Journal of Experimental Psychology:
Human Perception and Performance
2007, Vol. 33, No. 4, 753–763

0096-1523/07/$12.00 DOI: 10.1037/0096-1523.33.4.753

Initial Scene Representations Facilitate Eye Movement Guidance in Visual Search

Monica S. Castelhano
University of Massachusetts at Amherst

John M. Henderson
University of Edinburgh

What role does the initial glimpse of a scene play in subsequent eye movement guidance? In 4 experiments, a brief scene preview was followed by object search through the scene via a small moving window that was tied to fixation position. Experiment 1 demonstrated that the scene preview resulted in more efficient eye movements compared with a control preview. Experiments 2 and 3 showed that this scene preview benefit was not due to the conceptual category of the scene or identification of the target object in the preview. Experiment 4 demonstrated that the scene preview benefit was unaffected by changing the size of the scene from preview to search. Taken together, the results suggest that an abstract (size invariant) visual representation is generated in an initial scene glimpse and that this representation can be retained in memory and used to guide subsequent eye movements.

Keywords: scene perception, eye movements, gaze control, visual search

A good deal of information can be acquired from an initial brief glimpse of a real-world scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Oliva & Schyns, 1997; Potter, 1976; Potter & Levy, 1969; Schyns & Oliva, 1994; Thorpe, Fize, & Marlot, 1996; for reviews, see Henderson & Ferreira, 2004; Henderson & Hollingworth, 1999). At the same time, scene viewing typically involves active visual sampling, with observers moving their eyes approximately three times each second to fixate important objects and scene elements (for reviews, see Findlay & Gilchrist, 2003; Henderson, 2003; Henderson & Hollingworth, 1998; Rayner, 1998). This raises a question: What is the relationship between the information acquired within an initial glimpse and subsequent active scene exploration via eye movements?

In the present study, we contrasted two potential hypotheses concerning this relationship. First, the initial glimpse could interact with later active exploration by providing information useful for guiding eye movements to informative scene regions. There is some evidence that in complex scene viewing, an initial set of orienting fixations is followed by more task-specific eye movements (Land & Hayhoe, 2001; see also Antes, 1974; Loftus & Mackworth, 1978). According to this *orienting hypothesis*, the initial glimpse provides information that plays a similar orienting role. This hypothesis requires that information acquired during the initial glimpse be stored in a memory system that is not retinotopically organized (so that it survives changes in the image falling on the retina across saccades) and that continues to be available as scene processing unfolds over multiple eye fixations. According to this hypothesis, the relationship between the developing scene representation and eye movement guidance is interactive, with the initially generated scene representation interacting with task knowledge to guide the eyes to informative locations. Information acquired during subsequent fixations might then be added to the developing representation, which in turn could be used to guide future eye movements and so on (Henderson & Castelhano, 2005; Hollingworth, 2005).

A second possibility is that whereas an initial glimpse might be sufficient to allow rapid identification of objects and scene categories (Intraub, 1981; Potter, 1976; Schyns & Oliva, 1994; see also Li, VanRullen, Koch, & Perona, 2002; Thorpe, Fize, & Marlot, 1996), it may not be sufficiently detailed or long lived to support the guidance of subsequent eye movements. For example, an initial glimpse could provide general information about a scene's semantic gist (e.g., its basic level category, general meaning, expectations about semantic associations; Biederman et al., 1982; Friedman, 1979; Mandler & Johnson, 1976), but this generic representation might not be particularly useful for guiding eye movements through a particular instance of a scene. It is possible that scene category information alone may not be useful for guiding subsequent eye movements without more detailed information. Although there is now relatively good evidence that eye movements during visual search can draw on memory (Gibson, Li, Skow, Brown, & Cooke, 2000; Gilchrist & Harvey, 2000; Kristjánsson, 2000; Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Shore & Klein, 2000), this point has not been shown for eye movements in real-world scenes. Finally, to be useful across saccades, information acquired in an initial glimpse would have to be coded in

This research was supported by a Michigan State University graduate school fellowship (National Science Foundation Integrative Graduate Education and Research Traineeship Program Grant ECS-9874541) to Monica S. Castelhano. This work was supported by National Science Foundation Grant BCS-0094433 and Army Research Office Grant W911NF-04-1-0078. We thank Bryan Corpus and Devon Witherell for data collection help; Aaron Pearson, James Brockmole, and Mareike Wieth for helpful discussions and comments on drafts of this article; and Peter De Graef, Thomas Sanocki, and Alexander Pollatsek for their thoughtful comments. The opinions expressed in this article are those of the authors and do not necessarily represent the views of the Department of the Army or any other governmental organization.

Correspondence concerning this article should be addressed to Monica S. Castelhano, Department of Psychology, University of Massachusetts, Amherst, MA 01003, or to John M. Henderson, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, United Kingdom. E-mail: monica@eyelab.msu.edu or john.m.henderson@ed.ac.uk

a nonretinotopic coordinate system. If it is not, then it would not continue to be useful once the point of fixation has changed.

One way to determine whether information acquired in an initial scene glimpse can influence subsequent eye movement control is to manipulate the availability of the initial glimpse while also allowing participants to make eye movements through the scene. A difficulty arises in trying to meet both of these constraints, however, because if the complete scene remains visible during subsequent fixations, then it is not possible to determine whether eye movement control decisions are based on information acquired during the initial glimpse or on information acquired in subsequent fixations. To provide for both manipulation of a scene glimpse and extended scene viewing involving saccadic eye movements, while at the same time restricting scene information available following the initial fixation, we developed the new flash-preview moving-window paradigm. This new paradigm combines the brief tachistoscopic viewing method typically used in scene identification experiments with the moving-window technique typically used to investigate eye movements under restricted viewing conditions. In the paradigm, participants were asked to search for target objects in scenes while their eye movements were recorded. Prior to presentation of the search scene, a scene preview was briefly presented, and the nature of that preview was manipulated. The search then took place with the search scene visible only through an eye-contingent moving window with a 2° diameter centered at fixation within a scene that measured about 15° × 12° of visual angle (van Diepen, Wampers, & d'Ydewalle, 1998; see also Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981).

