VISUAL COGNITION, 2009, 17 (1/2), 273–291

Psychology Press
Taylor & Francis Group

Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location

Andrew Hollingworth
Department of Psychology, University of Iowa, Iowa City, IA, USA

The role of scene memory in visual search was investigated in a preview-search task. Participants saw a preview of a real-world scene prior to searching through that scene for a target object. Compared with a no preview baseline, search was facilitated following a preview scene that did not contain the ultimate target object, indicating that memory for the general context and layout of a scene exemplar facilitates later search. In addition, search was more efficient when the preview scene contained the ultimate target object than when it did not, demonstrating that memory for the specific binding of objects to locations facilitates later search. Both forms of memory were explicitly available for participant report.

Keywords: Visual search; Visual memory; Scene memory; Eye movements; Attention.

Recent research has demonstrated that people can generate robust visual memory representations of natural environments, retaining information about the visual features of objects in a scene (Castelhano & Henderson, 2005; Hollingworth, 2004, 2005b; Hollingworth & Henderson, 2002; Melcher, 2006; Tatler, Gilchrist, & Land, 2005; Tatler, Gilchrist, & Rusted, 2003) and the locations of those objects (Brockmole, Castelhano, & Henderson, 2006; Brockmole & Henderson, 2006a, 2006b; Hannula, Tranel, & Cohen, 2006; Hollingworth, 2005a, 2007). Such capability raises the question of the function of this visual memory: For what real-world, behavioural purposes is visual memory for natural scenes utilized?

Please address all correspondence to Andrew Hollingworth, Department of Psychology, University of Iowa, 11 Seashore Hall E, Iowa City, IA 52242-1407, USA.
E-mail: andrew-hollingworth@uiowa.edu

The present research was supported by NIH grants R03MH65456 and R01EY017356. Parts of this study were reported at the Munich Visual Search Symposium, Munich, Germany, 2003, and the fourth annual meeting of the Vision Sciences Society, Sarasota, Florida, 2004.

http://www.psypress.com/viscog DOI: 10.1080/13506280802193367

One plausible function for scene memory is the guidance of visual search through a scene (Brockmole & Henderson, 2006b; Chun & Jiang, 1998; Henderson, Weeks, & Hollingworth, 1999). Almost all real-world tasks require visual search of some sort. For example, when cooking a meal, one must search for many of the items (ingredients, utensils) that are necessary to complete the task (Land & Hayhoe, 2001). More generally, almost every behavioural operation requires attending to (and typically fixating) the target of the behaviour. If the relevant target object is not attended beforehand or does not recruit attention to itself directly, a visual search operation is necessary to orient attention and the eyes to that object. Thus, visual search is a fundamental process underlying object selection, and understanding visual search in real-world environments is therefore critical to understanding how humans are able to behave intelligently in the service of top-down goals (such as making a meal).

Memory for the visual form of objects could facilitate search by specifying the perceptual features that discriminate the target from distractors (Desimone & Duncan, 1995; Hollingworth, Richard, & Luck, 2008). Search for a pen is likely to be more efficient if one remembers that the pen is blue, as search can be limited to blue items. Perhaps more importantly, memory for the spatial properties of scenes could guide attention efficiently to a location that is known to, or is likely to, contain the relevant object.

Research over the last decade has probed various circumstances under which memory for the spatial properties of scenes guides visual search. First, memory for scenes and objects in general (e.g., that pedestrians tend to appear on pavements and pictures on walls) can guide attention and the eyes to locations within a scene where the target object is likely to appear (Henderson et al., 1999; Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006), even for scenes that participants have never viewed before. Second, repeated search through a particular scene exemplar generates more efficient search when the position of the target object remains constant, both for naturalistic scene stimuli (Brockmole et al., 2006; Brockmole & Henderson, 2006a, 2006b; Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006) and for scenes composed of randomly arrayed collections of simple objects (Chun & Jiang, 1998). Finally, a brief preview of a scene facilitates subsequent search through that scene when the search stimulus is degraded so that only the currently fixated portion of the scene is visible (Castelhano & Henderson, 2007).

My goal in the present study was to examine the means by which visual memory for a particular scene exemplar facilitates subsequent search through that scene. The present experiments served to isolate two forms of scene memory that might plausibly guide search: (1) Memory for the general context and layout of the scene and (2) memory for the particular binding of target object to scene location. Memory for the context of the scene (e.g., that the kitchen counter extends across the left side of the image) combined with knowledge of the target object (that toasters tend to appear on kitchen counters) could serve to direct attention to the region of the scene where the target is likely to appear (the left side). In addition, one's memory for the actual, observed location of the target could serve to guide attention directly to that location in the scene.

