VISION RESEARCH

An International Journal for Functional Aspects of Vision

ELSEVIER

Vision Research

Is visual attention required for robust picture memory?

Jeremy M. Wolfe [a,b,*], Todd S. Horowitz [a,b], Kristin O. Michod [c]

[a] *Visual Attention Lab, Brigham and Women's Hospital, 64 Sidney St. Suite 170, Cambridge, MA 02139-4170, USA*
[b] *Department of Ophthalmology, Harvard Medical School, USA*
[c] *Department of Psychology, University of Delaware, USA*

Abstract

Humans can remember many scenes for a long time after brief presentation. Do scene understanding and encoding processes require visual selective attention, or do they occur even when observers are engaged in other visual tasks? We showed observers scene or texture images while they performed a visual search task, an auditory detection task, or no concurrent task. Concurrent tasks interfered with memory for both image types. Visual search interfered more than effects of auditory detection even when the two tasks were equally difficult. The same pattern of results was obtained with concurrent tasks presented during the encoding or consolidation phases. We conclude that visual attention modulates picture memory performance. We did not find any aspect of picture memory to be independent of attentional demands.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Picture memory; Visual search; Attention; Dual-task interference; Gist; Scene processing

1. Introduction

The ability to rapidly perceive and understand complex visual scenes is one of the most fundamental and impressive aspects of human vision. Potter (1976) and Schyns and Oliva (1994) have shown that scenes can be identified in about 100 ms, indicating that the information necessary for scene identification can be extracted very quickly. Understanding this talent is one of the larger challenges faced by visual perception research. Not only do humans recognize scenes with ease, but they also remember those scenes remarkably well (Shepard, 1967). Standing's seminal studies (1973, 1970) showed that humans have a vast memory capacity for photographs of natural objects presented for only a few seconds. In the existing literature, picture or scene memory has been studied under conditions of full attention. However, outside of the laboratory, we rarely concentrate on encoding scenes for later recall. Rather, we are usually engaged in some other purposeful behavior,

such as navigation or visual search. Do the processes supporting robust picture memory proceed unimpeded outside the focus of attention? If not, what role might attention play? In this paper we address the role that visual selective attention plays in memory for pictures.

In order to answer that question, it would be very helpful if we knew what is being recognized when we recognize a scene as in a picture memory task. However, the nature of the mental representation formed after briefly viewing a scene is not fully understood. A secondary goal of this work is to add some insight into the nature of the representations that support picture memory. Our initial subjective experience of a scene is of a very detailed representation. As soon as the stimulus is removed, however, that representation begins to deteriorate. Work on change blindness shows just how fragile that representation must be (Pashler, 1988; Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997). Observers are very bad at recognizing substantial changes in scenes, even a very brief moment after that change has occurred unless they happened to be attending to the item that was changed. Indeed, Wolfe, Reinecke, and Brawn (2006) showed that observers can be very poor at detecting

* Corresponding author. Fax: +1 617 277 2085.
E-mail address: wolfe@search.bwh.harvard.edu (J.M. Wolfe).

change even if attention is directed to the critical item at the moment of change.

Clearly, memory for scenes is not the equivalent of an image file stored on your disk drive. In the literature, it has been proposed that people extract and encode the gist of an image, but what exactly is meant by "gist"? Colloquially, it might be described as the brief description of the scene. "A farm with cows, a beach on a sunny day, etc." Indeed, some have tried to codify this approach by collecting those verbal descriptions (Potter, Staub, & O'Connor, 2004). While the verbal description may give us insight into the contents of the gist, it is not, itself, the gist if by gist we mean the representation that can be extracted in a brief exposure to the stimulus and used to support picture memory. We can easily imagine two scenes that might elicit the same brief description (e.g., beach on a sunny day) and yet be easily distinguished in memory. Potter et al. have shown that picture memory, at least in the short term, involves more than the semantic description.

One possibility is that the gist is a list of attended objects with, perhaps, some information about their relative positions. Some early evidence suggested that memory for a set of disorganized objects was as good as memory for objects in organized scenes (Mandler & Johnson, 1976; Mandler & Parker, 1976) and more recent work shows excellent memory for attended objects in scenes (Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001; Newell, Brown, & Findlay, 2004). Nevertheless, as with the limitations of the verbal description, it is clear that picture memory is not based on a list of objects, alone—even if that list can include large background "objects" like a beach. First, it is intuitively clear that the same list of objects might be derived from two scenes that would be easily discriminated in memory. Second, there is abundant evidence for the role of scene context in memory (e.g., Brockmole & Henderson, 2006; Hollingworth, 2006; Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006).

Oliva and her colleagues have stressed the role of low spatial frequency information and global information (Oliva, 2005; Oliva & Schyns, 1997; Oliva & Torralba, 2006) in the initial analysis of images. They have shown that semantic information can be derived from low-level features of the scene as a whole, without any need to parse the scene into objects. Thus, a simple, feed-forward analysis of spatial frequency components in a scene can be enough to derive labels like "beach" or "street scene." It is unlikely that this representation can successfully support persistent memory for *this* beach or *this* street, but the spatial envelope can provide a different sort of information about the gist of a scene.