The specific question we asked in this study was whether eye movements during search would be more efficient given a brief preview of a search scene compared with a control preview. If the initially generated scene representation can be used to plan subsequent eye movements, then a *scene preview benefit* should be observed such that an informative preview of the search scene produces greater eye movement efficiency than an uninformative preview. If, on the other hand, eye movements during scene search are controlled independently of information acquired in an initial glimpse, then the nature of an initial scene preview should not matter for subsequent eye movements and so no scene preview benefit should be observed.

Experiment 1

The purpose of Experiment 1 was to establish whether the information acquired from a scene in an initial glimpse can facilitate subsequent eye movements. Experiment 1 contrasted identical, different, and control scene preview conditions. In the identical condition, the preview scene was the same as the search scene. In the different condition, a scene unrelated to the search scene was presented in the preview. In the control condition, a meaningless mask was presented as the preview. Both the different and control conditions were included because past studies have suggested that a meaningful scene may facilitate the processing of another scene compared with a meaningless prime, regardless of the similarity of those scenes (Sanocki, 2003; Sanocki & Epstein, 1997). We then expected some scene preview benefit from the different condition compared with a meaningless control condition simply because the different condition involved presentation of a scene. Alternatively, it could be that presentation of a different scene would lead to interference compared with the meaningless control condition. Inclusion of both control conditions allowed us to determine the degree to which the different condition could be used as a baseline.

During the search, participants were able to gather information about the scene only through a 2° window centered at fixation that was synchronized with movements of the eyes. If the representation generated from an initial scene glimpse can play a functional role in eye movement control, then a scene preview benefit should be observed such that eye movements more efficiently locate the search target in the identical condition compared with the different and control conditions. If, however, the scene representation generated from an initial fixation plays no role in subsequent eye movement planning, then no scene preview benefit should be found.

Method

Participants. Twelve Michigan State University undergraduates participated for course credit or for $7/hr.

Apparatus. Eye movements were recorded at 1000 Hz using a Generation 5.5 Dual-Purkinje-Image Eyetracker with a resolution of 1 min of arc and linear output over the range of the visual display. The position of the right eye was tracked, though viewing was binocular. The eyetracker and display monitor were interfaced with a computer that controlled the experiment.

Stimuli. The stimuli were digitized photographs of real-world scenes. The scenes were displayed at 800 × 600 pixels and true-color resolution on a cathode ray tube monitor, with a refresh rate of 143 Hz. At the viewing distance of 1.13 m, the scenes subtended visual angles of 15.20° (horizontal) × 11.93° (vertical). Thirty-six search scenes were shown, and the search scenes always contained the search target. The targets were selected to be at ~2° from the center of the scene and had an average size of 1.83° × 1.9° (see Table 1). Details of the targets (i.e., name, size, and eccentricity) can be found in Appendix A. In the identical preview condition, the preview and search scene was the same photograph. In the different preview condition, the preview scene differed in identity and spatial layout from the search scene. The control preview condition was created from scrambled sections taken from all of the scenes. The scrambled sections measured 10 × 10 pixels; the control preview was meaningless but with some local visual characteristics such as color, lighting, and contours. Each participant saw each search scene once, and search scenes were rotated through conditions across participants via a Latin square.

Table 1
Mean Size and Mean Eccentricity From the Center of the Screen for Target Objects in Each Experiment

	Experiment 1		Experiments 2, 3, and 4	
Variable (in degrees)	Horizontal	Vertical	Horizontal	Vertical
Size				
M	1.83	1.90	2.20	2.19
SD	0.70	0.80	0.81	0.88
Range	0.93–3.95	0.76–3.91	0.91–4.41	0.66–4.66
Eccentricity				
M	4.95		4.24	
SD	1.26		1.50	
Range	1.83–6.91		1.63–6.91	

Procedure. At the beginning of each session, the eyetracker was calibrated. Participants' viewing position was maintained with a bite bar and forehead rest. Calibration was deemed accurate when estimated fixation position was within ± 10 min of arc from each of five test points. Figure 1 depicts the trial events. Participants began each trial by fixating a center point. The preview image was then presented for 250 ms, followed by a visual mask for 50 ms. Following the mask, a word naming the target object was presented at the center of the display for 2 s. The total delay between preview scene offset and search scene onset was therefore 2,050 ms. The target was named after the scene preview to discourage participants from trying to identify target features while ignoring global scene information (Evans & Treisman, 2005). The search scene was then shown through a 2° diameter circular window centered at fixation that moved as the eyes moved (for further details on the implementation of the moving window, see Henderson, McClure, Pierce, & Schrock, 1997). The display screen outside of the moving window was gray. Eyetracking was checked before each trial and recalibrated if necessary to ensure accurate positioning of the moving window. Participants were instructed to search the scenes for the named targets and to press the response button when they had found and were looking at the target. The search scene was displayed for 15 s or until the participant pressed the response button. Prior to the experimental trials, six practice trials were given in which the preview and search scenes were identical. Participants were told that during the experimental trials, the preview would not always match the search scene. The entire experiment lasted approximately 40 min.

Results

Accuracy. A search was scored as correct when the participant both fixated on the target object and pressed the response button. This method of scoring the searches ensured that the participants were responding to the correct target object and reassured us that no participant was just randomly pressing the button after a given interval. Search accuracy averaged 43% (identical: 47%, different: 44%, control: 38%) and did not significantly differ across conditions, $F(2, 22) = 1.205$, $MSE = .025$. The response time (RT) and eye movement analyses included only correct trials to ensure that they reflected successful search. Table 2 summarizes the RT and eye movements measures across the conditions.

RT. RT was defined as elapsed time from search scene onset until the button press. None of the contrasts were statistically significant: identical versus different, $t(11) = -1.61$; identical versus control, $t(11) = -1.3$; different versus control, $t(11) = 0.803$. However, the numerical pattern was the same as the other measures reported below. Numerically, a scene preview benefit was observed in the identical condition compared with the different and control conditions.