To tease apart these two mechanisms, I used a preview-search paradigm, manipulating scene information contained in the preview before search through that scene. The basic method is illustrated in Figure 1. Participants saw a preview image of a scene (or no preview image in the control condition). This was followed by a target-probe object, which was displayed in isolation in the centre of the screen, indicating the object that was to be found during search. Next, participants saw a search scene. The target object was always present in the search scene, and it was either identical to the target probe or mirror reversed. Participants' task was to find the target in the search scene as quickly as possible and respond to indicate whether it was the same as the target probe or mirror reversed. This forced participants to search through the scene and fixate the target, because they needed to acquire fairly precise information from the target to assess its left/right (l-r) orientation. Search efficiency was measured using standard search RT data and eyetracking data.

There were three preview conditions. In the *no preview* condition, participants did not see a preview of the scene prior to search. In this case, participants could have used knowledge about scenes and objects in general to guide search (for example, directing attention and the eyes to the lower regions of the exercise room scene where a pair of sneakers were likely to appear; Torralba et al., 2006), but they did not have an existing visual memory representation of the particular scene exemplar through which they were searching. In the *target-absent preview* condition, participants saw a preview of the scene, but it did not contain the ultimate target object. In this case, participants could remember the general context and layout of the preview scene. If search were more efficient in the target-absent preview condition than in the no preview condition, this would indicate that participants can use memory for scene context to guide attention and the eyes to plausible target locations. Finally in the *target-present preview* condition, participants saw a preview of the scene, and the preview contained the target object in the same location as it would appear in the search scene. When the target appeared in the preview scene, its l-r orientation was the same as the target probe but was uncorrelated with the orientation of the target in the search scene. Thus, participants could not respond on the basis of memory for the orientation of the target in the preview scene; they were forced to find the target in the search scene. If search were more efficient in the target-present preview condition than in the target-absent preview

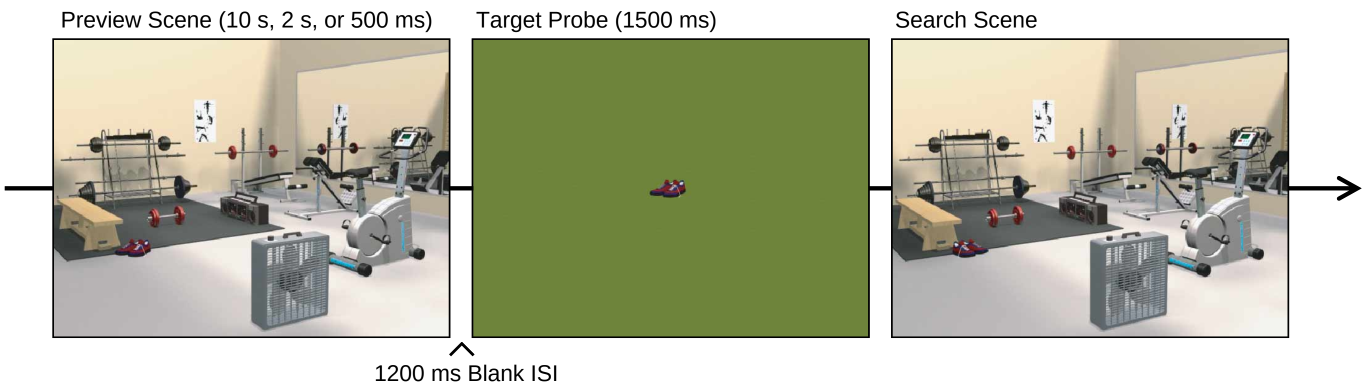

Figure 1. Sequence of events in a trial. Participants saw a preview of a scene (pictured) or no preview (not pictured), followed by a target probe and a search scene. Participants searched through the scene to find the target object and responded to indicate whether its orientation in the search scene was the same as the orientation of the target probe (in this example, the correct response was “no”). To view this figure in colour, please see the online issue of the Journal.

condition, this would indicate that memory for the specific binding of target object to scene location facilitates search, as the two conditions differed only in the presence of the target in the preview scene.

