probe that is a scene segment (not easily classifiable as an object) to better understand scene representation in VSTM. We use a

doi: 10.1167/8.16.8 Received February 11, 2008; published December 22, 2008 ISSN 1534-7362

series of such experiments to quantify the effect of key image properties on recognition performance. Because under natural conditions observers must often extract information rapidly, we examine local recognition for stimulus durations under 100 ms. We begin with what is known about this problem from the literature.

Previous research

Over the last three decades, there has been substantial interest in the influence of context information on object perception (e.g., Auckland, Cave, & Donnelly, 2007; Biederman, 1981; Biederman, Glass, & Stacey, 1973; Biederman, Rabinowitz, Glass, & Stacey, 1974; Boyce & Pollatsek, 1992; Boyce, Pollatsek, & Rayner, 1989; Davenport & Potter, 2004; De Graef, 1992; De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, 1992; Henderson, Pollatsek, & Rayner, 1987; Hollingworth & Henderson, 1998, 1999; Palmer, 1975). In these studies, observers must detect or name an object within its regular context, a context that is scrambled, or a context that is highly unlikely to be linked with the target object. These studies usually show better performance when an object is within its regular context, supporting the notion that both object and context information is automatically processed despite the fact that context is incidental to the task.

A natural question concerns whether this complete scene information is retained in VSTM after display offset. A variation of Biederman's paradigm (Biederman, 1972; Biederman et al., 1973, 1974) may partially address this question. In all of his experiments, participants viewed test images that consisted of a monochromatic photograph of a natural scene. The photographs were divided into six blocks that were either in their correct locations, thus maintaining the coherent global structure of the scene, or scrambled to disrupt this structure. The target was a complete object contained within one of the blocks. In a memory condition of the paradigm (Biederman, 1972), an observer viewed a coherent or scrambled test image (300, 500, or 700 ms), an arrow pointing to an area associated with an object (300 ms), and then object probes derived from the test image. The observer's task was to indicate the object that the arrow highlighted. Results showed a scene coherence effect: recognition performance was better for objects viewed within coherent rather than scrambled scenes. Because the arrow and object probes were presented after the test image, these results show that the global context somehow enhanced the representation of target objects in memory. It could be that global context acts to sharpen the representation of independent objects at the encoding stage or that objects and context are encoded and stored together in memory in a more reliable form than objects encoded and stored without a coherent global context.

Subsequent studies help to disambiguate these possibilities. For example, in a study by Antes and Metzger (1980), participants viewed briefly presented line drawings containing several objects with or without global context and were asked to indicate which of four objects had been in the image. Presentation of objects within a global context aided recognition, but only when the incorrect alternatives were inconsistent with the global context. This suggests that there must be some trace of the global context in memory, used during recall: the effects of context cannot be entirely at the encoding stage. However, it is still possible to explain these results based on parallel but independent encoding of objects and scene context, both brought to bear at the recall stage.

A more recent study using naturalistic scenes suggests that the memory encoding of objects and scene context cannot be independent. In this study (Hollingworth, 2006), observers viewed a natural scene, and memory for the visual properties of constituent objects was tested. Specifically, participants performed a two-alternative forced choice task. In each trial, one four second interval contained the target object, and another four second interval contained a distractor object which was the same object rotated 90° in depth or a similar object (a different token of the same object category). In addition, target and distractor objects were presented either within the original scene background (with arrows pointing to them) or in isolation. Discrimination performance was consistently superior when the target and distractor objects were presented within the original scene background compared to when they were presented in isolation. Because the scene context is not diagnostic for either the specific target object or its orientation, these results suggest that object and scene representations are not independent but are bound together in VSTM.

Will the same apply to local segments of a scene that are not easily identifiable as objects? A series of studies speaks to this issue (Tjan, Ruppertsberg, & Bülthoff, 1998, 1999, 2000). These experiments measured recognition performance for local segments of briefly presented natural scenes and how performance depended upon the nature of a priming stimulus presented immediately before the test image. After viewing the (primed) test image and a mask image, the task was to discriminate between a small probe segment of the test image and a probe segment taken from an unseen image. Several types of prime were found to be effective. A prime consisting of a pixel-scrambled version of the test image was found to raise performance relative to the non-primed condition, suggesting that the color histogram of the test image is one factor determining discrimination of the local segments. However, a block-scrambled prime was found to raise performance even higher, indicating some role for local configural cues. Finally, a prime identical to the test image was found to yield best performance, suggesting a role for global configural cues.

While these results are suggestive, there remain some uncertainties. First, it is unclear whether the superiority of globally coherent primes over block-scrambled primes derives from direct cues to the target segment or from indirect contextual cues. Further, there is increasing similarity between the prime and the cue as we progress from pixel-scrambled to block-scrambled primes. It is at least logically possible that in addition to helping, the scrambled primes are also partially hurting performance by forward-masking the test image. The differences in performance for the different primes may thus reflect differences in their effectiveness as masks, not as cues. Finally, if there is a role for multiple levels of cue in local VSTM, it is natural to inquire about the quantitative importance of each level.

In the present study, we will assess local VSTM recognition memory using both coherent and block-scrambled test images. By making probe blocks identical to blocks in the test image and by avoiding the priming paradigm, we will more directly assess the role of indirect contextual cues in local VSTM recognition. To aid quantitative assessment of direct local cues relative to indirect contextual cues, we will test recognition for local components that are completely occluded, therefore providing no local cues, and compare against recognition for unoccluded blocks for which both local and global cues are available.

Another goal of our study is to establish a better understanding of the role of color in local VSTM recognition. Recent work suggests that longer-term memory (at least 5 minutes between presentation and test) is better for chromatic than monochromatic images (Gegenfurtner, Wichmann, & Sharpe, 1998; Suzuki & Takahashi, 1997; Wichmann, Sharpe, & Gegenfurtner, 2002). This pattern of results extends to shorter delays in the VSTM range (Gegenfurtner & Rieger, 2000; Gegenfurtner et al., 1998; Spence et al., 2006). In the present study we specifically probe the relative importance of color to global configural cues and local content in VSTM for natural scenes.

A final question concerns the representational level of the cues on which local scene recognition is based: are these cues visual or semantic in nature? For example, a global scene composed mainly of earth tones may provide a strictly visual cue for recognition of a local component of similar hue, but the same scene may also provide a semantic cue, for example, recognition of the scene as a landscape may improve discrimination of a local component containing a tree from an alternative local component containing a chair. We know from classic work that semantic information can be extracted from a brief observation of a visual scene (e.g., Antes, 1977; Potter & Levy, 1969). For example, scenes can be categorized with image presentation times as short as 45–135 ms (Oliva & Schyns, 1997; Schyns & Oliva, 1994). A high-level description of a scene can be retained in memory if sufficient time exists for consolidation (Potter, 1976).

If semantic encoding plays a significant role, we would expect that making images less familiar would make semantic content less easily extracted and thus reduce recognition performance. Further, if the global representation underlying the coherence effect is primarily semantic, then this manipulation should also reduce the magnitude of the coherence effect. In this paper, we modulate access to semantic content by inverting the orientation and/or color of images, under the premise that semantic content will be harder to extract from images that are upside-down (Intraub, 1984; Klein, 1982; Rock, 1974; Shore & Klein, 2000) or colored in an unfamiliar way (Goffaux et al., 2005; Oliva & Schyns, 2000).

General methods

In this study, we use a local recognition task based on the studies by Biederman (1972) and Tjan et al. (1998, 1999) to examine VSTM for natural scenes. In four experiments, we examine the role of local and global cues, color, and semantics in local VSTM recognition. Test images are natural scenes that may be color or monochrome, coherent or scrambled, familiar or unfamiliar, or partially occluded or unoccluded. A combined analysis of the effects of these manipulations allows quantification of the relative importance of these factors in local VSTM.