To examine eye movement behavior as a function of the information available in an initial fixation, we calculated three dependent measures. These eye movement measures more directly reflected the efficiency of eye movement guidance during the search. *Latency to first target fixation* was defined as the elapsed time between search scene onset and the beginning of the first fixation on the target object. *Number of fixations to first target fixation* was defined as the number of discrete fixations on the scene from scene onset until but not including the first fixation on the target object.

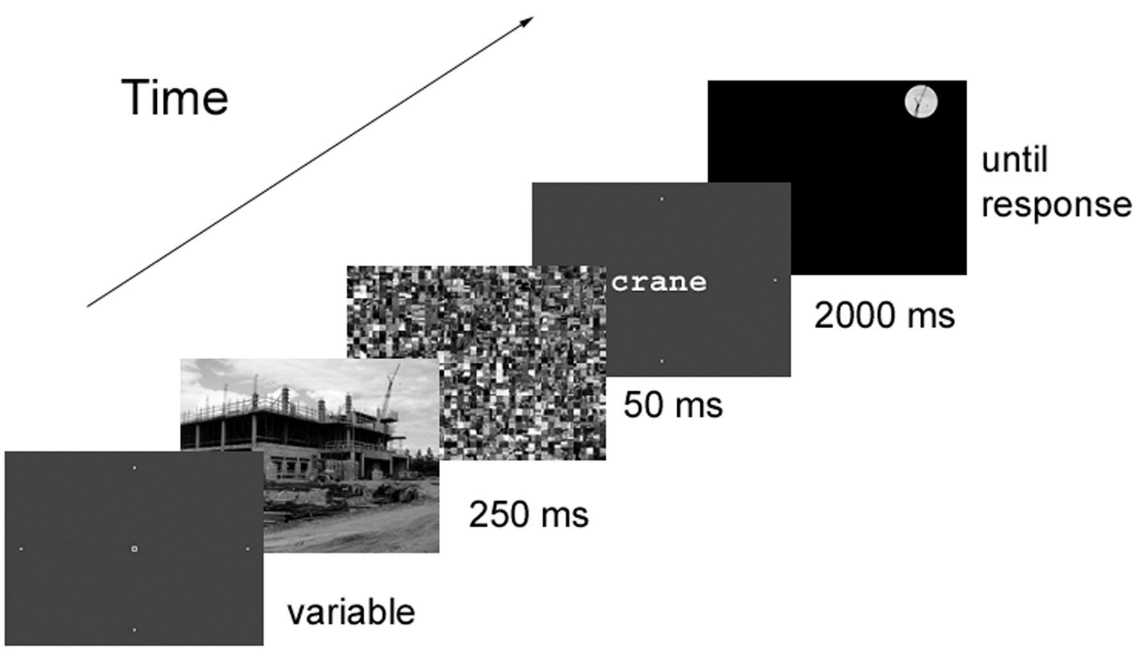

Figure 1. The trial sequence for all experiments. Participants fixated a center point on a calibration screen, then viewed the preview scene for 250 ms, followed by a visual mask for 50 ms, followed by a word naming the target object for 2,000 ms. The search scene was then shown through a 2° diameter circular window centered at fixation until response or for a maximum of 15 s.

Table 2
Mean Reaction Times (RTs) and Eye Movement Measures of Search With Moving Window for the Preview Condition in Experiment 1

Variable	Same	Different	Control
RT (ms)			
M	6,783	7,980	7,330
SE	577	718	510
Latency (ms)			
M	4,604	6,851	5,873
SE	503	650	424
Number of fixations			
M	11.30	16.40	14.50
SE	0.93	1.23	1.01
Pattern ratio (path taken: straight path)			
M	3.30	5.20	4.90
SE	0.24	0.62	0.40

Scan pattern ratio was defined as the length of the scan pattern taken by the participant's eyes through the scene on his or her way to the target object (computed as the summed distance between all fixations from scene onset to the first fixation on the target), divided by the most direct possible path (computed as the distance from the central fixation point to the center of the target object). The target object was defined as the smallest possible rectangle to encompass the target object. The boundary allowed us to be certain that the participants were fixating on the object itself and that the object was visible within the present window (further details about object size and eccentricity are shown in Table 1 and Appendix A).

For these measures, we conducted three planned comparisons: identical versus different, identical versus control, and different versus control. The first two comparisons tested whether the initial view of the search scene resulted in a scene preview benefit for the identical condition. The third comparison provided information about whether a benefit would be observed for any scene preview (Sanocki, 2003) and conversely also allowed us to determine whether an inappropriate scene preview might produce interference compared with the meaningless control condition. Statistical significance for these planned comparisons was set at $\alpha = .05$.

Latency to first target fixation. Latency in the identical condition was significantly shorter than in both the different and control conditions, $t(11) = -2.96$, $p < .05$, and $t(11) = -2.22$, $p < .05$, respectively; the different and control conditions did not differ, $t(11) = 1.22$.

Number of fixations to first target fixation. Participants made fewer fixations in the identical condition than in the different and control conditions, $t(11) = -3.06$, $p < .05$, and $t(11) = -2.91$, $p < .05$, respectively; the different and control conditions did not differ, $t(11) = 1.3$.

Scan pattern ratio. The identical condition produced a shorter scan pattern ratio than both the different and control conditions, $t(11) = -2.9$, $p < .05$, and $t(11) = -3.3$, $p < .05$, respectively; the different and control conditions did not differ, $t(11) = 0.34$.

Discussion

Across all eye movement measures, a scene preview benefit was observed such that participants searched scenes more efficiently in a moving-window display when a 250-ms preview was identical to the search scene than when it was a different scene or was meaningless. RT measure showed the same pattern numerically and may have had a smaller effect due to the variability in how long participants took to verify the fixated object as the target. Thus, Experiment 1 established that participants were able to generate, retain, and use an initial scene representation to facilitate eye movements during a subsequent search that began about 2 s later. These results provide support for the orienting hypothesis, and we conclude that the initial scene representation does in fact have an effect on the guidance of eye movements. In the next three experiments, we sought to further understand the nature of the initial scene representation.