The present study complements and extends existing research on the role of scene memory in visual search. Studies examining the phenomenon of contextual cueing have already demonstrated that memory for a particular target location can facilitate search (Brockmole & Henderson, 2006b; Chun & Jiang, 1998). However, contextual cueing depends on repeated search, and in the standard version of the contextual cueing task, significant facilitation is observed only after multiple repetitions of a particular search context. In the present study, the effect of scene memory was examined on the very first search through the scene. Because participants had no previous search trials for a particular scene item, they could not know the location of the target object during learning, and therefore the best strategy was to encode the locations of multiple objects from the preview scene. If robust facilitation were observed in the target-present preview condition, this would indicate that participants encoded a scene representation that maintained the binding of multiple object exemplars to scene locations (Hollingworth, 2007) and that could be used flexibly to facilitate search depending on the particular object that was the ultimate target of search. Such a result would also demonstrate single-trial learning of scene information functional in search, in contrast with the incremental learning over multiple trials of repetition in traditional contextual cueing studies.

In addition, an advantage in the present study for the target-present preview condition over the target-absent preview condition requires memory for the *binding* of particular objects to scene locations. In contrast, contextual cueing requires only memory for location. In contextual cueing tasks, the identity of the target varies across repeated search trials (as do the identities of the distractors, typically), and thus the information functional in facilitating search is limited to spatial properties of the search arrays and the target location. In the present study, however, learning only the spatial properties of the scene would be insufficient to generate an advantage for the target-present preview condition over the target-absent preview condition, because the two conditions were identical in their spatial properties (except for the presence of one additional occupied location in the former condition). In addition, learning the visual forms or identities of the objects in the preview scene without remembering their locations would also be insufficient to generate an advantage for the target-present preview condition. The 1500 ms presentation of the target probe ensured that, in all conditions, participants had an accurate search target template. And memory for the l-r orientation of the target in the preview scene could not influence performance directly, because the orientation of the target in the preview and search scenes was uncorrelated. Thus, an advantage for the target-present preview condition over the target-absent preview condition could plausibly arise only if participants remembered the specific binding between the visual properties and/or identity of object that was to be the target and its location within the scene (Hollingworth, 2005a, 2007).

Finally, the present preview manipulations are similar to those used in a study by Castelhano and Henderson (2007). In that paper, participants saw a brief, 250 ms preview of a scene, followed by a label describing the target object and a search scene. Participants' task was to orient the eyes to the target object and fixate it. Castelhano and Henderson were interested in the effects of a scene preview independently of the visual information that could be extracted from the scene during search. Thus, during search, only a very small portion of the search scene was visible, corresponding roughly to the foveal region at the centre of gaze. Although this "moving window" approach was necessary to answer Castelhano and Henderson's central research question, this method has limited applicability to understanding search under natural conditions, in which one has access to information across the visual field. In particular, degraded perceptual information during search may have caused participants to depend more heavily on memory than they would have if the search scene had been fully visible (Oliva, Wolfe, & Arsenio, 2004). In the present experiments, the search task could be performed solely on the basis of perceptual information acquired from the search scene, allowing for a more stringent test of the utility of memory in guiding visual search.

In a manner similar to the present study, Castelhano and Henderson (2007) included two preview conditions, one in which the preview scene contained the ultimate search target object and another in which the preview did not. A scene preview facilitated search in both of these conditions compared with a no preview control. There was a numerical trend towards more efficient search when the preview scene contained the search target than when it did not, but this effect did not approach statistical significance. Thus, one cannot determine from the Castelhano and Henderson data whether memory for the specific binding of target object to scene location generates search facilitation beyond that attributable to familiarity with the general context of the scene.

The present study consisted of two main experiments. Experiment 1a and 1b included only the target-present preview condition and the no preview condition to establish and quantify the basic preview effect. In addition, the length of the preview was varied to determine whether participants accumulate visual information from the preview scene over extended viewing. Experiment 2 added the target-absent preview condition so as to isolate the effects of memory for general scene layout and memory for the binding of target object to scene location.

EXPERIMENTS 1A AND 1B

Two preview conditions were contrasted: Target-present preview and no preview. In Experiment 1a, the preview duration was 10 s. In Experiment 1b, the preview duration was 2 s or 500 ms. Given evidence that scene structure and identity can be extracted from a very brief view of a scene (e.g., Castelhano & Henderson, 2007; Potter, 1976), facilitation of search was expected at all preview durations. In addition, given evidence that visual memory for individual objects accumulates over the course of multiple seconds of scene viewing (Hollingworth, 2004; Hollingworth & Henderson, 2002; Tatler et al., 2003), it was expected that the amount of facilitation would be larger with longer preview duration.