Participants

Ten participants received CAN$8/hour for participation. The same participants participated in all experiments. All participants were naïve to the purpose of the experiments and had normal or corrected-to-normal vision.

Apparatus

A 21-inch Sony Trinitron® CRT with a display resolution of 640 × 480 pixels and a refresh rate of 85 Hz was used to display the stimuli. The experiments were programmed with Matlab 5.2 using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) for stimulus presentation and synchrony between the graphics card and monitor. A head-chin rest was used to fix viewing distance at 40 cm.

Stimuli

Test and mask images were derived from a database of 14,267 JPEG color photographs of natural scenes that are part of the Print Artist Platinum Plus (v. 12.0) product from Sierra Home. The database incorporates photographs from a variety of natural indoor and outdoor scenes. Each

photograph was cropped to a resolution of 358 × 358 pixels and divided into 64 blocks that were either in their regular positions (coherent condition) or randomly scrambled (scrambled condition). Each image subtended 30.8°. Scrambling the images produces intensity and color discontinuities not present in the coherent versions. To make the two conditions more comparable, a black two-pixel-thick lattice was overlaid on all test and mask images to occlude block boundaries. For each participant, a previously unviewed photograph was used for each test and mask image. Mask images were to ensure that the task was not testing iconic memory (Sperling, 1960).

In the first two experiments, we tested the effects of both coherent and scrambled mask images. We employed only scrambled mask images for the remaining experiments.

The two probes consisted of (a) a target probe that was one of the 64 blocks from the test image and (b) a distractor probe that was one of the 64 blocks from an unseen image. The probes were randomly selected from the images and then presented with a two pixel black border. The two probes each subtended 3.9° and were presented centrally, on either side of fixation, with 7.8° center-to-center spacing. The left/right order of the target and distractor probes was randomized from trial-to-trial.

Procedure

Each participant completed the four experiments in random order. Across experiments, a participant viewed a scene only once to avoid scene repetition effects. Within each experiment, conditions were in random order, and each condition consisted of 20 practice and 100 experimental trials. Figure 1 illustrates a typical trial in one of our experiments. Each trial consisted of a fixation stimulus containing a 1.7° fixation mark (1 s), a stimulus image (70.6 ms), a mask image (517 ms), and a two-probe stimulus (until response). The participant pressed one of two keys on a computer keyboard to indicate which target probe they believed was present in the test image. There was feedback for each trial: a tone if the response was incorrect.

Experiment 1: Global configural cues

Object recognition appears to be influenced by the visual context in which objects are seen (e.g., Biederman, 1972). In our first experiment, we tested whether this finding extends to local image patches not necessarily identified with objects. In one condition, the local target image block was presented in the natural context of the image. In the second condition, the local target was presented within a scrambled image.

The main independent variable of interest in this experiment was the nature of the global context (coherent or scrambled). To facilitate discussion and modeling, we will operationally define the term "global configural cues" as recognition cues available in coherent natural images but not in their scrambled counterparts. This term is an approximation: natural images contain structure over a

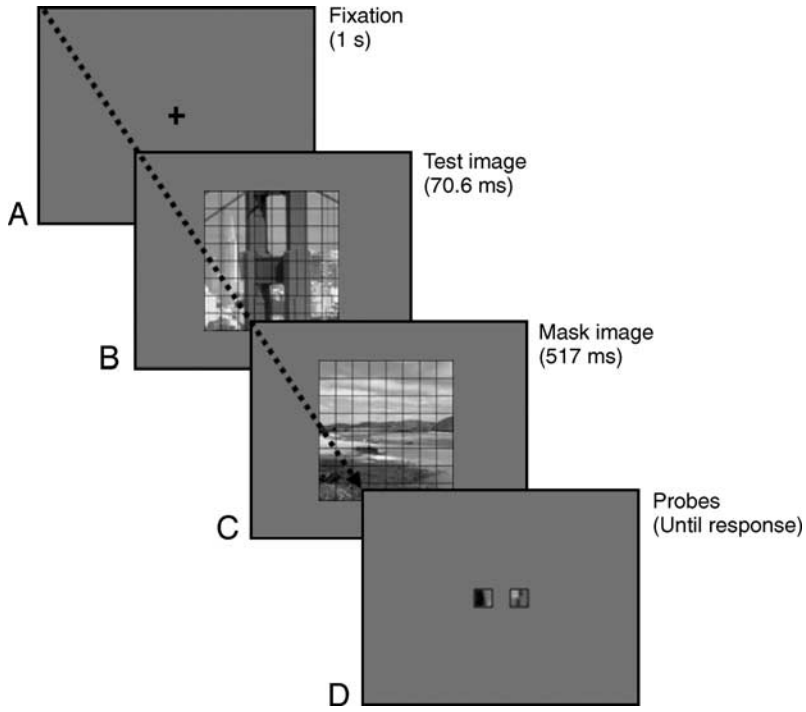


Figure 1. Stimulus sequence.

continuum of scales, and while we are disrupting structure primarily beyond the scale of a block, smaller scale structure is disrupted as well, but to a lesser degree. Note, however, that the manipulation leaves features within the target block completely unperturbed.

Both coherent and scrambled masking images were also tested, because increasing the similarity of the mask image to the test image has been shown to affect performance in some circumstances (e.g., McClelland, 1978). Thus, Experiment 1 was a 2 (test image: coherent or scrambled) × 2 (mask image: coherent or scrambled) within-subject design. All stimuli were monochromatic (Figure 2).

Results

Results are shown in Figure 3. We used a two-way within-subject ANOVA (test image × mask image) to evaluate the effect of global structure on d' for recognition of local regions in monochromatic natural images. Recognition rates were higher when test images were coherent rather than scrambled, $F(1, 9) = 14.58$, $p = .004$. The main effect of the mask type was not significant, $F(1, 9) = .60$, $p = .46$, and there was no significant interaction between mask type and image type, $F(1, 9) = .55$, $p = .48$. Our results suggest that the configural context effect discovered by Biederman (1972) is not specific to object recognition. Rather, the global configural context of a natural visual scene appears to provide an important anchor for the recall of arbitrary components of the scene.

We also note that, in the scrambled condition, recognition performance is still well above chance (grouped across mask type): $t(9) = 5.73$, $p < .001$. Thus even in the absence of global configural cues, observers are able to recognize local components of the scene. This may reflect direct coding of local cues present in the target block but may also reflect the encoding of non-configural global cues (e.g., luminance statistics) that assist in identifying the target probe.

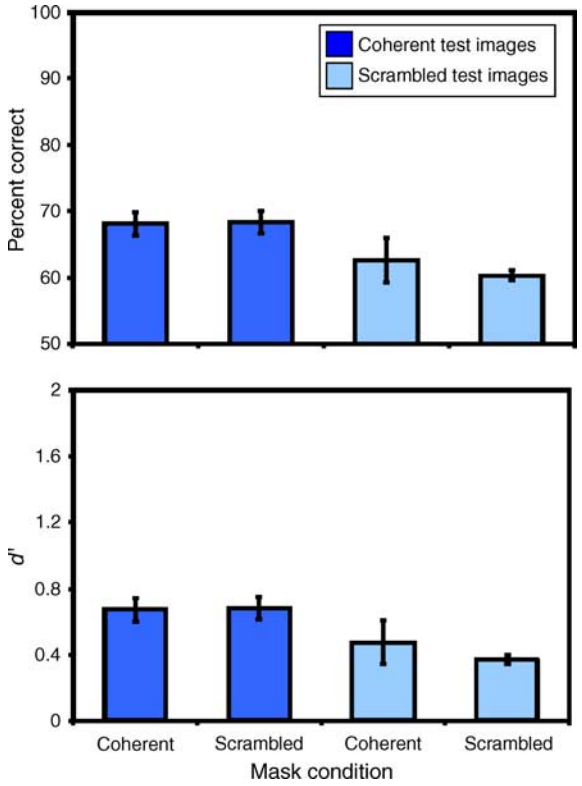


Figure 3. Recognition performance for monochrome images. Error bars indicate ±1 *SEM*.