Other low-level features, such as mean object size, can be extracted from visual images. A number of studies have shown that observers are capable of extracting the statistical properties of meaningless textures (Ariely, 2001; Chong & Treisman, 2003; Chubb, Econopouly, & Landy, 1994; Chubb & Landy, 1994). There is some evidence for some memory for these properties (e.g Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) but there have not been extensive studies of memory for properties like the orientation mean or the brightness variance in an image, especially in meaningful scenes.

There is no reason to assume that the remembered representation of a scene needs to be limited to one of these types of information. In this paper, we explore the possibility that the gist of a scene contains at least three components: (1) a set of *statistical* properties including the distribution of basic features like colors and orientations, (2) structural information about the *layout* of the scene (as in Oliva & Torralba's (2001) spatial envelope) and (3) a set of *objects*, which might be relatively small if the scene is presented briefly.

In this paper, we ask if memory for any or all of these elements is modulated by visual selective attention. By "visual selective attention," we mean that aspect of attention that is required when observers perform inefficient visual search tasks (Wolfe, 1998). Some basic feature information can be found while attention is occupied by a search task (Braun, 1994, 1998; Braun & Julesz, 1998). There is a limited ability to determine if a type of object (e.g., an animal) is present under these circumstances (Li, VanRullen, Koch, & Perona, 2002; VanRullen, Reddy, & Koch, 2004). Finally, it has been proposed that spatial layout/spatial envelope information does not require selective attention (Torralba et al., 2006). Thus, is it possible that all of the building blocks of scene memory might be encoded even if selective attention is otherwise engaged.

Accordingly, we had participants perform a visual search task during picture encoding. In three experiments, participants studied meaningful scenes and meaningless texture images under single-task conditions or while performing a concurrent task. Since we were particularly interested in the role of visual selective attention, as opposed to more central attentional limitations, the concurrent task could be either a visual search task or an auditory detection task. Previous research has shown that auditory shadowing can interfere with picture memory (Allport, Antonis, & Reynolds, 1972; Rollins & Thibadeau, 1973; Rowe & Rogers, 1975). The comparison between the visual dual-task and the auditory dual-task allows us to determine if diverting *visual* selective attention has an effect that is greater than a general dual-task cost.

In the three experiments reported here, we found a dual-task cost on picture memory. This cost was greater when the concurrent task was visual search, as compared to the auditory task. This was true even when the visual and auditory tasks were equated for difficulty (Experiments 2 and 3). Interference was observed for all stimulus classes, suggesting that encoding of all components of gist is modulated by attention. Furthermore, dual-task interference was obtained even when the concurrent task was only performed after the picture was removed (Experiments 3), indicating that visual selective attention is involved during consolidation as well as (or perhaps instead of) encoding.

2. Experiment 1

In this experiment, we assessed picture memory for a variety of stimulus materials with and without a concurrent attention task.

2.1. Method

2.1.1. Participants

Fourteen naïve observers were recruited from the Brigham and Women's Hospital Visual Attention Laboratory volunteer pool. Observers (4 males, 10 females) ranged in age from 18 to 50 years (mean = 29.3 years). Each participant passed the Ishihara test for color blindness and had normal or corrected to normal vision. All participants gave informed consent and were paid $10.00 for their time.

2.1.2. Stimuli and apparatus

In this study, we used four types of scenes: (1) Scenes, as they are typically defined in the literature: meaningful pic-tures of real places, indoor and outdoor. Such scenes have feature statistics, a spatial layout, and objects; (2) Texture images. These have feature statistics. However, they lack objects, unless the texture, itself, is an object, and they all have roughly the same flat, frontal layout making layout information useless; (3) Shuffled scenes. We disrupted the spatial layout aspect of scenes by randomly shuffling blocks of the image (Biederman, 1972; Biederman, Glass, & Stacy, 1973; Biederman, Rabinowitz, Glass, & Stacy, 1974). This created a set of images with objects (albeit fragmented in some cases) and feature statistics but significantly reduced layout information (local regions could be used to infer some scene structure but the global structure or spatial envelope was disrupted); (4) Shuffled textures. Applying the block shuffling procedure to the texture stimuli simply created another set of textures with a block-like structure. Examples of all four types of stimuli are shown in Fig. 1.