Experiment 2

Experiment 2 explored the contribution of scene identity and basic-level category membership to the scene preview benefit observed in Experiment 1. In Experiment 2, a concept preview condition was introduced in which the category of the search scene was preserved but the visual details differed (see Figure 2B). The concept preview was compared with the identical and different preview conditions used in Experiment 1. If the scene concept was responsible for the scene preview benefit observed in Experiment 1, then there should be no difference in scene preview benefit between the identical and concept conditions, and both of these conditions should produce a scene preview benefit compared with the different condition. If instead scene concept is not responsible for the scene preview benefit, then the identical condition should produce a preview benefit compared with both the concept and different conditions, which should not differ from each other. Finally, if scene concept contributes partially to the scene preview benefit, then the preview benefit in the concept condition should fall between the identical and different preview conditions, with a smaller scene preview benefit than the identical condition.

Method

The method used in Experiment 2 was the same as in Experiment 1 with the following exceptions.

Participants. Twelve Michigan State University undergraduates participated for course credit or for $7/hr. These participants did not take part in the previous experiment.

Stimuli. Search scenes remained constant from Experiment 1 to Experiment 2, but 17 of the 36 named search targets were changed in Experiment 2 to increase search accuracy and therefore the number of trials from which eye movement data could be analyzed. Some new targets were selected to be more easily identifiable by participants. Other new targets were chosen to be larger or slightly closer to the center of the scene. The newly selected targets had an average size of $2.25° \times 2.2°$. The changes made to the targets are summarized in Table 1, and specific details including size and eccentricity can be found in Appendix B. The meaningless control preview condition was replaced by the concept preview condition. For the concept preview condition, 36 preview scenes were added. Each concept preview scene came from the same conceptual category as its paired search scene but differed in its visual details. As in Experiment 1, each participant saw each search scene once, and search scenes were rotated through the three scene preview conditions across participants via a Latin square.

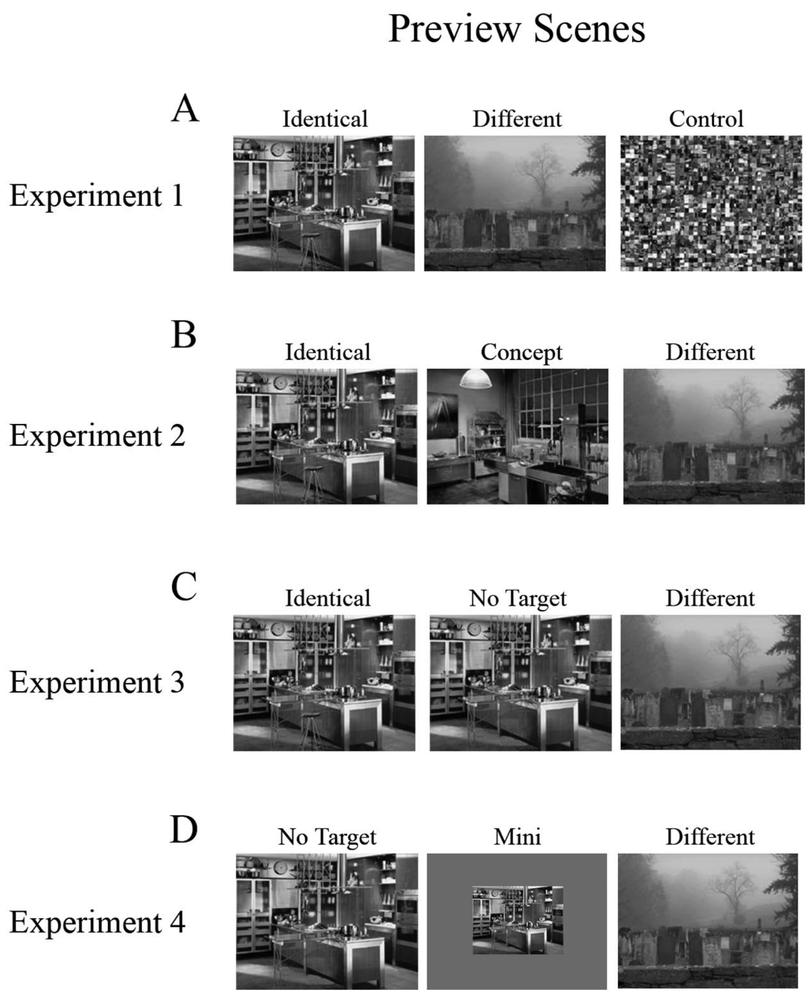

Figure 2. Example preview scenes for each experiment. A: In Experiment 1, the preview scene could be identical to the search scene, a different scene, or a meaningless mask. B: In Experiment 2, the preview scene was identical to the search scene, a different scene, or a scene with the same concept but different details. C: In Experiment 3, the preview scene was identical to the search scene, a different scene, or the same as the search scene with the target object removed. D: In Experiment 4, the preview scene was identical to the search scene without the target, a different scene, or the same as the no-target scene reduced in size (50% of width and height).

Results

Accuracy. Accuracy was scored using the same method described in Experiment 1 and averaged 69% (identical: 69%, concept: 70%, control: 67%; $F < 1$). RT and eye movement analyses were based on correct trials. Table 3 summarizes the RT and eye movement measures across the conditions.

RT. The identical preview condition was significantly faster than both the concept and different conditions, $t(11) = -3.89$, $p < .05$, and $t(11) = -3.93$, $p < .05$, respectively. The concept and different conditions did not significantly differ from each other, $t(11) = -0.37$.

Latency to first target fixation. Latency to first target fixation in the identical condition was significantly faster than both the concept and different conditions, $t(11) = -3.3$, $p < .05$, and $t(11) = -4.72$, $p < .05$, respectively; the concept and different conditions did not differ, $t(11) = 0.29$.