Experiment 2: Color

Recent work suggests that color may enrich the rapidly computed representation of a visual scene (Gegenfurtner

Figure 2. Coherent and scrambled monochromatic images.

Figure 4. Coherent and scrambled color image stimuli.

& Rieger, 2000; Spence et al., 2006); here, we examine the role of color in a local VSTM recognition task. We repeat our first experiment but with color stimuli to determine 1) whether color plays a role in the coding and recall of local components of a visual scene and 2) whether color mediates the global context effect witnessed for monochrome stimuli.

Experiment 2 was a 2 (test image: coherent or scrambled) × 2 (mask image: coherent or scrambled) within-subject design. Test images, mask images and probe stimuli were all chromatic (Figure 4).

Results

Results are shown in Figure 5. A two-way within-subject ANOVA (test image × mask image) was performed to examine the effect of coherence on d' for local recognition. Recognition rates were better when test images were coherent rather than scrambled, $F(1, 9) = 10.11$, $p = .01$, indicating an effect of global configural context. The main effect of the mask condition (scrambled or coherent) was not significant, $F(1,9) = .05$, $p = .82$, and no significant interaction with test image was observed, $F(1, 9) = .01$, $p = .91$. Grouping across mask type, results again show that local VSTM performance was well above chance for scrambled scenes, $t(9) = 16.16$, $p < .001$.

The absence of a difference between the coherent and scrambled mask conditions in this and the previous experiment lies in contrast to prior demonstrations of conceptual masking effects in scene memory (Intraub, 1984; Loftus & Ginn, 1984; Loftus, Hanna, & Lester, 1988; Potter, 1976) and scene gist recognition (Loschky et al., 2007). However, there are numerous methodological differences between the present paradigm and these prior studies, and future research would be required to investigate the difference in the pattern of results across studies.

We compared Experiments 1 and 2[2] using a two-way within subject ANOVA (chromaticity × coherence) on d' to examine the effect of color on recognition rates and on the global coherence effect (Figure 6).[3] There was a significant main effect of chromaticity (i.e., whether the images were monochromatic or color), $F(1, 9) = 121.94$, $p < .001$, reflecting better performance for color images,

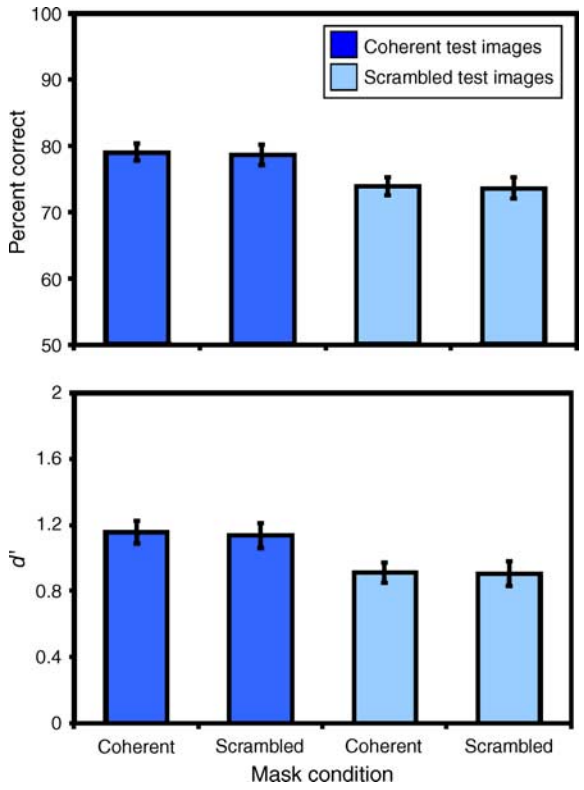


Figure 5. Recognition performance for chromatic stimuli. Error bars indicate ±1 *SEM*.

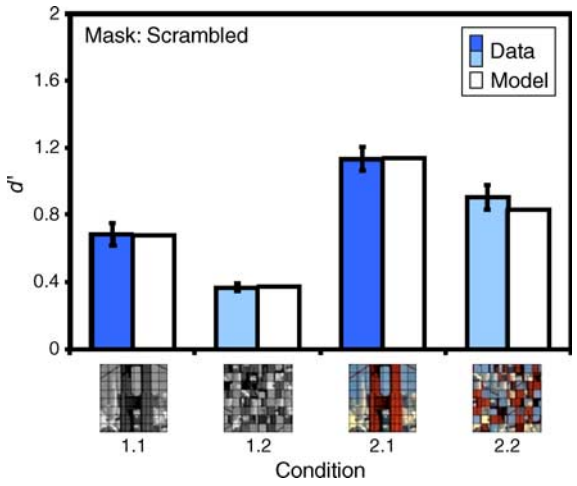


Figure 6. Recognition performance for monochromatic and color stimuli. Error bars indicate ±1 *SEM*. The model fit is described in Analysis section.

and a significant main effect of coherence, $F(1, 9) = 25.77$, $p < .001$, indicating better performance for coherent than scrambled images. The interaction between chromaticity and coherence was not significant, $F(1, 9) = .42$, $p = .53$, suggesting that global configural cues and color cues to local recognition act relatively independently.

Experiment 3: Familiarity

Our first two experiments show that both global configural context and color are significant cues for the coding and recall of local scene components. In our third experiment, we ask whether these cues are best described as visual or semantic in nature. Prior work suggests that semantic information can be extracted from a brief observation of a visual scene (e.g., Antes, 1977; Potter, 1976). Thus, the effects of color and global configural context may result from bottom-up visual cues as well as top-down semantic categorization mechanisms. Here we attempt to distinguish these factors by reducing access to semantic content in our natural image stimuli. A number of researchers (Intraub, 1984; Klein, 1982; Rock, 1974; Shore & Klein, 2000) have reported that access to semantic information is reduced if an image is made unfamiliar by inverting its orientation. We hope to further reduce familiarity and therefore access to semantic content by color inversion, because previous research suggests that atypical scene colors can affect scene categorization speed and accuracy (Goffaux et al., 2005; Oliva & Schyns, 2000). To facilitate discussion and modeling, we will operationally define the term "familiarity" as the set of recognition cues available in normal natural images but not in their inverted counterparts.

Experiment 3 was a 2 (test image: coherent or scrambled) × 3 (familiarity: orientation inversion, color inversion, or both color and orientation inversion) within-subject design (Figure 7). All stimuli were chromatic. Color inversion was achieved by taking the 8-bit complement of each color channel. Mask images were scrambled for all conditions. Mask images and probe stimuli were inverted in orientation and/or color in correspondence with the test images for each condition.

An effect of coherence for test images that are inverted in orientation and/or color would suggest that local VSTM performance derives to some degree from non-semantic configural cues. Comparison between Experiments 2 and 3 allows for an assessment of the role of semantic information: lower recognition rates for inverted stimuli would suggest a role for semantic information in the encoding and retrieval process. Any differences that are specific to coherent conditions would suggest a role for semantic information encoded through global configural properties.

