Perception & Psychophysics
2002, 64 (6), 882-895

Eye movements and scene perception: Memory for things observed

DAVID E. IRWIN
University of Illinois at Urbana-Champaign, Champaign, Illinois

and

GREGORY J. ZELINSKY
State University of New York, Stony Brook, New York

In this study, we examined the characteristics of on-line scene representations, using a partial-report procedure. Subjects inspected a simple scene containing seven objects for 1, 3, 5, 9, or 15 fixations; shortly after scene offset, a marker cued one scene location for report. Consistent with previous research, the results indicated that scene representations are relatively sparse; even after 15 fixations on a scene, the subjects remembered the position/identity pairings for only about 78% of the objects in the scene, or the equivalent of about five objects-worth of information. Report of the last three objects that were foveated and of the object about to be foveated was very accurate, however, suggesting that recently attended information in a scene is represented quite well. Information about the scene appeared to accumulate over multiple fixations, but the capacity of the on-line scene representation appeared to be limited to about five items. Implications for recent theories of scene representation are discussed.

Consider walking into a room that you have never been in before. The room contains more information than you can perceive in a single glance, so you direct your eyes, via rapid movements called *saccades*, so as to fixate objects around you. Each eye movement causes visual information to be swept across your retinas, producing a blur or a smear that is not perceived, because of saccadic suppression (see Matin, 1974, for a review). Because of saccadic suppression, visual information about the room is registered in isolated glimpses that are separated in time. Furthermore, the contents of these isolated glimpses are not identical, because the retinal positions of the objects in the room change from one fixation to the next. Despite this, you perceive the room as unified, stable, and continuous, with objects maintaining their positions in space (see Bridgeman, van der Heijden, & Velichkovsky, 1994, for a review). There is no feeling of "starting anew" with each fixation; rather, you develop an on-line mental representation, or situation model (De Graef, 1992; cf. Van Dijk & Kintsch, 1983), of the room that describes where you are, what objects are present, what they look like, and where they are located.

Or do you? In truth, surprisingly little is known about the nature of the mental representation that is built up across successive eye movements. Intuitively, it seems very detailed, but considerable research indicates that it is not. For example, people are unable to integrate visual patterns viewed in successive eye fixations so as to perceive some composite pattern (e.g., Bridgeman & Mayer, 1983; Irwin, Brown, & Sun, 1988; Irwin, Yantis, & Jonides, 1983; O'Regan & Levy-Schoen, 1983; Rayner & Pollatsek, 1983). Furthermore, stimulus displacements during saccadic eye movements are frequently undetected unless the displacements are large (e.g., Bridgeman, Hendry, & Stark, 1975), and changes in visual patterns are difficult to detect unless the patterns are simple (e.g., Irwin, 1991). Similarly, changing the visual characteristics of words (by varying letter case, for example) during eye movements has no effect on reading (e.g., McConkie & Zola, 1979) or word naming (e.g., Rayner, McConkie, & Zola, 1980), and changes in the spatial positions or visual properties of pictures during eye movements are frequently undetected (e.g., Currie, McConkie, Carlson-Radvansky, & Irwin, 2000; Grimes, 1996; Henderson, 1997; Henderson & Hollingworth, 1999b; Hollingworth & Henderson, 2000, 2002; Hollingworth, Williams, & Henderson, 2001; McConkie & Currie, 1996) and have little or no effect on picture naming (e.g., Pollatsek, Rayner, & Henderson, 1990). This insensitivity to change is not restricted to changes made during eye movements, however, but occurs also when a blank screen is interposed between an original and a changed display (e.g., Blackmore, Brelstaff, Nelson, & Troscianko, 1995; Pashler, 1988; Phillips, 1974; Rensink, O'Regan, & Clark, 1997; Simons, 1996), when film cuts occur in motion pictures (e.g., Hochberg, 1986; Levin & Simons, 1997), or when one's view of the world