Method

Participants. In all experiments, participants were recruited from the University of Iowa community. Each participant completed only one experiment. All participants reported normal, uncorrected vision. They either received course credit or were paid. Sixteen participants completed Experiment 1a, and twenty-four completed Experiment 1b.

Stimuli. Scene stimuli were rendered from 48 3-D models of real-world environments. An additional eight scenes were used for practice trials. In each scene model, a single target object was chosen. Two versions of the scene were rendered that differed only in the l-r orientation of the target object. For the target probe images, the target object was rendered at screen centre against a neutral green background (its appearance at screen centre was identical to its appearance embedded within the preview scene). Scene stimuli subtended $19.5^\circ \times 26.1^\circ$. Target objects had a mean width of 3.0° and mean height of 2.3°.

Apparatus. Stimuli were presented on a 17-inch computer monitor with a refresh rate of 120 Hz. Stimulus presentation was synchronized with the monitor's refresh cycle. Responses were collected by a serial button box. Stimulus presentation and response collection were controlled by a PC-compatible computer running E-Prime software. Eye position was monitored by a video-based, ISCAN ETL-400 eyetracker sampling at 240 Hz. A chin and forehead rest was used to maintain a constant viewing distance of 70 cm and to minimize head movement.

Procedure. The sequence of events in a trial is illustrated in Figure 1. In Experiment 1a, each trial was initiated by the experimenter after eyetracker calibration was checked. This was followed by a 1000 ms central fixation cross.

In the 10 s preview condition, the preview scene was then displayed for 10 s. In the no preview condition, instead of the preview scene, an otherwise blank screen with the words "No preview" was displayed for 2 s. Next, there was a blank, 1200 ms delay. The target-probe image (target object at screen centre) was then presented for 1500 ms. Finally, the search scene was displayed until response. Participants pressed one of two buttons to indicate whether the target object in the search scene was the same as the target probe or mirror reversed.

The sequence of events in Experiment 1b was the same, except a message appeared before the preview scene for 2 s informing the participant about the length of the preview ("2 second preview", "½ second preview", or "No preview"). The message was followed either by a preview scene for 2 s, a preview scene for 500 ms, or no preview.

Participants were instructed that they would or would not be given a preview of each scene before searching through it to find a particular object. They were told that after the preview, they would see the search target at the centre of a blank image. When the search scene appeared, they were to find this object in the scene and respond to indicate whether its orientation was the same as or different from the orientation of the target presented in isolation before the search. They were further instructed that the orientation of the target in the preview scene did not predict the orientation of the target in the search scene, and thus they must find the target in the search scene and respond on the basis of its orientation. Finally, they were instructed to respond as quickly and as accurately as possible.

When the participants arrived, they first gave informed consent, followed by instructions and eyetracker calibration. In Experiment 1a, participants completed a practice block of eight trials, evenly divided among the main conditions. Participants then completed 48 experiment trials, six in each of the eight conditions created by the 2 (10 s preview, no preview) $\times$ 2 (target same/different orientation) $\times$ 2 (initial target orientation) design.

In Experiment 1b, participants completed a practice block of six trials, evenly divided among the main conditions. Participants then completed 48 experiment trials, four in each of the 12 conditions created by the 3 (2 s preview, 500 ms preview, no preview) $\times$ 2 (target same/different orientation) $\times$ 2 (initial target orientation) design.

Each participant viewed each of the 48 scene items on one trial; there was no scene repetition. Across participants, condition-item assignments were counterbalanced by Latin square so that each scene item appeared in each condition an equal number of times. Trial order was determined randomly. The entire session required approximately 45 min.

Data analysis. Eyetracking data analysis was conducted offline using dedicated software. A velocity criterion (eye rotation $> 31°$/s) was used to define saccades. These data were then analysed with respect to the target object region. Target regions were rectangular, 0.65° degrees larger on each side than the smallest rectangle encompassing the target object.

Results

Data for Experiments 1a and 1b are reported in Tables 1 and 2, respectively. The initial orientation of the target object did not produce any reliable effects, and the analyses collapsed this factor. In addition, the match between target probe and search target (same orientation, different orientation) also did not produce any reliable effects, and the analyses collapsed this factor as well.