Results

Results are shown in Figure 8. We used a two-way within-subject ANOVA (test image × inversion) to evaluate the effect of reducing access to semantic content on local recognition performance. The main effect of the type of inversion, $F(2,18) = .89$, $p = .43$, and the interaction between coherence and the type of inversion, $F(2,18) = .08$, $p = .92$, were not significant. The main effect of coherence approached significance, $F(1, 9) = 3.76$, $p = .08$: recognition rates were moderately higher for coherent than for scrambled scenes, suggesting that local recognition may benefit from non-semantic global configural content in the scene. Further, pooling across inversion types, results show that recognition rates were well above chance for scrambled scenes, $t(9) = 15.48$, $p < .001$, suggesting that local VSTM performance may also derive from non-configural and non-semantic cues (e.g., textural cues, luminance, and color histograms).

To further examine the effect of semantics, planned comparisons were conducted using within-subject ANOVAs between Experiments 2 and 3 (Figure 9). For coherent images, recognition performance was generally superior when images were not inverted (i.e., access to semantic content was maintained): $F(1, 9) = 4.95$, $p = .053$ (orientation), $F(1, 9) = 5.65$, $p = .04$ (color), and $F(1, 9) = 19.67$, $p = .002$ (orientation and color). For scrambled test images, a similar trend is evident for the first two comparisons, and a significant benefit of maintaining normative orientation and color is evident for the last comparison: $F(1, 9) = 1.17$, $p = .31$ (orientation), $F(1, 9) = 2.50$, $p = .15$ (color), and $F(1, 9) = 5.29$, $p = .047$ (orientation and color). Thus, while semantic cues appear to be most readily extracted and exploited from globally coherent images, some effects are also evident in

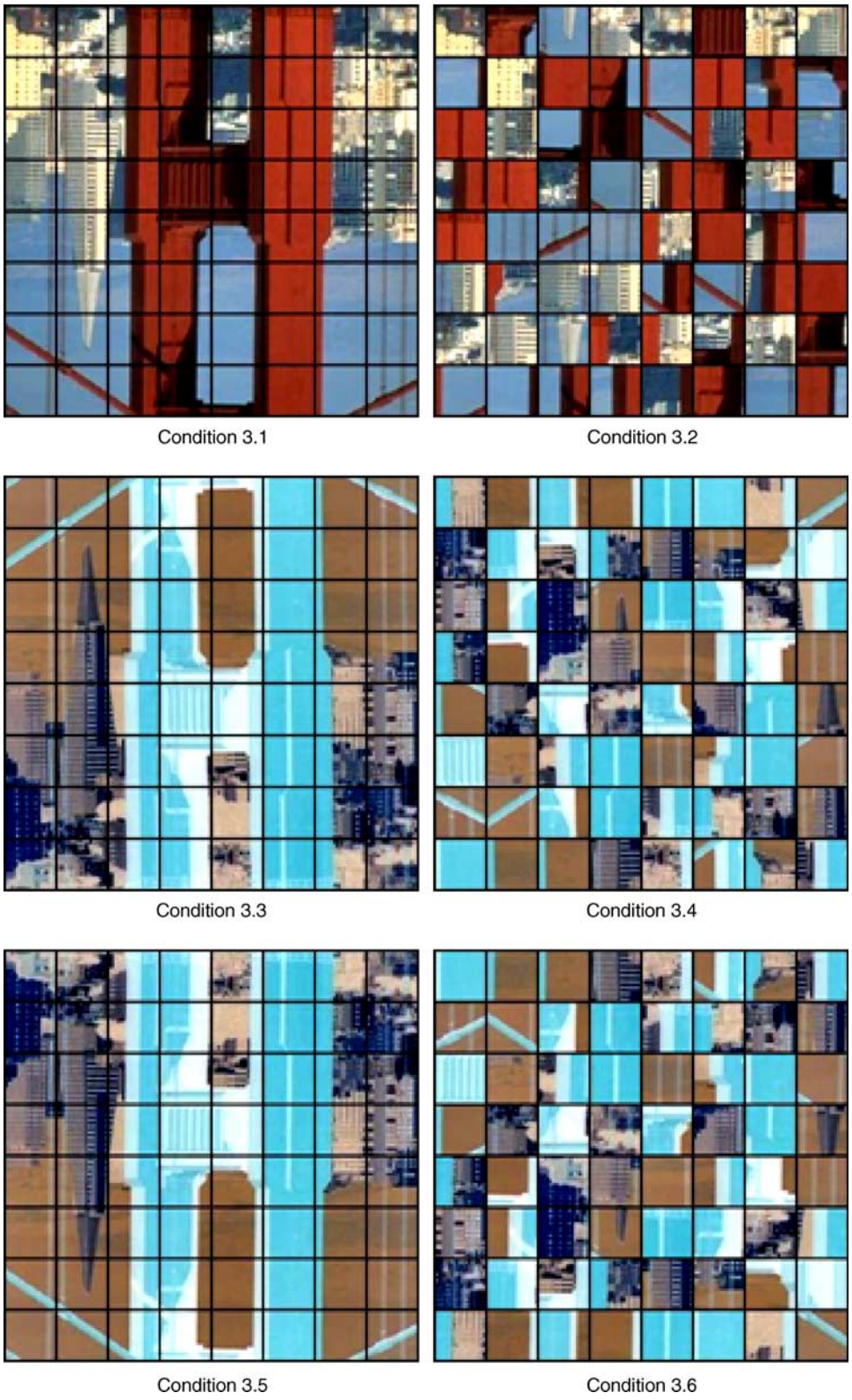


Figure 7. Images inverted in orientation (3.1 and 3.2), color (3.3 and 3.4), and both orientation and color (3.5 and 3.6).

scrambled images without coherent global configural cues. We conclude that the effect of inverting the images in orientation and color may be mediated partly through less efficient encoding or decoding of unfamiliar or unusual local configurations and color combinations. For this reason, we prefer to characterize the effects of inversion as "familiarity" effects rather than semantic effects, because some component of the effect seems

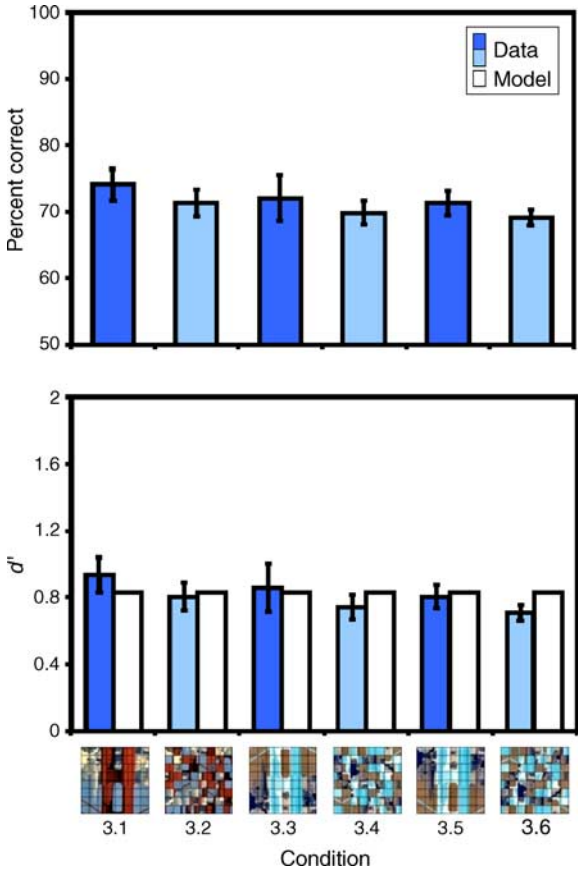


Figure 8. Recognition performance for orientation- and/or color-inverted stimuli. Error bars indicate ±1 *SEM*. White bars indicate model fit (Analysis section).

likely to be pre-semantic. We will consider stronger conditions for semantic activation shortly in the Analysis section.