Stimuli were presented on a 21 in. monitor set to a resolution of 1024 by 768 at a refresh rate of 75 Hz, and controlled by a Macintosh G4 computer running Mac OS 9.2.2.
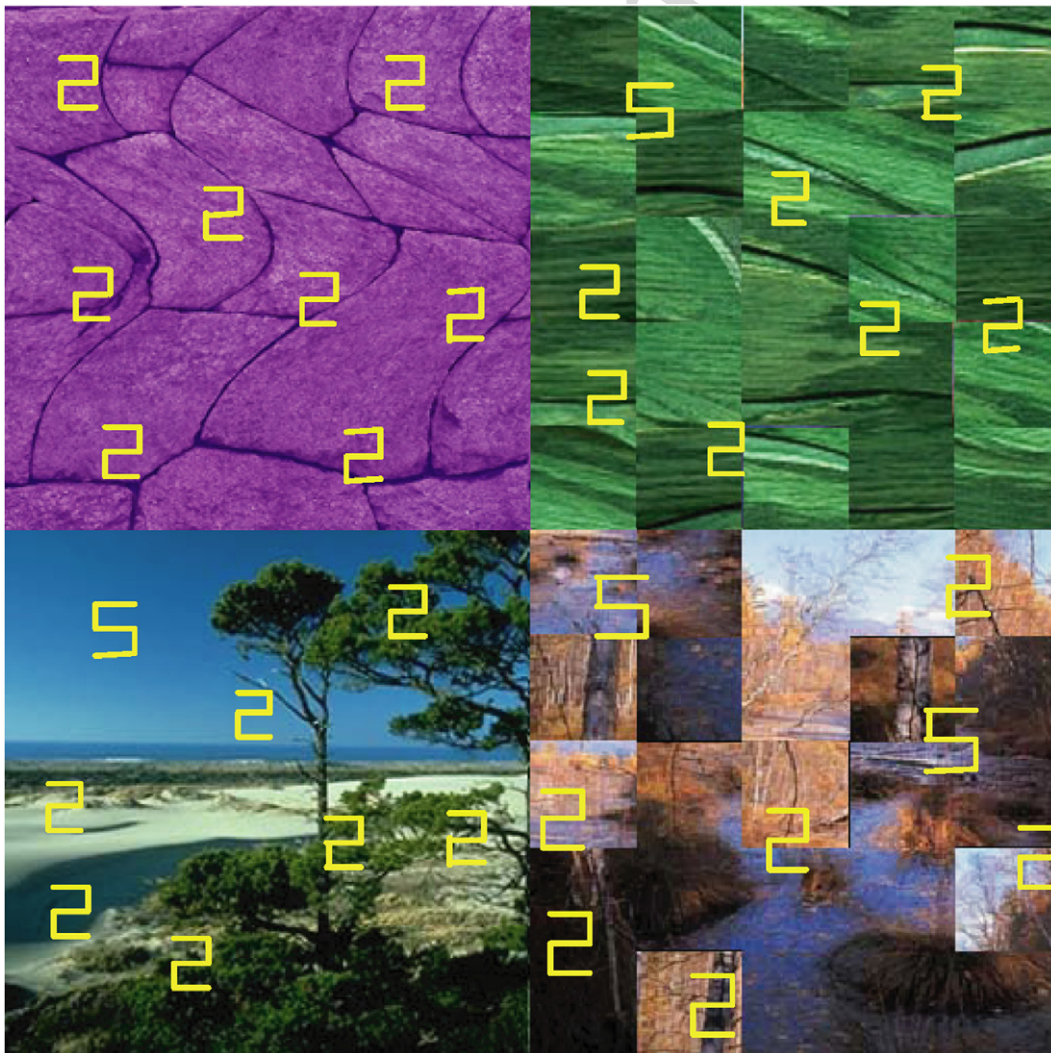

Fig. 1. Stimulus types: top row whole and scrambled textures. Bottom row: whole and scrambled scenes. The 2 s and 5 s are the stimuli for the concurrent visual search task.

The experiment was programmed in Matlab 5.2.1 (The MathWorks) using the Psychophysics Toolbox routines (Brainard, 1997; Pelli, 1997). Participants were seated 57.4 cm from the monitor; at this distance, 1 cm on the screen subtends one degree of visual angle (°). All images were presented in 8-bit color and subtended 20° by 20° (512 × 512 pixels).

Four sets of visual stimuli were used for this experiment; these were derived from 370 scenes and 340 textures. Scenes were taken from the VisTex database (http://www.white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html, the Oliva & Torralba (2001) database and from the photographic works of Hans Hendriksen (http://www.hanshendriksen.net). Texture images were chosen from the photographic works of Steven Duty and Ulfar Harris Eliasson (http://icestory.com). Shuffled versions of each image were created by dividing the images into a 5 × 5 grid and randomly rearranging the sectors. In all conditions, a search array of eight yellow digital 2s and 5s, each subtending 0.8 degrees in width and 1.8 degrees in height, was superimposed over the image (see Fig. 1). The search stimuli were placed randomly on an invisible jittered 5 × 5 grid. There were 0, 1, or 2 5s in each array.

When an auditory task was added, three 50 ms tones were presented on each trial. High tones had a frequency of 800 Hz and low tones had a frequency of 500 Hz.

2.1.3. Procedure

This experiment consisted of 14 conditions. Twelve were produced by crossing the four stimulus classes (intact scenes, shuffled scenes, intact textures and shuffled textures) with the three task sets (single-task picture memory, dual-task visual search, and dual-task auditory discrimination).

The other two conditions measured baseline (single task) accuracy in the visual search and auditory discrimination; tasks shuffled scene images not used in other conditions were used as the background for these conditions. Each participant completed all fourteen conditions in a randomly generated order.

Except for the visual search and auditory discrimination baseline conditions, each condition consisted of 32 training trials followed by 32 memory test trials. For each of the 32 training trials, the observer viewed an image for 500 ms, followed by a response screen in the dual-task conditions, or a blank interval in the single-task, picture memory conditions. For each of the 32 memory test trials, participants were asked to indicate whether the presented picture was new or old. Sixteen of the images in this phase had been presented in the training phase and 16 were novel images. Order of presentation was randomized. If a picture was used in either the training or test phase of one condition, it was never shown again in any of the remaining conditions. For the baseline conditions, the test phase was omitted. A diagram of the experimental sequence for the dual-task conditions is shown in Fig. 2.