Eye movement analysis. The data of interest reflected the efficiency with which the eyes were directed to the target object region. Thus, trials on which the participant did not fixate the target prior to response were eliminated. In Experiment 1a, the target region was fixated prior to response on 96.2% of trials. In Experiment 1b, the target region was fixated prior to response on 95.7% of trials. Trials were also excluded from eyetracking analysis if the participant provided an incorrect manual response (see later). Altogether, 6.3% of the trials was eliminated from Experiment 1a and 6.9% from Experiment 1b.

Eye movement scan paths for one of the scene items in Experiment 1a are shown in Figure 2. The top panel depicts the scan paths during search for the eight participants who had a 10 s preview of the scene prior to search. The bottom panel depicts the scan paths of the eight participants who did not have a preview of the scene prior to search.

Elapsed time to target fixation. The primary eyetracking measure of search efficiency was the elapsed time from the onset of the search scene

TABLE 1
Experiment 1a search data

	Preview condition	
	10 s preview	*No preview*
Elapsed time to target fixation (ms)	390	591
Path ratio	1.22	1.65
Manual reaction time (ms)	1232	1487
Manual response accuracy (%)	93.3	95.0

TABLE 2
Experiment 1b search data

	Preview condition		
	2 s preview	*500 ms preview*	*No preview*
Elapsed time to target fixation (ms)	441	489	608
Path ratio	1.31	1.47	1.62
Manual reaction time (ms)	1241	1299	1539
Manual response accuracy (%)	91.1	95.3	94.9

to the beginning of the first fixation on the target region.[1] In Experiment 1a, there was a reliable effect of scene preview, $F(1, 15) = 53.8$, $p < .001$, with shorter elapsed time to target fixation in the 10 s preview condition than in the no preview condition. In Experiment 1b, there was also a reliable effect of scene preview, $F(2, 46) = 10.7$, $p < .001$. Elapsed time to first fixation in both the 2 s preview condition and 500 ms preview condition was reliably shorter than that in the no preview condition, $F(1, 23) = 28.0$, $p < .001$, and $F(1, 23) = 8.87$, $p = .007$, respectively. Elapsed time did not reliably differ in the 2 s and 500 ms preview conditions, $F(1, 23) = 1.50$, $p = .23$, but the numerical trend was towards shorter elapsed time with longer preview duration.

Elapsed time to target fixation was related to eye movement behaviour during the scene preview. The target object was fixated during the 10 s preview on 92% of trials, during the 2 s preview on 54% of trials, and during the 500 ms preview on 5% of trials. In the 5 s preview condition (which included a significant proportion of both preview-fixated and preview-unfixated trials), mean elapsed time to target fixation during search was 351 ms when the target had been fixated during study, and 570 ms when the target was not fixated during study.

Path ratio. Path ratio (Henderson et al., 1999) provides a converging measure of search efficiency. Path ratio is the length of the actual eye movement scan path from scene centre to target object fixation (summing the straight-line distances of each of the saccades) divided by the absolute, Euclidean distance from scene centre to target. A path ratio of 1 would indicate that the eyes were oriented on a direct path to the target. Path ratios greater than 1 indicate an indirect path. These data are reported in Tables 1 and 2. In Experiment 1a, there was a reliable effect of scene preview, $F(1, 15) = 41.9$, $p < .001$, with a shorter mean path in the 10 s preview condition

[1] An equivalent pattern of results was observed for the number of fixations on the scene prior to target fixation.

Figure 2. Experiment 1a. Eye movement scan paths for all 16 participants over one of the scene items. Dots represent fixations and lines saccades. The first fixation on the scene tended to lie at the centre of the image. The target object was the framed photograph, and all participants fixated the target prior to manual response. The top panel shows the scan paths of the eight participants who saw this item in the 10 s preview condition. The bottom panel shows the scan paths of the eight participants who saw this item in the no preview condition. To view this figure in colour, please see the online issue of the Journal.

than in the no preview condition. In Experiment 1b, there was also a reliable effect of scene preview, $F(2, 46) = 6.65$, $p = .003$. Path ratio was shorter in the 2 s preview condition than in the no preview condition (1.62), $F(1, 23) = 13.5$, $p = .001$, and there was a trend towards shorter path in the 500 ms preview condition than in the no preview condition, $F(1, 23) = 3.46$, $p = .08$. Finally, there was also a trend towards a shorter path in the 2 s preview condition than in the 500 ms preview condition, $F(1, 23) = 3.15$, $p = .09$.