Experiment 4: Occlusion

There are at least two ways that participants could solve the local recognition problem of Experiments 1, 2, and 3. They could (a) recognize the target probe directly from detailed local information or (b) indirectly infer the target probe from the global structure of the image, global image statistics, or local configurations surrounding the probe block. Our fourth experiment distinguishes these possibilities.

In this experiment, only alternating blocks of the test image were displayed in a checkerboard fashion (Figure 10). We refer to the other half of the blocks as *occluded*. In one condition, one of the visible blocks was selected as the target. In the comparison condition, one of the occluded blocks was selected. Thus, Experiment 4 was a 2 (test image: coherent or scrambled) × 2 (probe target: visible or occluded) within-subject design. All stimuli were chromatic. The checkerboard was randomly chosen to be either as shown in Figure 10 or its complement. Mask images were scrambled.

Comparison between the conditions within Experiment 4 allows the relative importance of direct and indirect cues to be assessed. If participants indirectly infer the target probe from the scene context or general scene statistics, recognition performance should not differ between the visible and occluded target groups. If participants encode detailed local information from each target probe, recognition performance should improve when the target probe is visible rather than occluded. To facilitate discussion and modeling, we operationally define the term "local cues" as recognition cues extracted directly from the target block. Note also that the occlusion of alternate blocks may have an overall impact on performance, and this can be measured by comparison with Experiment 2. We will refer to this as a reduction in the "global visibility" of the stimulus.

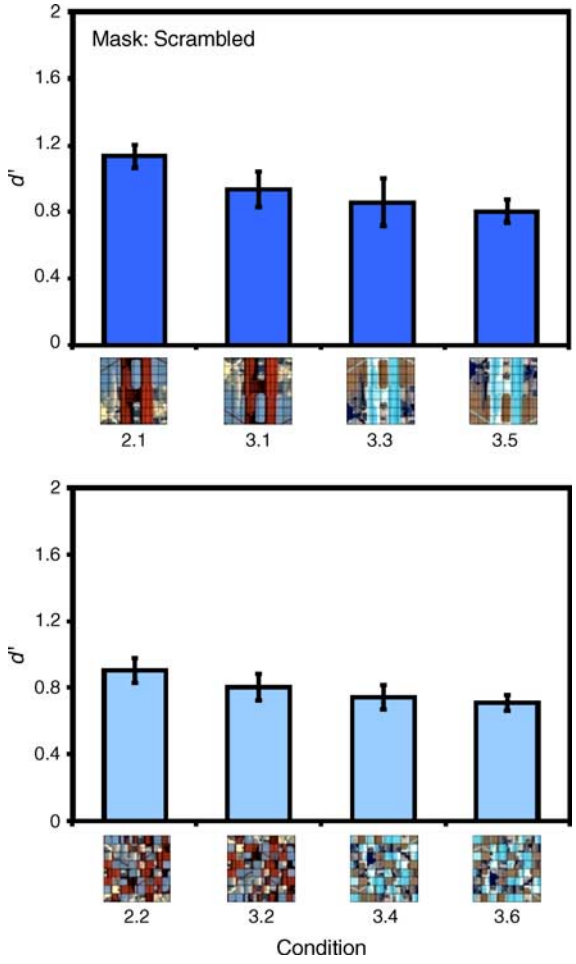


Figure 9. Comparing recognition performance for Experiments 2 and 3 (Inverted and Non-Inverted stimuli). Error bars indicate ±1 *SEM*.

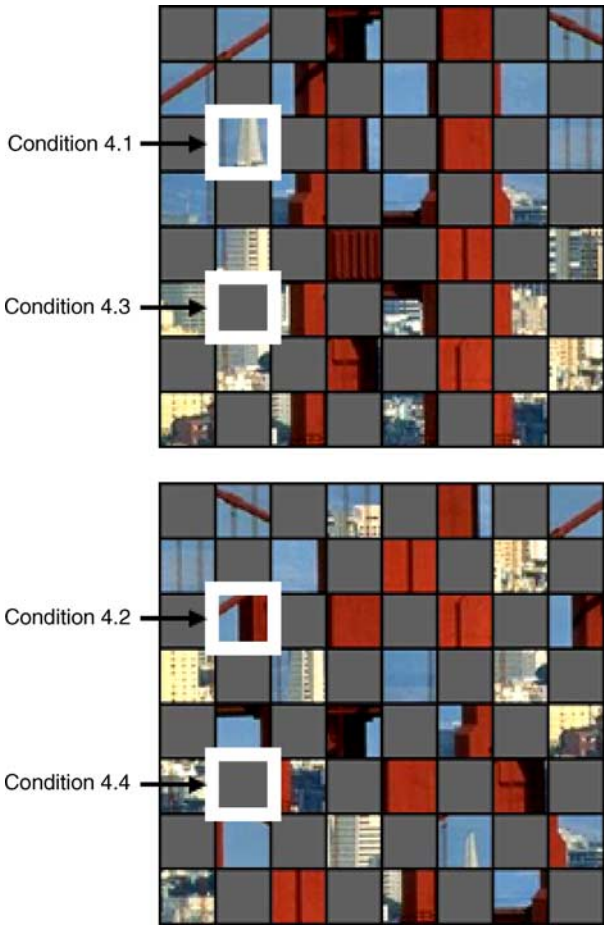


Figure 10. Example of a coherent (top) and scrambled (bottom) test image with visible (4.1 and 4.2) and occluded (4.3 and 4.4) targets.

Results

Figure 11 shows the results for this experiment. We used a two-way within subject ANOVA (test image × probe target) to evaluate whether the coherence effect relies on the direct processing of local cues or whether when such cues are absent, they can be inferred from the surrounding region. Overall recognition rates were higher when test images were coherent rather than scrambled, $F(1, 9) = 6.50$, $p = .03$, demonstrating a benefit of global configural context to local VSTM recognition despite the fact that test images were half-occluded. Recognition rates were not significantly affected overall by whether the specific local target was visible, $F(1, 9) = 1.71$, $p = .22$, although the interaction between global coherence and local target visibility approached significance, $F(1, 9) = 4.50$, $p = .06$. Furthermore, planned comparisons indicated that for coherent images, recognition performance did not depend on whether the target block was visible, $F(1, 9) = .01$, $p = .92$. One possible account of this result is that for coherent test image stimuli, occluded information can be interpolated from surrounding visible blocks, thus reducing the importance of direct visibility. For scrambled test images, performance grouped across target visibility was well above chance, $t(9) = 11.04$, $p < .001$. More importantly, the effect of occlusion approached significance, $F(1, 9) = 4.40$, $p = .065$: recognition rates were higher for visible target blocks, suggesting that when global configural cues are unavailable, observers may rely on specific visual detail within the target block.