This research was supported by National Science Foundation Grant SBR 96-15988 to D.E.I. We thank Bruce Bridgeman, Neal Cohen, Heiner Deubel, Peter De Graef, Albrecht Inhoff, Art Kramer, Gordon Logan, Lester Loschky, George McConkie, and Jennifer Ryan for helpful comments on the research, Gary Wolverton for programming assistance, and Emily Kyro for assistance with data collection. Correspondence concerning this article should be addressed to D. E. Irwin, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel St., Champaign, IL 61820 (e-mail: dirwin@s.psych.uiuc.edu).

is interrupted by an occluding object (Simons & Levin, 1998). In short, people are strikingly bad at detecting changes in the visual world whenever low-level motion cues are eliminated by eye movements, flicker, or other disruptive visual events (see Simons & Levin, 1997, for a review). These findings indicate that people do not have unlimited access to a richly detailed mental representation of a scene; otherwise, these changes would be quite noticeable.

So what is contained in one's on-line representation of a scene? It seems clear that some information is represented; we are not startled every time we move our eyes and take in a new view of the world, for example. The purpose of the present research was to begin to determine the characteristics of on-line scene representations. In particular, we investigated how much people remember from a simple scene, what factors influence their memory, and how their on-line scene representation changes as a function of the number of fixations on the scene.

To address the nature of on-line scene representations, a partial-report technique (Averbach & Coriell, 1961) was used to assess people's memory for pictures of simple scenes. The scenes always used a baby's crib as a background and contained seven objects arrayed in seven fixed locations in the crib. The subjects began each trial by fixating a plus sign in the bottom center of the crib scene. They were instructed to inspect the scene and to try to remember which objects were located where. The subject's eye position was monitored with an eyetracker, and during some critical saccade (the 1st, the 3rd, the 5th, the 9th, or the 15th), the scene was erased, a delay ensued until the subject's eye settled into its new position, and then a partial-report probe, a bar marker, was presented near one of the previously occupied crib locations. The subject's task was to report the object that had appeared in the probed position.

METHOD

Subjects

Twelve observers participated in this experiment. Six subjects completed a version of the experiment in which 1, 3, or 5 fixations were allowed on the scene, whereas the other 6 subjects completed a version of the experiment in which 3, 9, or 15 fixations were allowed on the scene. With the exception of the second author, all of the observers were naive as to the specific questions under investigation, and all were paid $8/h for their participation. None of the subjects required visual correction and all had normal color vision (both by self-report).

Stimuli

The stimuli were scenes consisting of seven objects (teddy bear, baby bottle, toy car, box of crayons, baby doll, rubber ducky, and toy trumpet) appearing in a baby's crib. These objects were arranged in a semicircle around the observer's initial fixation point; a 7° visual angle was subtended by this point and each object's center. The angular positions of these seven objects relative to the fixation point were fixed at 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, and 157.5°, although the identity of the object appearing at a particular position varied from trial to trial. As a result of these placement constraints, the minimum and maximum center-to-center object separations were 2.7° and 13.0°, respectively. Figure 1 shows a grayscale reproduction of one of the actual scenes used in the experiment. Each 756 × 486 pixel image subtended 18° of visual angle horizontally and 11.6° vertically at a viewing distance of 82 cm. The component objects, although irregular in shape, could each fit within a 2.4° square bounding box. Both the crib background and the objects were displayed at 72 pixels/in. in 16-bit RGB color. Details regarding the construction and software manipulation of these images can be found elsewhere (Zelinsky, 1999, 2001). The stimuli were displayed on a Princeton Ultra-Synch monitor controlled by an ATVista display controller card in a 386 computer and refreshed at 60 Hz. The pictures were displayed at a comfortable luminance in a dimly illuminated room; shutter tests (McConkie & Currie, 1996) confirmed that visible phosphor persistence decayed from the display screen within 12 msec after stimulus offset.