It is useful to compare these path ratio data with those from Brockmole and Henderson (2006a). Brockmole and Henderson examined eye movements in a contextual cueing paradigm using small letter targets embedded within photographs of real-world scenes. Path ratio significantly declined with scene repetition. However, even after nine previous searches through a scene, path ratio was no lower than 1.8, on average. The significantly shorter path ratios in the present study (even in the no preview condition) are likely to have been caused by multiple factors, including: The use of relatively large, natural objects that could be detected in the periphery; the 1500 ms presentation of the target probe before search, which provided a precise visual search template; and the placement of objects in contextually appropriate locations.

Manual RT analysis. In Experiment 1a, there was a reliable effect of preview, $F(1, 15) = 31.3$, $p < .001$, with shorter mean correct RT in the 10 s preview condition than in the no preview condition. In Experiment 1b, there was also a reliable effect of scene preview, $F(2, 46) = 11.5$, $p < .001$. RT in both the 2 s preview condition and 500 ms preview condition was reliably shorter than that in the no preview condition, $F(1, 23) = 19.8$, $p < .001$, and $F(1, 23) = 8.69$, $p = .007$, respectively. RT did not reliably differ in the 2 s and 500 ms preview conditions, $F(1, 23) = 1.72$, $p = .20$, but the numerical trend was again towards shorter RT with longer preview duration.

Accuracy analysis. In Experiment 1a, there was no effect of preview on the accuracy of the manual response, $F < 1$. In Experiment 1b, there was a trend towards an effect of preview, with lower accuracy in the 2 s preview condition (91.1%) than in the 500 ms preview condition (95.3%) and no preview conditions (94.9%), $F(2, 46) = 2.73$, $p = .08$. Lower accuracy in the 2 s preview condition, which had the fastest RTs in Experiment 1b, raises the possibility of a speed–accuracy tradeoff, potentially complicating interpretation of the RT data in that condition. However, the significant effects of preview in the 500 ms and 10 s preview conditions cannot be attributed to speed–accuracy tradeoff, and these two durations framed the 2 s duration. In addition, the eyetracking data are not compromised by manual response accuracy, because they measure a discrete event, target fixation, that can be analysed independently of the participant's criterion for generating a manual response.

Discussion

Experiments 1a and 1b showed that a scene preview facilitates search even when the scene is fully visible during the search and participants could perform the task without consulting memory. These results are broadly inconsistent with claims that participants fail to consult memory when search can be accomplished on the basis of visual inspection of the scene (Oliva et al., 2004; Wolfe, Klempen, & Dahlen, 2000). In addition, the effect of the preview tended to be larger as preview duration increased, suggesting that during the preview, participants accumulated visual information relevant to the subsequent search. With a 10 s preview of the scene, participants came to fixate the target object only 390 ms after the onset of the search scene, which was quite remarkably efficient considering that when viewing the preview scene, they did not know which of the objects would be the target. Even a very brief preview of 500 ms was sufficient to facilitate subsequent search through the scene, consistent with the finding of Castelhano and Henderson (2007).

EXPERIMENT 2

In Experiment 2, the preview duration was set at 10 s, and a target-absent preview condition was added to isolate the effects of (1) memory for general scene context and (2) memory for the binding of target object to scene location, both of which could have facilitated search in Experiments 1a and 1b.

Method

Participants. Twenty-four participants completed the experiment.

Stimuli and apparatus. The stimuli and apparatus were the same as in Experiments 1a and 1b, except for the target-absent preview scenes. These scenes were identical to the target-present preview scenes, except the target object was removed before rendering the image. The software automatically filled in the scene information that had been occluded by the target object.

Procedure. The events in a trial were the same as in Experiment 1a. When viewing a scene preview, participants did not know whether the target object was present in the scene or not present in the scene. Participants were informed that some preview scenes would contain the target object and some would not. Participants completed a practice block of six trials, evenly divided among the main conditions. Participants then completed 48 experiment trials, four in each of the 12 conditions created by the 3 (preview type) × 2 (target same/different orientation) × 2 (initial target orientation) design. The entire session required approximately 45 min.

Results

Data from Experiment 2 are reported in Table 3. Neither the initial orientation of the target object nor the match between target probe and search target produced a reliable effect, and the analyses collapsed these factors.

Eye movement analysis. On 2.9% of the trials, participant responded without fixating the target, and these were eliminated from the eyetracking analysis. Trials were also excluded from eyetracking analysis if the participant provided an incorrect manual response. Altogether, 6.2% of the trials was eliminated.