To investigate the effect of occlusion due to the checkerboard mask, we compared performance across Experiments 2 and 4. Specifically, for this analysis, we used coherent and scrambled test images from Experiment 2 (i.e., 2.1 and 2.2) and coherent and scrambled occluded images with visible target blocks from Experiment 4 (i.e., 4.1 and 4.2). Thus, across all conditions the target block was visible and what was manipulated was coherence and the degree to which context blocks were visible (i.e., global visibility). Accordingly, a two-way within-subject ANOVA (test image × global visibility) was performed (see Figure 12 for results). There was a significant effect

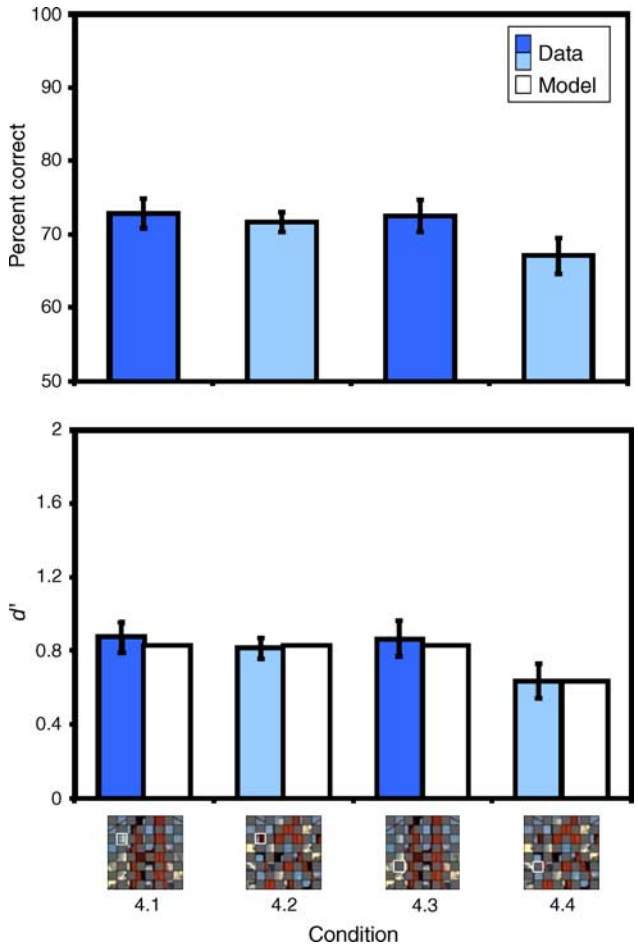


Figure 11. Recognition performance as a function of target and test image condition. Error bars indicate ±1 *SEM*. White bars indicate model fit (Analysis section).

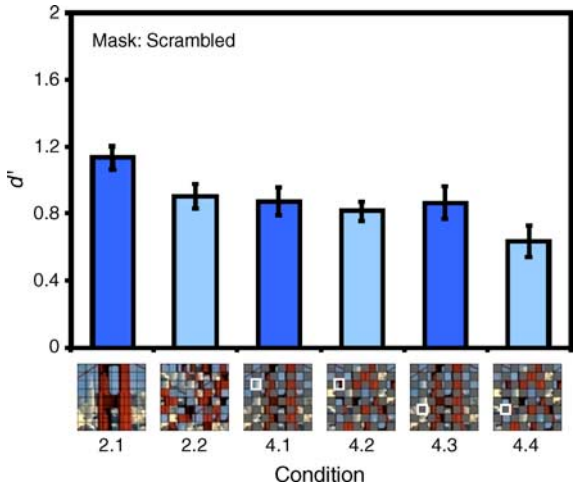


Figure 12. Recognition performance for Experiments 2 and 4. Error bars indicate ±1 *SEM*.

of global visibility, $F(1, 9) = 10.31$, $p = .01$, reflecting better performance with unoccluded images, and coherence $F(1, 9) = 6.88$, $p = .03$, indicating better performance with coherent rather than scrambled images. The interaction between global visibility and coherence was not significant, $F(1, 9) = 2.34$, $p = .16$. One might have predicted otherwise. For example, were the effect of the checkerboard mask only to interfere with the rapid extraction of global configural information, one would expect the effect to be more pronounced in the coherent condition than the scrambled condition.

A closer examination of the data in Figure 12, however, suggests a more complex interaction. Specifically, for visible target blocks, the coherence effect (contrasting Condition 4.1 and 4.2) is reduced, $F(1, 9) = .89$, $p = .37$, suggesting that the checkerboard mask may lead participants to rely less on global configural cues and more on direct local cues when they are available. In contrast, the effect of coherence is highly significant for invisible target blocks (contrasting Condition 4.3 and 4.4), $F(1, 9) = 9.21$, $p = .01$, suggesting that some global configural cues are still used, at least when direct local cues are unavailable. One way to explain this pattern of results is that the checkerboard mask does lead the participant to rely more on local, or at least regional cues, and these regional cues may be available either directly from the target block when visible or indirectly from neighboring blocks when the target is occluded but the image is coherent.

Analysis

In four experiments, we have examined VSTM for local scene content in images, and how it depends on global coherence, color, familiarity, local cues, and visibility. The goal of the present section is to try to extract more precisely from these data the principal factors that appear to affect recognition performance. The factors we consider are: Global Configural Cues, Color Cues, Familiarity, Local Cues, and Global Visibility. Table 1 delineates the presence or absence of these characteristics for each condition. We use only conditions with scrambled masks to examine two statistical accounts of the data.

The first model consists of all five of these factors. We modeled the data by regressing d' against the five factors and testing all subsets of the model factors using the prediction residual sum of squares (*PRESS*) statistic in a leave-one-out cross-validation design (Allen, 1974; Jobson, 1991). This statistic is simply the sum of squares of the prediction residuals for observations omitted from the regression (see Appendix A for calculation). The *PRESS* statistic allows fair comparison between models of differing complexity.

We found the minimum *PRESS* statistic to be achieved by including four of the five factors: Color Cues, Global Visibility, Familiarity, and Global Configural Cues. Including the Local Cues factor increased the *PRESS* statistic. The coefficient of determination for the 4-factor model was $r^2 = 0.23$.

Figure 13 shows the factor weights with their 95% confidence intervals (in d' units). Color is clearly the most important factor with a weight roughly double that of the other factors. Without any of these four cues (i.e., an inverted, scrambled, half-occluded monochrome image in which the target block is not visible), the model predicts that d' for the task would be 0.08, not significantly greater than chance ($p > .05$).

We also tested a second model that included the Familiarity, Color Cues, and Global Visibility factors as well as compound factors composed of conjunctions or disjunctions of individual properties shown in Table 1. The compound factors are activated by the following combinations of properties:

a. The *Semantics* factor is activated by images with the properties of Global Configural Cues *and* Familiarity *and* Global Visibility.
b. The *Regional Cues* factor is activated by images with the property of Local Cues *or* Global Configural Cues.

The construction of the Semantics factor was based on the hypothesis that top-down or semantic effects may depend upon viewing familiar, unoccluded, coherent images. The construction of the Regional Cues factor was based on the hypothesis that local, direct cues to the target are available if either the target is visible *or* the target is occluded but there is locally coherent context that can be used to locally interpolate these target cues.

Again we modeled the data by regressing d' against the 5 factors and testing all subsets of the model factors using

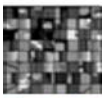
Conditions	Properties				
	Global configural cues	Color cues	Familiarity	Global visibilty	Local cues
1.1 	✓	✗	✓	✓	✓
1.2 	✗	✗	✓	✓	✓
2.1 	✓	✓	✓	✓	✓
2.2 	✗	✓	✓	✓	✓
3.1 	✓	✓	✗	✓	✓
3.2 	✗	✓	✗	✓	✓
3.3 	✓	✓	✗	✓	✓
3.4 	✗	✓	✗	✓	✓
3.5 	✓	✓	✗	✓	✓
3.6 	✗	✓	✗	✓	✓
4.1 	✓	✓	✓	✗	✓
4.2 	✗	✓	✓	✗	✓
4.3 	✓	✓	✓	✗	✗
4.4 	✗	✓	✓	✗	✗

Table 1. Test image properties in each condition across the experiments.