Table 3
Mean Reaction Times (RTs) and Eye Movement Measures of Search With Moving Window for the Preview Condition in Experiment 2

Variable	Same	Concept	Different
RT (ms)			
M	3,976	5,443	5,582
SE	336	293	379
Latency (ms)			
M	2,687	4,151	4,045
SE	329	300	301
Number of fixations			
M	7.40	11.40	10.70
SE	0.82	0.87	0.71
Pattern ratio (path taken: straight path)			
M	3.40	4.20	4.60
SE	0.43	0.30	0.30

Number of fixations to first target fixation. Participants made fewer fixations before fixating the target object in the identical preview condition than in both the concept and different conditions, $t(11) = -2.82$, $p < .05$, and $t(11) = -4.4$, $p < .05$, respectively; the concept and different preview conditions did not differ, $t(11) = 0.73$.

Scan pattern ratio. Although the pattern was similar to that produced by the other measures, the differences were not statistically significant. The identical preview condition produced a marginally smaller ratio than the different condition, $t(11) = -2.14$, $p = .056$, but did not differ from the concept condition, $t(11) = -1.46$. The concept and different conditions did not differ, $t(11) = 0.79$.

Discussion

A clear scene preview benefit was observed when the scene presented in an initial fixation was identical to the search scene, replicating Experiment 1. However, there was no evidence in any measure that scene concept or identity accounted for this scene preview benefit. It seems that although scene category is acquired within the first fixation (Biederman et al., 1982; Potter, 1976), this information by itself is not sufficient to facilitate eye movement guidance during search.

Experiment 3

In Experiments 1 and 2, the identity of the search target was specified after the preview scene, so participants could not visually search the brief preview display for the target. However, it is conceivable that participants might have identified a small number of objects during the preview, enabling them to find the target faster once the target label was presented when it happened to match one of those identified objects. In Experiment 3, we directly tested whether the scene preview benefit was the result of identifying target objects in the identical preview condition. A no-target preview condition was introduced in which the preview scene was identical to the search scene except that the target object had been deleted from it (see Figure 2C). As in Experiments 1 and 2, the target was always present in the search scene. If the scene preview benefit observed in the first two experiments was due to target object identification in the identical condition preview, then the identical condition should show a scene preview benefit, but the no-target condition should not. If, however, the scene preview benefit was at least partly due to scene information other than identification of the specific target, then the no-target preview should also produce a scene preview benefit over the different condition.

Method

The method used in Experiment 3 was the same as in the previous experiments with the following exceptions.

Participants. Twelve Michigan State University undergraduates participated for course credit or for $7/hr. These participants did not participate in any of the previous experiments.

Stimuli. The concept preview condition in Experiment 2 was replaced with a no-target preview condition. In the no-target condition, the identical preview scenes were modified by digitally deleting the target objects. Again, each participant saw each search scene once, and search scenes were rotated through the three scene preview conditions across participants via a Latin square.

Results

Accuracy. Accuracy averaged 68% (identical: 69%, no target: 69%, different: 66%; $F < 1$). RT and eye movement analyses included only correct trials. Table 4 summarizes the RT and eye movements measures across the conditions.

RT. The identical condition was significantly faster than the different condition, $t(11) = -2.69$, $p < .05$. There was no significant difference between the no-target condition and the other conditions (vs. identical: $t[11] = -0.787$; vs. different, $t[11] = -1.72$).

Latency to first target fixation. Latency to first target fixation in the identical and no-target conditions was significantly faster than in the different condition, $t(11) = -3.29$, $p < .05$, and $t(11) = -2.29$, $p < .05$, respectively; the identical and no-target conditions did not differ, $t(11) = -0.98$.

Number of fixations to first target fixation. Fewer fixations were made before fixating the target object in the identical and no-target conditions than in the different condition, $t(11) = -3.65$, $p < .05$, and $t(11) = -2.89$, $p < .05$, respectively; the identical and no-target conditions did not differ, $t(11) = -0.87$.

Scan pattern ratio. Although the pattern was similar to that of the other eye movement measures, only the identical and different conditions differed significantly: identical versus different condition, $t(11) = -2.33$, $p < .05$; identical versus no-target condition, $t(11) = -0.85$; and no-target versus different condition, $t(11) = -1.46$.

Discussion

The latency to the first fixation on the search target and the number of fixations to that first fixation were both significantly facilitated in the no-target preview condition compared with the different condition. The ratio of the search pattern and overall response latencies also showed a similar tendency, though these differences were not significant. Overall, these results suggest that the scene preview benefit is not entirely a result of identifying the target in the preview. At the same time, the pattern of means seen

Table 4
Mean Reaction Times (RTs) and Eye Movement Measures of Search With Moving Window for the Preview Condition in Experiment 3

Variable	Same	No target	Different
RT (ms)			
M	4,657	5,093	6,032
SE	296	409	447
Latency (ms)			
M	2,990	3,459	4,595
SE	233	343	412
Number of fixations			
M	8.40	9.40	12.40
SE	0.64	0.92	1.03
Pattern ratio (path taken: straight path)			
M	3.50	3.80	4.60
SE	0.25	0.36	0.36

in all measures in Experiment 3 showed a greater scene preview benefit for the identical over the no-target previews, suggesting that target object identification in the preview image may also have played at least some role in the preview benefit. A subsidiary analysis was conducted to examine the role of target object identification more fully. However, these results show that even when highly salient target objects were removed from the analysis, the pattern of results was the same. Removing target objects from the preview scenes did not eliminate the scene preview benefit. This finding rules out the possibility that the scene preview benefits observed in the three experiments solely were due to identifying targets in the preview scenes. Therefore, we conclude that the results were not due to the identification of the location of highly salient target objects.