Procedure

The observers viewed the above-described scenes in preparation for a recognition judgment following a spatial probe. Preceding each trial, there was a fixation target at the point corresponding to the origin of the semicircle in the study scene. The observers were asked to look carefully at this target (a 0.5° X in a box) and then to press a hand-held button to initiate the trial. Once a scene appeared (Figure 2A), it remained visible for a variable period of time, an interval constituting our primary experimental manipulation. The observers were allowed to freely view the scenes for exactly 1, 3, or 5 fixations (in Version 1 of the experiment) or for exactly 3, 9, or 15 fixations (in Version 2 of the experiment), depending on the condition being instantiated on that particular trial. As the observer's gaze moved away from one of these critical fixations (e.g., the saccade following the 3rd fixation in a 3-fixation trial, as illustrated in Figure 2A), the scene was immediately replaced with a dark screen before the eye reached its next intended destination. Note that the initial fixation on the scene counted as one of these critical fixations, meaning that the observers would not be permitted to fixate an object in the 1-fixation condition.

Just as the actual time that the observers had to inspect the scene depended on both the fixation criteria and the durations of the individual fixations, so the blank delay immediately following the study scene also depended on the observer's oculomotor behavior. The reason for this variable delay was to ensure that the eye was stable prior to the onset of the spatial probe. If the probe appeared while the eye was still in motion, this motion might mask the sudden onset of the probe, causing an ineffective direction of attention to the desired location. Our criteria for stability required that the eye be moving at a velocity less than 12 deg/sec for at least 50 msec following the end of the critical saccade. If the eye exceeded this velocity threshold at any time during the 50 msec, the counter would be reset, and another delay would be initiated. As a result of these stability criteria, the actual delay between study scene offset and probe onset averaged 153 msec. Following the postscene delay, the target was designated by a perfectly valid spatial probe flashed for a brief 50 msec (Figure 2B). The probe was a white bar, 0.75° in length, positioned along an imaginary line passing through the fixation point and the center of the target object. The distance between the fixation point and the nearest point on the probe was 8.7°, placing the probe just beyond the region described by the target bounding box (i.e., the target and the probe did not spatially overlap). A 2-sec blank interval followed the offset of the probe, after which appeared a seven-alternative forced-choice response grid (Figure 2C). The grid depicted the seven study objects in fixed display locations (e.g., the teddy bear always appeared in the upper-left position of the arch-shaped configuration). Object identity was mapped consistently to display position so as to minimize the need for the subjects to search for the target during response. The observers' task was to select the object in the response grid indicated