Elapsed time to target fixation. There was a reliable effect of scene preview on elapsed time to target fixation, $F(2, 46) = 12.6$, $p < .001$, with shorter elapsed time in the target-present preview condition than in the target-absent preview condition, $F(1, 23) = 5.62$, $p = .03$, and shorter elapsed time in the target-absent preview condition than in the no preview condition, $F(1, 23) = 6.20$, $p = .02$.

Path ratio. There was a reliable effect of scene preview on path ratio, $F(2, 46) = 8.28$, $p < .001$. There was not a significant difference between path ratios in the target-present preview and target-absent preview conditions, $F(1, 23) = 1.96$, $p = .18$, though the numerical trend was towards a shorter path in the target-present preview condition. Path was shorter in the target-absent preview condition than in the no preview condition, $F(1, 23) = 5.4$, $p = .02$.

Manual RT analysis. Consistent with the eyetracking analysis, there was a reliable effect of preview on correct RT, $F(2, 46) = 12.6$, $p < .001$, with shorter RT in the target-present preview condition than in the target-absent preview condition, $F(1, 23) = 7.20$, $p = .01$, and shorter RT in the target-absent preview condition than in the no preview condition, $F(1, 23) = 5.75$, $p = .03$.

TABLE 3
Experiment 2 search data

	Preview condition		
	Target-present preview	*Target-absent preview*	*No preview*
Elapsed time to target fixation (ms)	404	492	576
Path ratio	1.29	1.41	1.62
Manual reaction time (ms)	1207	1425	1546
Manual response accuracy (%)	94.4	95.7	92.4

Accuracy analysis. Accuracy on the search task did not vary as a function of preview condition, $F(2, 46) = 1.59$, $p = .22$.

Discussion

Search was more efficient in the target-present preview condition than in the target-absent preview condition, which in turn was more efficient than search in the no preview condition, both for correct RT and for elapsed time to target fixation. Memory for general scene context and memory for the specific binding of target object to scene location guided search.

One of the issues that arises when examining the role of scene memory in search is the extent to which the learning is implicit or explicit. In experiments examining contextual cueing, the contextual learning that facilitates search is typically implicit, at least for arrays of simple stimuli (Chun & Jiang, 1998). For real-world scenes, repeated search generates memory for target location that is explicitly available (Brockmole & Henderson, 2006b). To test whether the learning in the present experiments was explicit, I conducted an additional experiment ($N = 12$). The method was identical to that in Experiment 2, except instead of searching for the target object, participants had to estimate the position of the target in the scene. Specifically, participants saw a preview scene (or no preview scene in the no preview condition). Next, the target probe was presented at screen centre. The target probe was removed, and a blank screen was displayed with a mouse cursor. Participants moved the cursor and clicked to mark their estimate of the position of the target object in the scene. This experiment was a replication of Hollingworth (2005a) but using the present set of scene stimuli and a 10 s preview duration (instead of 20 s) so as to allow direct comparison with the present Experiment 2.

In the target-present preview condition, participants were instructed to click on the location in the blank image where the target object had appeared in the preview scene. In the target-absent preview condition, participants were instructed to place the mouse cursor at the position where the target object would have appeared in the preview scene had it been present. Finally, in the no preview condition, participants were instructed to imagine a scene that could contain the target object and choose a position where that object might likely appear. Even with no preview, participants could use knowledge about objects and scenes in general (e.g., that paintings tend to appear higher in scenes than do shoes) to guide placement.

On each trial, position error was calculated as the Euclidean distance (in degrees of visual angle) between the participant's position estimate and the centre of the target object as it appeared in the actual scene. Mean error was reliably lower in the target-present preview condition (3.79°) than in the target-absent preview condition (4.82°), $F(1, 11) = 11.7$, $p = .006$, and error was reliably lower in the target-absent preview condition than in the no preview condition (7.81°), $F(1, 11) = 72.1$, $p < .001$. These data replicate the results of Hollingworth (2005a). Participants had explicit access to both forms of memory that were functional in guiding search in Experiments 1 and 2. Memory for general scene context allowed them to select a plausible target location that was quite close to the actual location of the target when it appeared in the scene, and memory for the specific binding of target object to location enabled even more accurate estimates. Although the present results do not rule out effects of implicit learning on search in Experiments 1 and 2, the most parsimonious account is that explicit learning of scene context and target position was functional in guiding attention and the eyes to the target location.