In all conditions, a 2 versus 5 search array was superimposed over the pictures during the training phase. In the picture memory alone condition, participants were instructed to ignore the search array, and to try to remember as much about the pictures as possible. For the dual-task conditions, participants were instructed to give priority to the search or auditory task. In the visual search dual-task conditions, participants were asked to count the number of 5s presented on each training trial. After viewing each picture, participants were asked to indicate the number of 5s using the number keypad. There were no target 5s on 45% of the
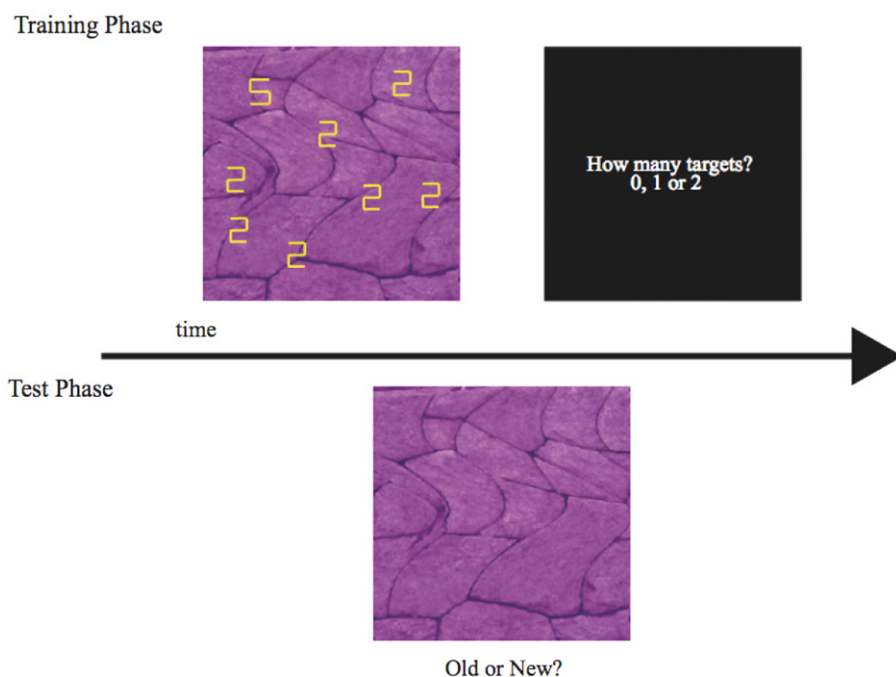

Fig. 2. Stimulus and task for Experiment 1. This is an example of the texture stimuli.

trials, one target on 45% of the trials and two targets on 10% of the trials. In the auditory dual-task conditions, a three tone sequence, lasting 500 ms, was presented simultaneously with the image. The task was to report the number of high tones played, which could be zero, one or two in the proportions give above. The termination of the last tone coincided with the image being removed from the screen. As in the visual search, participants used the numeric keypad to indicate the number of targets (here, high tones).

In the search and tone baseline conditions, participants performed 32 trials of the search or tone task while ignoring the scene. Participants were given two practice trials at the start of each condition, so they could familiarize themselves with the task.

2.2. Results and discussion

Fig. 3 shows the error rates on the concurrent tasks: search and tone discrimination. Participants were consistently more accurate on the tone task than on the visual search task but performance was quite good overall (chance would yield 58.5% error), indicating that participants were following instructions and attending to the concurrent tasks. Note that we wanted a task that would keep performance well below 100% correct because we wanted observers to be searching (or listening) throughout the picture presentation. Otherwise, observers might finish the primary task and have free time to devote to the encoding of the scene. The error rates indicate that participants were performing at the same level in the dual-task conditions as in the baseline conditions. For the search task, performance is actually somewhat better in the dual-task conditions, while for the auditory task, baseline performance is in the middle of the narrow range of performance in the dual task conditions.

Picture memory recognition performance is shown in terms of d' for intact images in Fig. 4a and for shuffled images in Fig. 4b. In analyzing the dual task conditions, we excluded all dual-task trials in which errors were made on the concurrent task.

Consider, first, the single-task picture memory data. There was a strong effect of stimulus type on performance
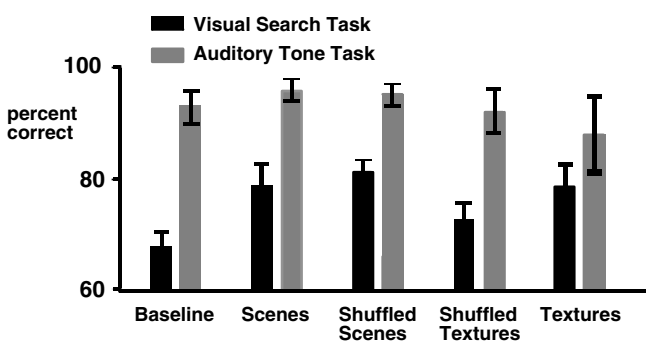