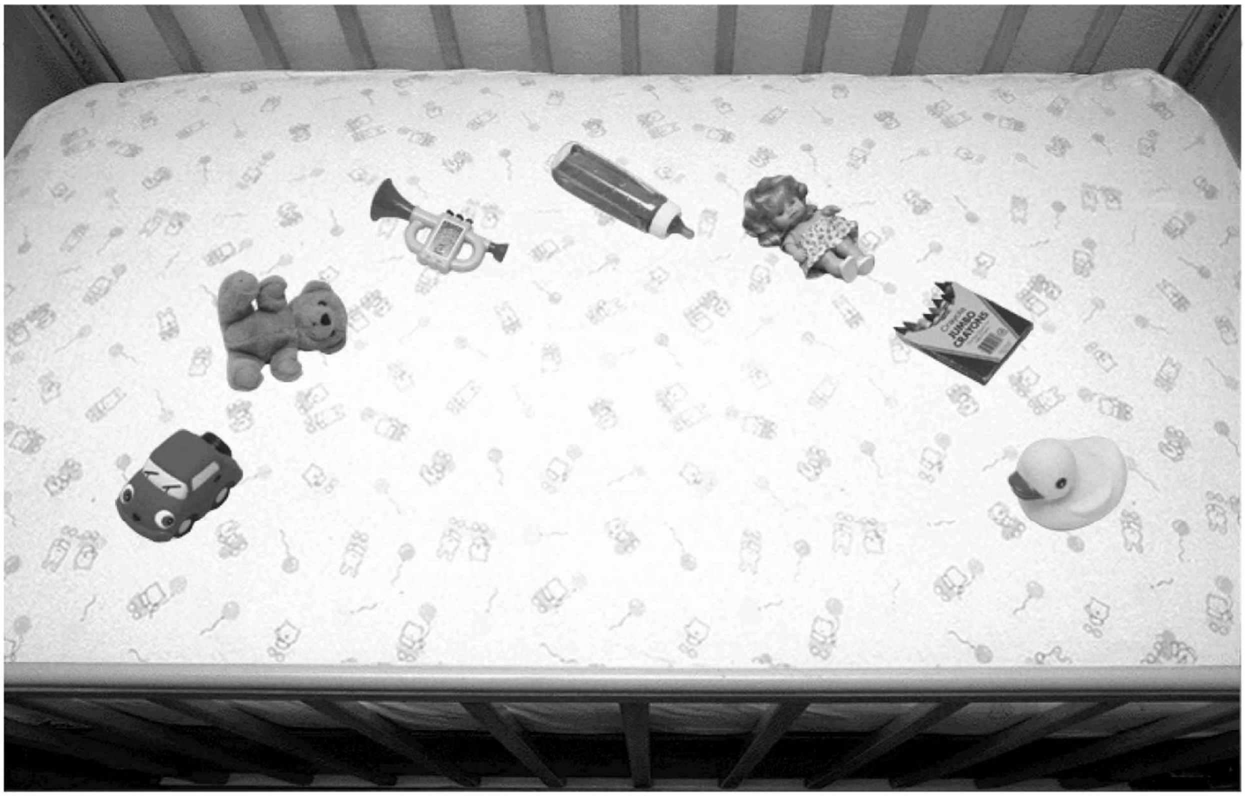

Figure 1. Grayscale reproduction of one of the scenes used in the experiment.

by the spatial probe. Because the observers were on a bite-bar and could not easily speak, their method of indicating a selection was simply to look at the desired object, an act that caused a white box to be drawn around the fixated item, and then to press a hand-held button when satisfied with their selection. There was no accuracy feedback, and the time commitment per subject was less than 2 h.

Design

Each observer participated in 147 trials, with these trials evenly divided into three randomly interleaved fixation conditions (1, 3, or 5 in Version 1 of the experiment and 3, 9, or 15 in Version 2 of the experiment). Postexperiment questioning revealed that the naive observers were unaware of these variable scene presentation times being contingent upon their oculomotor behavior. Within each condition, the spatial probe and, therefore, the target appeared equally often (seven times) at each of the seven display locations. Each object also served as the target seven times, resulting in 147 trials (3 fixation conditions × 7 locations × 7 objects). Except for the above constraints, the seven objects in each scene were randomly assigned to the seven locations.

Eye Movement Recording

A Fourward Technologies Generation V dual Purkinje-image eyetracker, interfaced with a Tecmar 8-bit 12-channel analog-to-digital converter, was used to sample eye position every millisecond throughout each trial. The spatial precision of this tracker during fixation is ±1 min of visual angle. To achieve this level of precision, a dental impression bite-bar was constructed for each subject, who was asked to remain on the bite-bar until scheduled breaks after every 30 trials. Prior to the start of the experiment, and again before every block of trials, the observers performed a calibration procedure consisting of accurately fixating each of nine targets demarcating the 18° × 11.6° field of view. As an internal validity check, the calibration procedure would not terminate until repeated fixations on each target resulted in an eye position error of less than 0.3°. Other than these calibration instructions and a reminder to look at the fixation cross whenever it appeared on the screen, the observers were not told how to direct their gaze in this experiment. Saccadic eye movements were extracted on line, using a velocity-based algorithm implementing a roughly 12.5 deg/sec detection threshold. Fixation duration was defined as the period between the offset of one saccade and the onset of the next. Fixations that fell within the virtual 2.4° × 2.4° bounding box within which each object was presented were scored as being *on the object*.