GENERAL DISCUSSION

In the present experiments, the role of visual memory in search through natural scenes was examined using a preview-search task. Before searching for a particular object in a scene, participants were or were not provided a preview of the scene. The visual memory representation derived from the scene preview significantly facilitated later search through that scene, even though the task could have been performed by a purely visual search process that did not consult memory. Effects of a preview were observed on the very first search through the scene. Therefore, repeated search and target detection (as in contextual cueing experiments) is not necessary to facilitate search in natural scenes. Further, the effect of a preview increased with longer preview durations, consistent with evidence that participants accumulate visual scene information over the course of viewing a scene (Hollingworth, 2004; Hollingworth & Henderson, 2002; Melcher, 2006; Tatler et al., 2003, 2005). Finally, effects of scene memory were observed both on manual search RT and eye movement measures of search efficiency. The latter data add to a large body of literature demonstrating that eye movements when viewing natural scenes are strongly driven by—and often dominated by—cognitive factors, such as memory and top-down goals (Brockmole & Henderson, 2006a, 2006b; Castelhano & Henderson, 2007; Henderson, Brockmole, Castelhano, & Mack, 2007; Henderson et al., 1999; Neider & Zelinsky, 2006; Summerfield et al., 2006; Torralba et al., 2006; Yarbus, 1967).

The main goal of the study was to tease apart two means by which scene memory could facilitate visual search. First, memory for the general context and layout of a scene could guide attention and the eyes to scene regions where the target object was likely to be found. Second, memory for the specific binding of target object to scene location could guide attention and the eyes to the remembered location of that object. To assess the former type of memory, the target-absent preview condition used a scene preview that did not contain the ultimate target object. This preview allowed participants to encode information about the scene context but not information about the specific location of the target. In the target-present preview condition, the target object was present in the preview scene, allowing the encoding of both scene context and target location. Search was more efficient in the target-present preview condition than in the target-absent preview condition, which was in turn more efficient than search in the no preview control condition. Thus, both forms of memory facilitated search.

In the present task, participants could not have known which of the objects in the preview scene would be the target. Thus, the task demanded that they remember the binding of multiple individual objects to scene locations from the preview. Evidence of more efficient search in the target-present preview condition than in the target-absent preview condition therefore demonstrates that the scene memory representation functional in guiding search is not limited to spatial layout, but also preserves information about the form and/or identity of the objects occupying those locations. In addition, such object-to-location binding is likely to be specific to the particular scene in which the object appeared (Hollingworth, 2007).

Finally, both types of scene memory were explicitly available for participant report. Accuracy in a position-estimation task mirrored the results observed in visual search. This finding is consistent with contextual cueing studies using real-world scene stimuli, in which memory for target location is also explicitly available (Brockmole et al., 2006; Brockmole & Henderson, 2006b). Explicit learning of the structure of natural scenes contrasts with implicit learning of spatial structure in contextual cueing studies using random arrays of simple stimuli (Chun & Jiang, 1998). The present results do not rule out a contribution from implicit learning, but they do raise the possibility that traditional search arrays fail to instantiate important features of the behavioural task under investigation (real-world search for natural objects). More generally, as the visual search literature evolves from using search as a methodological tool (e.g., for understanding binding in early vision; Treisman & Gelade, 1980) to investigating search in its own right as an important real-world behaviour, research stimuli and tasks must replicate core properties of that real-world behaviour, including the use of scene stimuli that instantiate the spatial structure and semantic coherence of natural environments.

REFERENCES

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*(4), 699–706.

Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *59*(7), 1177–1187.

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

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017–1040.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The long and the short of it: Relational memory impairments in amnesia, even at short lags. *Journal of Neuroscience*, *26*(32), 8352–8359.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). UK: Elsevier: Oxford.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 519–537.

Hollingworth, A. (2005a). Memory for object position in natural scenes. *Visual Cognition*, *12*, 1003–1016.

Hollingworth, A. (2005b). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*(3), 396–411.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47.

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., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*, 163–181.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559–3565.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6*(1), 8–17.

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

Oliva, A., Wolfe, J. M., & Arsenio, H. C. (2004). Panoramic search: The interaction of memory and vision in search through a familiar scene. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(6), 1132–1146.

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

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, *49*(6), 905–916.

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: Human Experimental Psychology*, *58A*(5), 931–960.

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, *32*(5), 579–592.

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 in object search. *Psychological Review*, *113*, 766–786.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 693–716.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.