Fig. 3. Performance (percent correct) on the concurrent tasks. Chance performance would produce 58.5% errors.
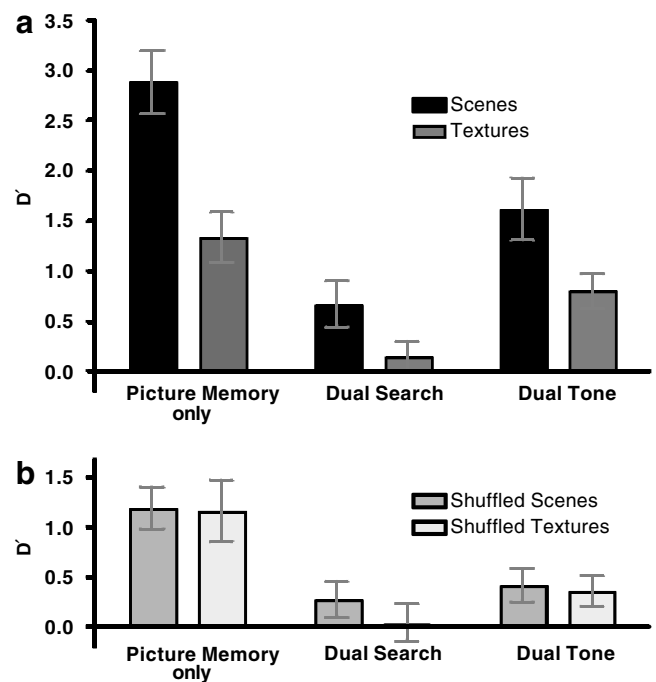

Fig. 4. Sensitivity (d′) in the picture memory test as a function of stimulus type and dual task condition.

(ANOVA $F(3,52) = 9.4$, $p < .001$). Post hoc comparisons show that unshuffled scenes were remembered better than any other stimulus type (Bonferroni corrected t-tests, all $p < .001$). No paired comparisons between texture, shuffled scenes, and shuffled textures approach statistical reliability.

Clearly, there is information available in intact scenes which is not available in any of the other three image types. We identified three types of information that may be useful: features, objects, and layout. It is highly unlikely that feature statistics can account for the performance differences, since all image types contain useful featural information. Scenes also vary widely in the number and type of objects present, while textures typically contain a single object (the texture) or at most a collection of highly similar objects (e.g., the flagstones in the upper left image in Fig. 1). However, the shuffled scenes also contain objects. Some of these objects may have been arbitrarily fragmented by the shuffling operation. Nevertheless, many objects in the shuffled scenes are clearly visible, even if broken. Shuffled scenes may have somewhat less object information than intact scenes, but they have far more object information than the texture images. Yet there was no difference in performance between shuffled scenes and textures. This outcome suggests that objects play very little role in picture memory when those objects are not situated in a meaningful spatial layout.

Finally, scenes vary widely in their layout or spatial envelope. Textures have some layout, but for the most part they all have the same 2D frontal plane layout, and a shallow spatial envelope, while the layout information in the shuffled scenes is substantially disrupted. Differences in

available layout or spatial envelope information, therefore, would seem to explain the differences in memory performance among the four image classes.

Turning to the effects of the concurrent tasks, we performed an ANOVA with three factors: dual task (none, visual, auditory), stimulus type (scene, texture), and stimulus status (normal, shuffled). There was dual task interference with memory performance for all stimulus types (One-way ANOVAs on each stimulus type, all $F(2,39)$ > 6.6, all $p < .003$). In all but one case, Bonferroni corrected t-tests found a significant decrease in performance between the single task and each of the dual-task conditions. The one exception was that the auditory dual task did not significantly interfere with memory for textures ($p = .19$). This argues for a role of attention in the encoding of scenes into memory. Moreover, diverting attention has an impact even on memory for the texture stimuli, which are presumably remembered on the basis of feature statistics since those statistics would differentiate between textures far more effectively than any minimal layout or object information.

While these results show a clear dual-task cost on picture memory scenes and textures, they do not clearly indicate whether visual selective attention is specifically required for normal picture memory. Performance on the picture memory recognition task was better in the auditory dual-task condition than in the visual dual task condition, (main effect of concurrent task in an ANOVA comparing just the concurrent search and tone tasks: $F(1,3) = 7.4$, $p < .005$). However, since the auditory task was apparently easier than the visual search task, reduced interference by the auditory task might only reflect its less demanding nature. Accordingly, we replicated the experiment with an effort to equate the dual tasks for difficulty.

3. Experiment 2

In this experiment, we sought to equate the visual search and auditory tone discrimination tasks in difficulty. Prior to the study phase of the experiment, participants performed a short tone discrimination calibration experiment. We varied the difficulty of the auditory task by varying the total number of tones. Since the mean error rates for the visual search control conditions were approximately 30% in Experiment 1, we selected the level of difficulty for each observer that would lead to roughly 30% errors on the tone task. With difficulty for search and tone tasks thus equated, we then replicated the design of Experiment 1, with the exception of the shuffled stimuli.

3.1. Method

Fourteen naïve observers (6 males, 8 females) were recruited as described above, ranging in age from 18 to 53 years (mean = 28.9 years). Apparatus and stimuli were as described previously, though we did not use the shuffled stimuli.

3.1.1. Procedure

This experiment began with a 50 trial auditory discrimination session. In separate blocks, we varied the total number of tones presented: 3, 4, 5, 6. Participants were asked to report the number of high tones in a sequence, which varied from three fewer than the total number of tones to one fewer (e.g., with 6 total tones, there could be 3, 4, or 5 targets). We then selected the total number of tones for each observer that yielded performance closest to 30% error. This value was used for that observers' tone task throughout the remainder of the experiment.