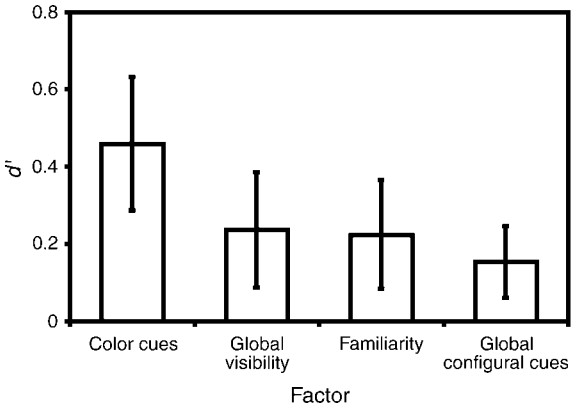


Figure 13. Model 1 factor weights. Error bars represent 95% confidence intervals.

the *PRESS* statistic. The minimum *PRESS* statistic was achieved by including only 3 of the 5 factors: Color Cues, Semantics, and Regional Cues. The coefficient of determination for this 3-factor model was $r^2 = 0.244$, slightly more than for our initial 4-factor model. Figure 14 shows the factor weights with their 95% confidence intervals (in d' units). The predicted performance based on Model 2 is shown in Figures 6, 8, and 11. Without any of these three cues (i.e., an inverted, scrambled, half-occluded monochrome image in which the target block is not visible), the model predicts that d' for the task would 0.18, not significantly greater than chance ($p > .05$).

Note that this model accounts for the results in Figure 12 with only three levels of performance: High (Condition 2.1), Medium (Conditions 2.2, 4.1, 4.2, 4.3) and Low (Condition 4.4). The drop from High to Medium is accounted for by a loss of the Semantics factor (either through loss of coherence or global visibility). The drop from Medium to Low is accounted for by loss of Regional Cues (either direct or interpolated local cues). The smaller variations between conditions were found to not be statistically reliable based upon our cross-validation results.

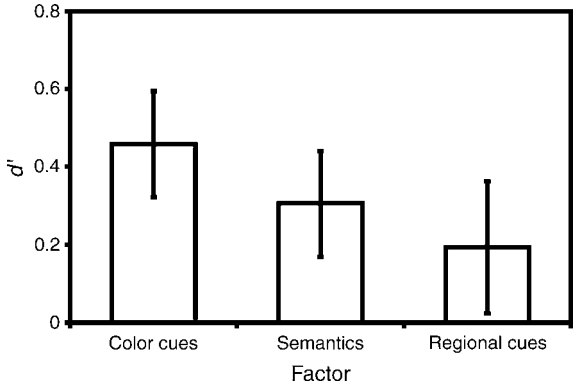


Figure 14. Model 2 factor weights. Error bars represent 95% confidence intervals.

Discussion

A typical natural scene contains a multitude of colors, textures, objects, and surfaces. While prior studies of global contextual effects in recognition performance have focused on the recognition of objects, in this study, we have examined recognition of arbitrary segments of a natural scene to better understand complete scene representation in memory. We used a local recognition task to examine the effect of a scene's coherence while varying the availability of configural cues, color cues, semantic cues, and direct local cues.

We have shown that the visual recognition of local segments of a natural image is strongly mediated by the global structure of the image. This effect is as strong for monochromatic as chromatic scenes, and the benefit does not disappear when the scene is inverted in orientation or color. These results suggest that the VSTM representation of local components of a scene hinges, at least in part, on an encoding of the global scene structure. This is consistent with the suggestion of Antes, Penland, and Metzger (1981) that global processing of images is a dominant influence on object recognition performance under brief viewing exposures (100 ms).

Although global scene structure is important for the task, we also note that, in the scrambled condition, recognition performance is still well above chance. Thus, even in the absence of global configural cues, observers are able to recognize local components of the scene. This may reflect direct coding of local cues present in the target block but may also reflect the encoding of non-configural global cues such as luminance and color statistics. In fact, for the parameters of this task, color was found to be the strongest factor determining recognition performance. This is consistent with prior results showing a significant effect of color cues on recognition and memory performance (Gegenfurtner et al., 1998; Suzuki & Takahashi, 1997; Wichmann et al., 2002).

The action of color cues could occur at encoding and/or retrieval stages. During encoding, color can aid segmentation of objects through a bottom-up sensory process that is independent of color knowledge (Mollon, 1989; Polyak, 1957; Walls, 1942). Color could also be coded in memory and used in the retrieval process as a low-level cue (Gegenfurtner & Rieger, 2000; Spence et al., 2006). Additionally, color could be used as a cue to rapidly index scene categories, thereby providing higher-level facilitation of the recognition process (Humphrey, Goodale, Jakobson, & Servos, 1994). Gegenfurtner and Rieger (2000) demonstrated that during short image presentation times (16 ms), the benefit of color was in segmentation, whereas at longer image presentation times (64 ms), the advantage of color was from its representation in memory. Gegenfurtner and Rieger's results would suggest that the color benefit in our experiment (with image presentation times of 70.6 ms) is manifested in the coding and retrieval (memory) process, not the segmentation process.

The fact that recognition performance decreases when coherent and scrambled test images are inverted in color and/or orientation suggests that our VSTM system is specifically tuned to the familiar colors and orientations typical of natural scenes. This tuning may be mediated in part through semantic or high-level representations that are indexed more effectively by familiar than unfamiliar stimuli. We know from prior work that semantic information can be extracted from brief observation of visual scenes (Antes, 1977; Potter & Levy, 1969) and retained in memory if sufficient time exists for consolidation (Potter, 1976). The observed trend toward better performance for inverted images that are coherent rather than scrambled suggests that this semantic information may be combined with pre-semantic configural cues.

The fact that recognition performance is generally better for coherent scenes than for scrambled scenes suggests that global configural cues are encoded explicitly and used in retrieval to discriminate the local target or enhance the encoding of the local target probe in the test image. While it seems likely that the coherence effect is largely effected through explicit coding of contextual cues, the results of Experiment 4 suggest that there may also be a direct enhancement in the coding of the local target. Specifically, we found a weak interaction between the effects of modulating the availability of the local target (through occlusion) and modulating the availability of global cues (through coherence): while recognition was unaffected by occlusion of the target block in coherent images, in scrambled images, occluding the target block lead to a drop in performance. This suggests that spatial coherence may allow indirect inference of local properties of the target block from neighboring blocks when the target block is occluded but the image is coherent.

Conclusion

We have shown that the visual recognition of local segments from natural images is strongly mediated by the global context of the image, thus extending prior results specific to object recognition. Through a variety of stimulus manipulations we have assessed the relative importance of multiple image properties determining local recognition performance. A cross-validation analysis suggests a concise account of local VSTM performance for natural scenes based on Color Cues, Semantics, and Regional Cues, in decreasing order of importance.

Appendix A

To calculate the PRESS statistic for each candidate model, we excluded observation i, fit the model to the remaining n-1 points, and calculated the predicted value $\hat{y}(i)$ for the excluded observation $y(i)$ and the residual for this observation: $e(i) = y(i) - \hat{y}(i)$. We repeated this procedure for each observation $i = [1,2,\ldots, n]$, producing a set of n *PRESS* residuals $[e(1), e(2),\ldots, e(n)]$. The *PRESS* statistic for each model is the sum of the squared *PRESS* residuals: $PRESS = \sum_{i=1}^{n} e(i)^2$.

Acknowledgments

We thank the reviewers for their valuable comments and suggestions. We also thank Ricardo Tabone for programming the experiments. This research was supported by grants from NSERC and PREA.