Experiment 4

Experiment 2 established that the general concept of the scene does not play a role in the scene preview benefit. Experiment 3 ruled out the possibility that the effect is solely due to identification of the target object. What, then, is the nature of the information that is driving the preview benefit? Prior research on short-term visual memory suggests that abstracted (i.e., noniconic) visual representations can persist over time and across eye movements (Carlson-Radvansky, 1999; Irwin, 1993; McConkie & Zola, 1979; Pollatsek, Rayner, & Collins, 1984; Rayner, McConkie, & Zola, 1980; Verfaillie & De Graef, 2000). These abstract visual representations are thought to preserve visual details in a nonmetric, nonmaskable, and nonretinotopic visual store (Hollingworth & Henderson, 2002). On the one hand, if such abstracted visual representations are the source of the scene preview benefit demonstrated here, then this benefit should persist across a metric transformation of the scene from preview to search. On the other hand, if the representation is more visually precise or metrically based, then a metric transformation should disrupt the preview benefit. These competing predictions were tested in Experiment 4 by adding a preview condition in which a smaller mini version of the search scene was presented as the preview in which the search scene was reduced in size by 50% in width and height (75% in area). As in Experiment 3, no-target and different preview conditions were also included in the design.

If the initial information giving rise to the scene preview benefit is image specific, we would expect the no-target preview condition to result in a scene preview benefit compared with both the mini and different conditions. However, if the information producing the preview benefit is abstracted from precise metric visual information, then the no-target and mini previews should produce scene preview benefits compared with the different condition because the same abstract scene-based information would be available in both previews.

Method

The method used in Experiment 4 was the same as in the previous experiments with the following exceptions.

Participants. Eighteen Michigan State University undergraduates participated for course credit or for $7/hr. These participants did not take part in any of the previous experiments.

Stimuli. No-target, mini, and different preview conditions were contrasted in Experiment 4. The mini preview condition was created by reducing the no-target previews to 25% of their original size. The mini previews measured 400 × 300 pixels and subtended 7.91° horizontally × 5.94° vertically. Each mini preview was presented on a gray background that created a 800 × 600-pixel image (see Figure 2D). Again, each participant saw each search scene once, and search scenes were rotated through the three scene preview conditions across participants via a Latin square.

Results

Accuracy. Accuracy averaged 68% (no target: 72%, mini: 64%, different: 69%; $F < 1$). RT and eye movement analyses included only correct trials. Table 5 summarizes the RT and eye movements measures for Experiment 4.

RT. The identical and mini conditions were significantly faster than the different condition, $t(17) = -2.61$, $p < .05$, and $t(17) = -2.68$, $p < .05$, respectively. There was no significant difference between the no-target and mini conditions, $t(11) = -0.194$.

Latency to first target fixation. Latency to first target fixation in the no-target and mini conditions was significantly faster than in the different condition, $t(17) = -2.976$, $p < .05$, and $t(17) = -2.87$, $p < .05$, respectively; the no-target and mini conditions did not differ, $t(16) = -0.16$.

Number of fixations to first target fixation. Fewer fixations were made before fixating the target object in the no-target and mini conditions than in the different condition, $t(17) = -2.83$, $p < .05$, and $t(17) = -2.78$, $p < .05$, respectively; the no-target and mini conditions did not differ, $t(17) = -0.31$.

Scan pattern ratio. Numerically, the pattern was similar to that of the other eye movement measures, in that the no-target and mini conditions showed a slightly smaller ratio; however, only the mini and different conditions differed significantly: no-target versus different condition, $t(17) = 1.55$; no-target versus mini condition, $t(17) = -0.96$; and mini versus different condition, $t(17) = -2.21$, $p < .05$.

Discussion

The results from Experiment 4 reveal that the scene preview benefit is present even when the preview differs markedly in scale.

Table 5
Mean Reaction Times (RTs) and Eye Movement Measures of Search With Moving Window for the Preview Condition in Experiment 4

Variable	No target	Mini	Control
RT (ms)			
M	4,479	4,409	5,326
SE	256	250	303
Latency (ms)			
M	3,545	3,491	4,457
SE	252	228	299
Number of fixations			
M	10.00	9.70	12.40
SE	0.78	0.60	0.92
Pattern ratio (path taken: straight path)			
M	3.90	3.50	4.30
SE	0.30	0.23	0.35

This finding strongly suggests that the representation that is preserved from the initial glimpse and that is available to facilitate eye movements in search is visually abstract. These results are consistent with other findings suggesting that visual short-term memory codes abstracted visual representations that can persist over time and across eye movements (Carlson-Radvansky, 1999; Irwin, 1993; McConkie & Zola, 1979; Pollatsek et al., 1984; Rayner et al., 1980; Verfaillie & De Graef, 2000). In addition, the results of Experiment 4 replicate those of Experiment 3 showing clear scene preview benefits despite the absence of the target objects in the previews. These results again demonstrate that the scene preview benefit is not simply a matter of target object identification from the preview.

General Discussion

Two important facts about scene perception are that the general concept of a scene can be apprehended very rapidly and that scene perception involves sequential eye fixations that are extended over time. The purpose of the present study was to investigate the relationship between initial scene perception and temporally extended scene viewing. Specifically, we investigated the hypothesis that one function of an initially generated scene representation is to facilitate subsequent eye movement control.

To investigate this issue, we introduced a flash-preview moving-window paradigm in which participants searched for a target object in a scene photograph. During search, the view was restricted to a 2° area of the scene visible through a moving window that was centered on the current fixation position. Prior to the moving window search, participants were shown a 250-ms scene preview. The preview was manipulated to determine whether the information acquired in this initial fixation could be used to plan later eye movements.