In the main portion of the experiment, images were presented for 500 ms with a superimposed search array, as in Experiment 1. In the dual task conditions, participants counted either 5 s or high tones. In the single task picture memory conditions participants rated the "likeability" of the scene or texture on a scale from 0 ("not at all") to 2 ("liked it a lot"). This ensured that the temporal structures of the single and dual-task conditions were equated. Since we did not use shuffled images in this experiment, there were six conditions (scene/texture X dual-task tone, dual-task search, or single task).

3.2. Results and discussion

The manipulation of the difficulty of the tone task was a success. Error rates for visual search and tone discrimination were not significantly different from each other (t (13) < 1.0). As in previous experiments, whenever an observer made an error on the concurrent task in the dual task conditions, the image presented on that trial was excluded from the picture memory analysis for that observer. Picture memory performance as a function of image type and dual-task condition is shown in Fig. 5.

The results of Experiment 2 closely replicated the results of the comparable conditions of Experiment 1. A 2-way ANOVA reveals a main effect of stimulus type: scenes are remembered better than textures ($F(1,14) = 53.9$, $p < 0.001$) and a main effect of dual-task ($F(2,28) = 102.2$, $p < .001$). There is no interaction between the two ($F(2,28) = 0.2$, $p = .77$). Diverting attention disrupted picture memory.
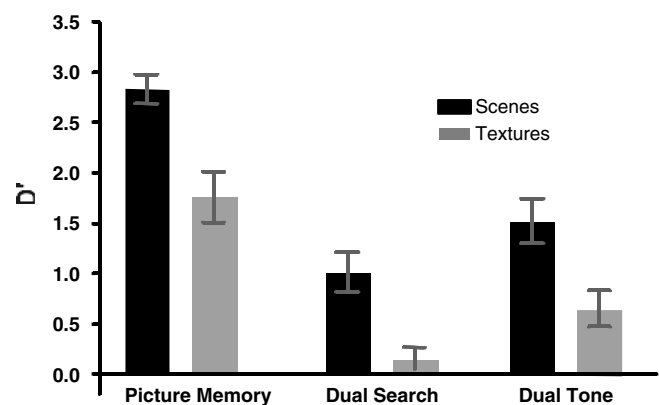

Fig. 5. Results of picture memory test in Experiment 2.

The goal of Experiment 2 was to determine whether the effect of the concurrent task had a specifically *visual* component. With the two concurrent tasks equated for difficulty, we can test this hypothesis more convincingly than in Experiment 1. A 2×2 ANOVA restricted to the dual task conditions revealed that the search task did, indeed, produce a larger effect than the tone task ($F(1,14) = 16.8$, $p < .001$). There was also a main effect of stimulus type: scenes were remembered better than textures under dual task conditions ($F(1,14) = 23.6$, $p < .001$). There was no interaction ($F(1,14) = 0$). These results support the conclusion that diverting visual selective attention interferes with picture memory. Furthermore, this interference is comparable for scenes, possessing objects and spatial layout, and for textures, that lack objects and share the same, flat, frontal plane layout information. This result was not obvious. It could have been that search interfered with attention-demanding object recognition processes but not with "parallel" assessment of image statistics. Nevertheless, whatever resources are tied up by the search task appear to be used in encoding both scenes and textures. In addition to the cost of diverting visual attention, there is also a more central dual-task cost associated with the tone task.

4. Experiment 3—interference during consolidation

Experiments 1 and 2 showed that picture memory is impaired if visual attention is occupied during the initial study of the to-be-remembered images. Successful picture memory requires that information be encoded into memory and then consolidated. Potter and colleagues (Potter, 1976, 1993; Potter & Levy, 1969) have shown that a picture can be encoded to a point that allows an observer to understand the basic content of a scene (e.g a picnic) within 150 ms. However, several hundred additional ms are required to sufficiently consolidate the memory to permit good recall after a delay. The design of Experiments 1 and 2 allowed the concurrent tasks to interfere with either encoding or consolidation. In Experiment 3, picture encoding and the concurrent task were separated in time. If the specifically visual dual-task costs in Experiments 1 and 2 were due to diversion of attention during stimulus encoding, then the cost of a visual search task during consolidation should be equivalent to that of the auditory tone task presented during consolidation.

4.1. Method

Experiment 3 followed the methods of Experiment 2. Indeed, the same 14 observers were tested in both experiments. Apparatus and stimuli were as described for Experiment 2.

4.1.1. Procedure

Like Experiment 2, this experiment began with a 50 trial auditory discrimination session that determined the num-

ber of tones used in the concurrent tone task of the main experiment. The procedure was the same as in Experiment 2 except that pictures and search arrays were temporally separated. During the training phase, images were shown for 125 ms without a superimposed search array. After 125 ms, the image was replaced by a search display of yellow 2 s and 5 s on a black background for 375 ms. In the auditory dual-task condition, the tone sequence was presented for the same duration as the visual search stimuli while participants viewed the blank screen.