Commercial relationships: none.
Corresponding authors: Ljiljana Velisavljević and James H. Elder.
Email: lvelisavljevic@gmail.com and jelder@yorku.ca.
Address: Centre for Vision Research, 4700 Keele Street, Room 0009 CSEB, Toronto, Ontario, Canada, M3J 1P3.

Footnotes

[1]Terms such as visual short-term memory (e.g., Hollingworth, 2006; Melcher, 2006), short-term memory (e.g., Potter, Staub, & O'Connor, 2004), visual working memory (e.g., Liu & Jiang, 2005; Vogel, Woodman, & Luck, 2006), working memory (e.g., Gajewski & Henderson, 2005; Krawczyk, Gazzaley, & D'Esposito, 2007), online memory (e.g., Hollingworth, 2005), and transaccadic memory or integration (e.g., Henderson, 1997) have all been used to refer to the selective maintenance of behaviorally relevant natural scene information in memory. Although such terms are used to address similar or overlapping theoretical concepts, the differentiation between these terms is seldom made explicit. In this paper, the term visual short-term memory will be used to refer to the memory representations created as a part of processing a visual scene. In particular, the term visual is exclusively used to denote the stimulus input modality (i.e., visual scenes) rather than to characterize the nature of the emerging representations (e.g., visual/pictorial vs. verbal/semantic/conceptual). Thus, no a priori assumptions are made regarding the types of memory traces created. Instead experimental manipulations are employed to study this issue.

[2]All cross-experimental comparisons use only the scrambled mask conditions, because mask type did not have a significant effect on recognition performance and the scrambled mask was used in all four experiments.

[3]Figures 6, 8, and 10 also show a fit to the data of a factor model that we detail in Analysis section.

References

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics, 16,* 125–127.

Antes, J. R. (1977). Recognizing and localizing features in brief picture presentations. *Memory & Cognition, 5,* 155–161.

Antes, J. R., & Metzger, R. L. (1980). Influences of picture context on object recognition. *Acta Psychologica, 44,* 21–30.

Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychological Research, 43,* 277–292. [PubMed]

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review, 14,* 332–337. [PubMed]

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77–80. [PubMed]

Biederman, I. (1981). Do background depth gradients facilitate object identification? *Perception, 10,* 573–578. [PubMed]

Biederman, I., Glass, A. L., & Stacey, E. W., Jr. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97,* 22–27. [PubMed]

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacey, E. W., Jr. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103,* 597–600. [PubMed]

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 531–543. [PubMed]

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 556–566. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15,* 559–564. [PubMed]

De Graef, P. (1992). Scene-context effect and models of real world perception. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 243–259). New York: Springer-Verlag.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317–329. [PubMed]

Epstein, R., DeYoe, E. A., Press, D. Z., Rosen, A. C., & Kanwisher, N. (2001). Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cognitive Neuropsychology, 18,* 481–508.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601. [PubMed]

Gajewski, D. A., & Henderson, J. M. (2005). Minimal use of working memory in a scene comparison task. *Visual cognition: Special issue on real-world scene perception, 12,* 979–1002.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10,* 805–808. [PubMed] [Article]

Gegenfurtner, K. R., Wichmann, F. A., & Sharpe, L. T. (1998). The contribution of color to visual memory in X-chromosome-linked dichromats. *Vision Research, 38,* 1041–1045. [PubMed]

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neuropsychological evidence. *Visual Cognition, 12,* 878–892.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception.* Oxford: Oxford Press.

Henderson, J. M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology, 46,* 319–341. [PubMed]

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8,* 51–55.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1987). Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 449–463. [PubMed]

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 396–411. [PubMed]

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 58–69. [PubMed]

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception?

Journal of Experimental Psychology: General, 127, 398–415. [PubMed]

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica, 102,* 319–343. [PubMed]

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 113–136.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8,* 761–768. [PubMed]

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception, 23,* 1457–1481. [PubMed]

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 115–125. [PubMed]

Jobson, J. D. (1991). *Applied multivariate data analysis, vol. I: Regression and experimental design.* New York: Springer.

Klein, R. M. (1982). Patterns of perceived similarity cannot be generalized from long to short exposure durations and vice-versa. *Perception & Psychophysics, 32,* 15–18.

Krawczyk, D. C., Gazzaley, A., & D'Esposito, M. (2007). Reward modulation of prefrontal and visual association cortex during an incentive working memory task. *Brain Research, 1141,* 168–177. [PubMed]

Liu, K., & Jiang, Y. (2005). Visual working memory for briefly presented scenes. *Journal of Vision, 5*(7):5, 650–658, http://journalofvision.org/5/7/5/, doi:10.1167/5.7.5. [PubMed] [Article]

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 435–441. [PubMed]

Loftus, G. R., Hanna, A. M., & Lester, L. (1988). Conceptual masking: How one picture captures attention from another picture. *Cognitive Psychology, 20,* 237–282. [PubMed]

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1431–1450. [PubMed]

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. [PubMed]

McClelland, J. L. (1978). Perception and masking of wholes and parts. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 210–223. [PubMed]

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*(1):2, 8–17, http://journalofvision.org/6/1/2/, doi:10.1167/6.1.2. [PubMed] [Article]

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66,* 1268–1281. [PubMed]

Mollon, J. D. (1989). "Tho' she kneel'd in that place where they grew…" The uses and origins of primate colour vision. *Journal of Experimental Biology, 146,* 21–38. [PubMed] [Article]

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34,* 72–107. [PubMed]

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210. [PubMed]

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46,* 461–488. [PubMed]

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519–526.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

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

Polyak, S. L. (1957). *The vertebrate visual system.* Chicago: University of Chicago Press.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522. [PubMed]

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81,* 10–15. [PubMed]

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 478–489. [PubMed]

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53,* 245–277. [PubMed]

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230,* 78–85. [PubMed]

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science, 5,* 196–200.

Shore, D. I., & Klein, R. M. (2000). The effect of scene inversion on change blindness. *Journal of General Psychology, 127,* 27–43. [PubMed]

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301–305.

Spence, I., Wong, P., Rusan, M., & Rastegar, N. (2006). How color enhances visual memory for natural scenes. *Psychological Science, 17,* 1–6. [PubMed]

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74,* 1–29.

Suzuki, K., & Takahashi, R. (1997). Effectiveness of color in picture recognition memory. *Japanese Psychological Research, 39,* 25–32.

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58,* 931–960. [PubMed]

Tjan, B. S., Ruppertsberg, A. I., & Bülthoff, H. H. (1998). Early use of configural information in rapid scene perception. *Perception 27 ECVP Abstract Supplement*.

Tjan, B. S., Ruppertsberg, A. I., & Bülthoff, H. H. (1999). Early use of color but not local structure in rapid scene perception. *Investigative Ophthalmology & Visual Science, 40,* 414.

Tjan, B. S., Ruppertsberg, A. I., & Bülthoff, H. H. (2000). Local structure facilitates rapid scene perception. *Perception 29 ECVP Abstract Supplement*.

VanRullen, R., & Koch, C. (2003). Competition and selection during visual processing of natural scenes and objects. *Journal of Vision, 3*(1):8, 75–85, http://journalofvision.org/3/1/8/, doi:10.1167/3.1.8. [PubMed] [Article]

Varakin, D. A., & Levin, D. T. (2006). Change blindness and visual memory: Visual representations get rich and act poor. *British Journal of Psychology, 97,* 51–77. [PubMed]

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1436–1451. [PubMed]

Walls, G. L. (1942). *The vertebrate eye and its adaptive radiation*. Bloomfield Hills, MI: Cranbrook Institute of Science.

Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 509–520. [PubMed]