In Experiment 1, the scene preview was identical to the search scene, a different scene, or a meaningless control image. Results across all eye movement measures showed a scene preview benefit in which eye movements were more efficient in searching for the target following an identical preview. These results suggest that the information acquired from an initial fixation can play a functional role in planning eye movements. Experiment 2 investigated whether the conceptual category or identity of the scene could account for the scene preview benefit. The results reveal that a preview that provided the scene's category–identity but not its visual details did not produce a scene preview benefit. These results suggest that something more specific about the scene was responsible for the scene preview benefit. Experiment 3 investigated whether the preview benefit was due to the presence of the target object in the preview scene. The presence of the search target in the preview scene was manipulated, and a scene preview benefit was observed regardless of whether the preview scene included the target. These results suggest that the initial scene representation contains information additional to the target object that can guide eye movements. Experiment 4 used a manipulation of the preview scene's size to test whether the scene representation used to guide eye movements was image based or visually abstract. The results show that a change to the size of the scene from preview to search produced a scene preview benefit that was very similar to that produced when the size remained constant. Taken together, the results of the four experiments demonstrate that the scene representation generated from a preview available in an initial fixation provides information that can be used to plan subsequent eye movements through a scene.

Recent studies have shown that the visual system can store spatial layout information and use it to improve later task performance (Sanocki, 2003; Sanocki & Epstein, 1997). For example, Sanocki and Epstein (1997) found that participants were better able to judge which of two targets in a scene was closer in depth when the scene was primed. In a more recent study, Sanocki (2003, Experiment 2) found that changes in lighting between the prime and probe scenes produced equivalent layout priming, suggesting that the priming effect was not based on detailed image properties. The results of the present study complement these findings and further suggest that a scene representation formed within an initial glimpse is abstract enough to be useful across spatial scale changes. These data converge with other demonstrations that a rapidly acquired and represented scene structure lingers in memory and can facilitate later perceptual and cognitive operations and behavior (Henderson & Hollingworth, 1999).

A number of past studies and recently proposed computational models posit that eye movements are strongly guided by immediately available visual information (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Parkhurst, Law, & Niebur, 2002; Parkhurst & Niebur, 2003). The present study demonstrates that an initially generated scene representation can linger and that this representation is capable of continuing to guide eye movements. A remaining issue is whether, under normal viewing conditions in which the complete scene is available in each fixation, representations from prior fixations (including the initial fixation) continue to exert an influence on eye movement control. On the one hand, one might expect that representations generated in later fixations would take precedence over initially acquired information. However, there are two ways in which the initially acquired information may continue to play an important or even primary role in natural scene perception. First, the initially generated representation may provide a scaffold or frame onto which subsequent information can be added and combined (Friedman, 1979). That is, an initial glimpse may provide hypotheses (or priors in a Bayesian framework) about what objects are present and where they are likely to be found (Torralba, Oliva, Castelhano, & Henderson, 2006). Second, as observers move their eyes (and bodies), some of the information that was available in prior views is no longer visible. Representations generated from these prior glimpses are likely to continue to be available and allow more efficient eye movements to those areas of a scene that are no longer in view (Oliva, Wolfe, & Arsenio, 2004). The results of the present study can be seen as offering an existence of proof for the availability and functionality of such lingering representations.

The results of the current study also bear on the more general issue of scene representation and memory. Phenomena such as change blindness (Simons, 2000) and inattentional amnesia (Wolfe, 1999) initially led to the hypothesis that scene representations and memory are sketchy and conceptual at best. More recent evidence, in contrast, has shown that when appropriate behavioral probes are used to assess scene memory, evidence for relatively detailed representation is found (Castelhano & Henderson, 2005; Hollingworth, 2005; Hollingworth & Henderson, 2002; see reviews by Henderson & Castelhano, 2005; Hollingworth, 2006). The results of the present study are consistent with these latter findings in the sense that a representation generated from an initial scene glimpse was detailed enough to be useful in guiding subsequent eye movements. Furthermore, the scene representation generated from an initial glimpse was stable enough to linger over an initial 2-s delay (from preview display to onset of search scene) and to continue to be functional over multiple fixation–saccade cycles. In contrast, according to the hypothesis that scene representations are fragile and short lived, there would be no reason to expect that memory for scene information acquired in an initial glimpse would continue to be available and functional for subsequent eye movement control.

Finally, from a methodological perspective, the results of this study demonstrate that the flash-preview moving-window paradigm is useful for investigating how eye movements can be used to reveal the presence and nature of the representation generated from a brief initial glimpse of a scene. The paradigm allows for flexibility in how the preview is presented (e.g., the content and duration of the preview can be manipulated), providing leverage for asking additional questions concerning the nature of the representation that is initially generated from a scene. Furthermore, the flash-preview moving-window paradigm uses eye movements to provide an implicit measure of the representation of behaviorally relevant information present in the initially generated representation.

References

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103,* 62–70.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Carlson-Radvansky, L. A. (1999). Memory for relational information across saccadic eye movements. *Perception & Psychophysics, 61,* 919–934.

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

Evans, K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476–1492.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing.* Oxford, England: Oxford University Press.

Friedman, A. (1979). Framing pictures: The roles of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316–355.

Gibson, B. S., Li, L., Skow, E., Brown, K., & Cooke, L. (2000). Searching for one versus two identical targets: When visual search has memory. *Psychological Science, 11,* 324–327.

Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanism in visual search. *Current Biology, 10,* 1209–1212.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences, 7,* 498–504.

Henderson, J. M., & Castelhano, M. S. (2005). Eye movements and visual memory for scenes. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 213–235). New York: Oxford University Press.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action* (pp. 1–58). New York: Psychology Press.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance while reading and while watching dynamic scenes* (pp. 269–293). Oxford, England: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271.

Henderson, J. M., McClure, K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics, 59,* 323–346.

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

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition, 14,* 781–807.

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

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 604–610.

Irwin, D. E. (1993). Perceiving an integrated visual world. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 121–142). Cambridge, MA: MIT Press.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model for saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Kristjánsson, Á. (2000). In search of remembrance: Evidence for a memory for locations in visual search. *Psychological Science, 11,* 328–332.

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

Li, F.-F., VanRullen, R., Koch, C., & Perona, P. (2002). Natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, USA, 99,* 9596–9601.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565–572.

Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 529–540.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception & Psychophysics, 25,* 221–224.

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

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

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107–123.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16,* 125–154.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science, 12,* 287–292.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113,* 426–442.