For the picture memory, single task training conditions, participants were instructed to ignore the search display and try to remember as much about the picture as possible, and then to rate how much they liked the picture on a scale from 0 ("not at all") to 2 ("liked it a lot"). As in Experiment 2, this ensured that the temporal structure of the single and dual-task conditions were equated.

Two image types (scene vs. texture) crossed with three tasks (single task picture memory, dual task visual search, and dual task auditory) yielded six conditions.

4.1.2. Results

Again, the effort to equate the tone and search tasks was successful. Performance on the search task was not significantly harder than performance on the tone task. Indeed, the tone task produced slightly higher error rates (16%) than the search task (12%), but this was not statistically reliable ($t(13) = 0.92$, $p = .33$).

Picture memory results for Experiment 3 are shown in Fig. 6. There was a main effect of image type, such that scenes were remembered better than textures ($F(1,14) = 11.5$, $p < .005$), and a main effect of concurrent task ($F(2,28) = 39.7$, $p < .001$). There was no significant interaction ($F(2,28) = 2.5$, $p = 0.09$). Since we deliberately equated concurrent task difficulty in Experiment 3, we can directly compare the effects of tone and search concurrent tasks. In a 2×2 ANOVA on just the dual-task conditions, the effect of visual search was reliably larger ($F(1,14) = 8.8$, $p < .01$). The effect of concurrent task did not interact with the type of search stimulus ($F(1,14) = 1.2$, $p = .3$).
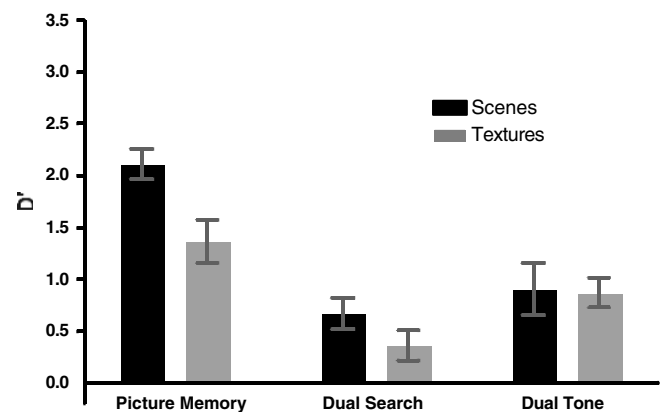
Fig. 6. Picture memory results for Experiment 3.

5. General discussion

The three experiments described here show that picture memory is not independent of the demands of visual spatial attention. In each experiment, the requirement to perform a visual search task reduced d' by more than 50%. Some portion of this loss of sensitivity can be attributed to a general dual task cost, since a concurrent tone identification task also reduced d'. However, in each of the experiments, the effect of the visual search task was reliably greater than the effect of the auditory tone task. In Experiment 1, this might be attributed to differences in task difficulty, but this explanation cannot account for Experiments 2 and 3, in which the difficulties of the two concurrent tasks were matched. Therefore, it is reasonable to conclude that attention to the visual search stimulus withdraws some resource from the processing of scenes and textures that would otherwise contribute to memory for those pictures. In a similar, complementary result, Allport et al. (1972) found that auditory shadowing interfered more with recognition of auditorily presented words than pictures.

It is unlikely that the visual search stimulus produced simple visual interference by obscuring the pictures. First, the search array was also present in the control and tone task conditions, in which participants did not need to attend to the search array. Second, in Experiment 3, the search array was presented only after the visual stimulus was removed, yet the concurrent search task still had a greater effect than the concurrent tone task.

We had hoped that visual selective attention would have differential effects on memory for different types of images. In the introduction, we noted that the gist of an image might be divisible into features statistics, layout, and object information. The texture images have very similar, flat layouts and essentially no object information. The shuffled scenes of Experiment 1 have substantially disrupted layout and some objects may be fragmented. It always seemed likely that diverting visual selective attention would interfere with object encoding, since most object recognition seems to demand attention (Biederman, Blickle, Teitelbaum, & Klatsky, 1988; Stankievich & Hummel, 1995; Treisman, 1988). However, it seemed possible that feature statistics and/or layout might reach memory via a nonselective pathway (Wolfe, 2006) that would not be disrupted by a search task. Some aspects of scene processing are not disrupted by concurrent visual search. For example, if shown a scene, observers are likely to believe that they have seen a wider angle view than was actually presented ("boundary extension", Intraub & Berkowits, 1996; Intraub & Richardson, 1989). This tendency to construct the world beyond the boundaries of the current stimulus is not disrupted by concurrent visual search (Intraub, Daniels, Horowitz, & Wolfe, 2006). If memory for feature statistics were similarly immune to the effects of concurrent search, we might have found that the visual search task interfered with memory for scenes but not with memory for textures. More realistically, given that any dual task would produce some interference, we might have found that the visual and auditory tasks would have the same effect on memory for textures but that the visual task would have had a greater effect on memory for scenes.