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

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., Inhoff, A. W., Morrison, R., Slowiaczek, M. L., & Bertera,

J. H. (1981). Masking of foveal and parafoveal vision during fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 167–179.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology, 12,* 206–226.

Sanocki, T. (2003). Representation and perception of scenic layout. *Cognitive Psychology, 47,* 43–86.

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Scienc*e, *8,* 374–378.

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

Shore, D. I., & Klein, R. M. (2000). On the manifestations of memory in visual search. *Spatial Vision, 14,* 59–75.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1–16.

Thorpe, S., Fize, D., & Marlot, C. (1996, June 6). Speed of processing in the human visual system. *Nature, 381,* 520–522.

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, 113,* 766–786.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337–355). Oxford, England: Elsevier Science.

Verfaillie, K., & De Graef, P. (2000). Transsaccadic memory for position and orientation of saccade source and target. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1243–1259.

Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.

Appendix A

Size and Eccentricity of Target Objects for Experiment 1

Item	Scene	Targets	Height (in degrees)	Width (in degrees)	Eccentricity (in degrees)[a]
1	Hardware store	Air conditioner	0.83	0.93	5.63
2	Fishing boats	Bell	1.53	0.93	6.91
3	Storefront	Bench	1.55	2.22	6.03
4	Backyard garden	Birdhouse	2.90	2.43	2.90
5	Venice street	Boat	1.26	1.52	1.83
6	Living room	Boot	3.88	2.40	4.93
7	Living room	Bowl	1.24	1.50	4.34
8	Mountains	Bridge	2.02	1.91	4.65
9	Living room	Candle	1.03	0.95	4.17
10	Backyard pool	Cap	2.05	1.60	5.26
11	Castle	Flag	0.90	1.24	4.91
12	Kitchen	Clock	2.09	2.57	5.41
13	Construction site	Crane	3.26	1.48	5.44
14	Houses	Fire hydrant	1.95	1.86	6.81
15	Lake	Ship	1.79	3.95	5.17
16	City line	Flag	1.21	0.97	4.18
17	Bedroom	Flower	1.55	1.43	6.36
18	Patio	Vase	1.34	1.26	2.92
19	Dining room	Bust	2.09	1.97	4.71
20	Porch	Cow	1.79	1.19	4.24
21	Boats under pier	Flag	3.91	2.43	4.94
22	Houses	Recycling box	1.76	2.47	6.04
23	Bedroom	Pen	1.31	2.26	5.67
24	Living room	Candles	2.62	1.72	6.86
25	Church–temple	Piano	1.69	2.17	6.01
26	Classroom	Overhead machine	1.71	1.26	5.56
27	City line	Pool	0.76	1.19	5.42
28	Living room	Lamp	1.81	1.50	6.25
29	Restaurant	Red bicycle	2.26	3.60	6.39
30	Living room	Ship	1.57	2.17	5.11
31	Kitchen	Fruit bowl	2.31	2.66	5.14
32	Living room	Stereo	2.14	1.93	4.45
33	Kitchen	Stool	3.16	1.93	3.94
34	Street	Traffic cone	1.17	1.19	4.22
35	Living room	Television	1.07	1.76	3.13
36	Living room	Banjo	2.81	1.33	2.14

[a] Eccentricity was calculated as the distance from the center of the scene to the center of the target object region.

Appendix B

Size and Eccentricity of Target Objects for Experiments 2, 3, and 4

Item	Scene	Changed items	Targets	Height (in degrees)	Width (in degrees)	Eccentricity (in degrees)[a]
1	Hardware store	Yes	Snow shovel	1.95	1.28	3.36
2	Fishing boats		Bell	1.53	0.93	6.91
3	Storefront		Bench	1.55	2.22	6.03
4	Backyard garden	Yes	Hanging flower pot	2.36	1.95	2.77
5	Venice street		Boat	1.26	1.52	1.83
6	Living room		Boot	3.88	2.40	4.93
7	Living room	Yes	Pillow	2.88	2.64	5.75
8	Mountains	Yes	Truck	0.66	0.91	2.69
9	Living room	Yes	Flowers	1.81	1.72	5.18
10	Backyard pool	Yes	Drink glass	1.74	1.69	5.06
11	Castle		Flag	1.66	1.12	4.56
12	Kitchen	Yes	Stool	3.29	3.00	3.63
13	Construction site	Yes	Portable toilet	2.21	1.93	3.20
14	Houses		Fire hydrant	1.95	1.86	6.81
15	Lake		Ship	2.52	4.17	5.06
16	City line		Flag	1.00	1.07	4.06
17	Bedroom	Yes	Big bird	1.84	1.64	1.63
18	Patio		Vase	2.83	2.38	2.55
19	Dining room	Yes	Fruit bowl	4.66	4.41	4.62
20	Porch	Yes	Door mat	1.50	3.40	2.26
21	Boats under pier		Flag	3.91	2.43	4.94
22	Houses		Recycling box	1.76	2.47	6.04
23	Bedroom	Yes	Flowers	1.88	1.91	3.05
24	Living room	Yes	Cabinet	2.78	1.91	6.43
25	Church–temple		Piano	1.69	2.17	6.01
26	Classroom	Yes	Wall clock	1.55	1.69	2.03
27	City line	Yes	Harbor	2.52	3.62	2.78
28	Living room	Yes	Painting	3.03	3.43	3.47
29	Restaurant		Red bicycle	2.26	3.60	6.39
30	Living room		Ship	1.57	2.17	5.11
31	Kitchen		Fruit bowl	2.31	2.66	5.14
32	Living room		Stereo	2.14	1.93	4.45
33	Kitchen		Stool	3.16	1.93	3.94
34	Street		Traffic cone	1.17	1.19	4.22
35	Living room		Television	1.07	1.76	3.13
36	Living room	Yes	Statue	2.91	2.12	2.61

[a] Eccentricity was calculated as the distance from the center of the scene to the center of the target object region.

Received August 1, 2005
Revision received August 23, 2006
Accepted August 25, 2006 ■