We found no such result. The visual search task had an effect on memory for all types of images used in these experiments and that effect was greater than the effect of the tone task. There are, at least, three possible accounts of this null result (beyond the ever-present possibility that it is merely a null result waiting for a more powerful experiment). First, our analysis of picture memory could be wrong. Picture memory might not be memory for objects, layout, and feature statistics. Second, textures and scenes might not be qualitatively different from the vantage point of picture memory. The visual system might find something like objects and layout in textures even if we, the experimenters, thought otherwise. Finally, it might be that the visual search concurrent task interfered with some visual processing stage beyond selection. There is evidence for at least two attentional bottlenecks in the visual pathway. One would be the selective bottleneck apparently involved in object recognition. The other would be a later bottleneck into visual working memory that produces phenomena like the attentional blink (Chun & Potter, 1993; Raymond, Shapiro, & Arnell, 1992). Evidence from priming studies shows that the blink bottleneck lies after object recognition, since objects that are not reported due to the blink nevertheless produce priming based on their recognized identity (Luck, Vogel, & Shapiro, 1996; Shapiro, Driver, Ward, & Sorenson, 1997). The bottleneck involved in the attentional blink seems to affect all types of visual processing (e.g., Marois, Yi, & Chun, 2004). If the search task invoked the blink bottleneck as well as the earlier selective bottleneck tied to object recognition, then it might not be surprising that visual search disrupted memory for all types of images used here. This hypothesis is consistent with the fact that visual search disrupted consolidation in Experiment 3.

The main conclusion of this work remains clear. Scenes are not coded into memory via some pathway that is immune from competing demands of other visual tasks. To effectively remember a scene, you need to attend to that scene.

Acknowledgment

References

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. Quarterly. *Journal of Experimental Psychology, 24*(2), 225–235.

Ariely, D. (2001). Seeing Sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(43), 77–80.

Biederman, I., Blickle, T. W., Teitelbaum, R. C., & Klatsky, G. J. (1988). Object search in nonscene displays. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*(3), 456–467.

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

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W., Jr. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103*(3), 597–600.

Braun, J. (1994). Visual search among items of different salience: Removal of visual attention mimics a lesion of extrastriate area V4. *Journal of Neuroscience, 14*(2), 554–567.

Braun, J. (1998). Divided attention: Narrowing the gap between brain and behavior. In R. Parasuraman (Ed.), *The attentive brain* (pp. 327–351). Cambridge, MA: MIT Press.

Braun, J., & Julesz, B. (1998). Dividing attention at little cost: detection and discrimination tasks. *Perception and Psychophysics, 60*(1), 1–23.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 443–446.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13*(1), 99–108.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404.

Chubb, C., Econopouly, J., & Landy, M. S. (1994). Histogram contrast analysis and the visual segregation of IID textures. *Journal of the Optical Society of America A, 11*, 2350–2374.

Chubb, C., & Landy, M. S. (1994). Orthogonal distribution analysis: A new approach to the study of texture perception. In *Computational models of visual processing* (pp. 291–301). Cambridge, MA: MIT Press.

Chun, M. M., & Potter, M. C. (1993). Interference in detecting multiple targets in a sequence: A dissociation between the attentional blink and repetition blindness. *Investigative Ophthalmology and Visual Science, 34*(4), 1232.

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 58–69.

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*(1), 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*(4), 761–768.

Intraub, H., & Berkowits, D. (1996). Beyond the edges of a picture. *American Journal of Psychology, 109*(4), 581–598.

Intraub, H., Daniels, K. K., Horowitz, T. S., & Wolfe, J. M. (2006). Looking at scenes while searching for numbers: Dividing attention multiplies space. *Paper presented at the Visual Sciences Society, Sarasota, FL.*

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*(2), 179–187.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9596–9601.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 382*, 616–618.

Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology [Hum Learn], 2*(5), 529–540.

Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology [Hum Learn], 2*(1), 38–48.

Marois, R., Yi, D.-J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron, 41*, 465–472.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614–621.

Newell, F. N., Brown, V., & Findlay, J. M. (2004). Is object search mediated by object-based or image-based representations? *Spatial Vision, 17*(4-5), 511–541.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–257). San Diego, CA: Academic Press/Elsevier.

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*(1), 72–107.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. J. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–745.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics, 44*, 369–378.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509–522.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21*(2), 156–161.

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

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*(3), 478–489.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849–860.

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.

Rollins, H. A., & Thibadeau, R. (1973). The effects of auditory shadowing on recognition of information received visually. *Memory & Cognition, 1*(2), 164–168.

Rowe, E. J., & Rogers, T. B. (1975). Effects of concurrent auditory shadowing on free recall and recognition of pictures and words. *Journal of Experimental Psychology: Human Learning & Memory, 1*(4), 415–422.

Schyns, P. G., & Oliva, A. (1994). From Blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5*(2), 195–200.

Shapiro, K., Driver, J., Ward, R., & Sorenson, R. E. (1997). Priming from the attentional blink. *Psychological Science, 8*(2), 95–100.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261–267.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73–74.

Stankievich, B. J., & Hummel, J. E. (1995). The role of attention in priming of left–right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology: Human Perception & Performance.*

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes:

The role of global features on object search. *Psychological Review, 133*(4), 766–786.

Treisman, A. (1988). Features and objects: The 14th Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, 40A*, 201–237.

VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience, 16*(1), 4–14.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–74). Hove, East Sussex, UK: Psychology Press Ltd.

Wolfe, J. M. (2006). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems.* New York: Oxford.

Wolfe, J. M., Reinecke, A., & Brawn, P. (2006). Why don't we see changes? The role of attentional bottlenecks and limited visual memory. *Visual Cognition, 19*(4-8), 749–780.