RESULTS AND DISCUSSION

Discarded Data

Trial data were deleted from analysis if the eye was not within the fixation box when the trial started or if the eyetracker lost track of the eye during scene, blank, or probe presentation. The proportion of deleted trials did not vary significantly across conditions in Version 1 of the experiment; 18.7% of the trials were deleted from the 1-fixation condition, 17.0% from the 3-fixation condition, and 17.7% from the 5-fixation condition. A procedural change was implemented in Version 2 of the experiment in order to reduce the number of trials lost owing to inaccurate fixation on the fixation cross that preceded scene presenta-

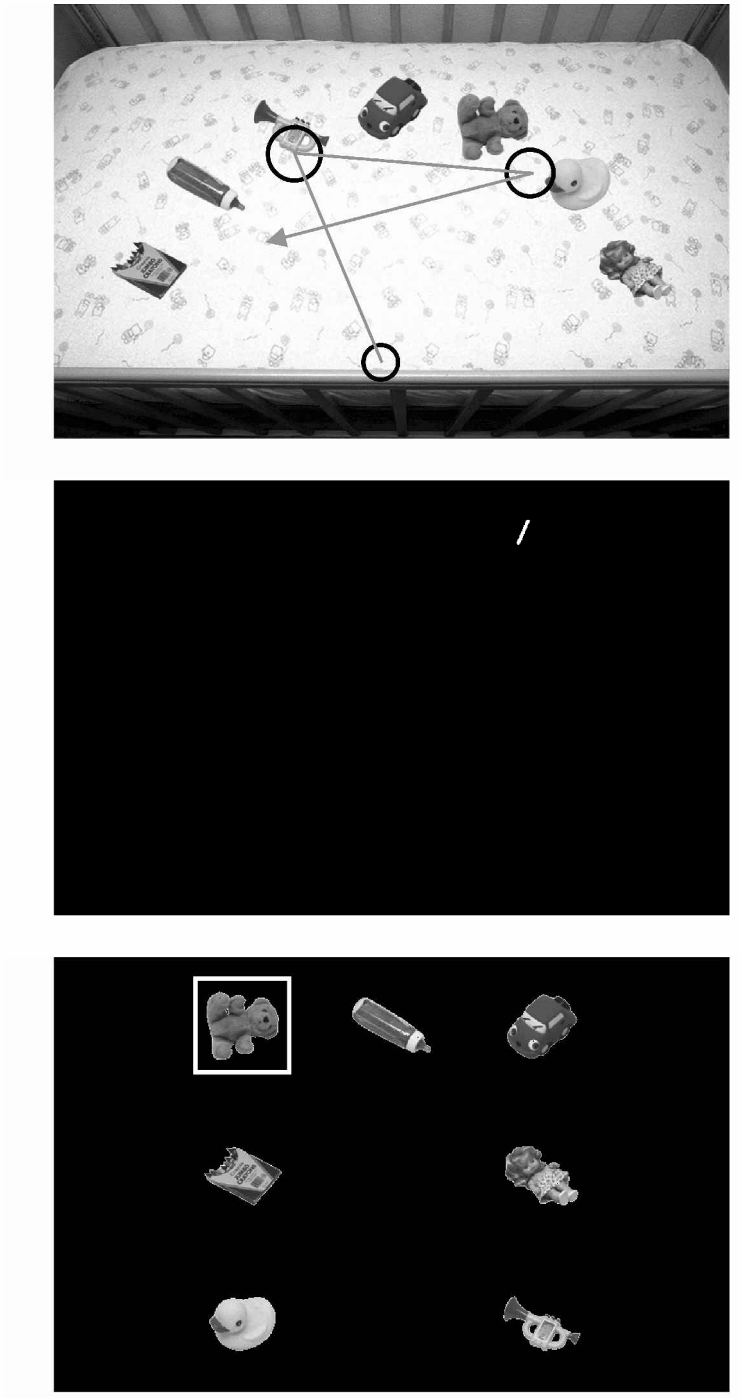

Figure 2. Highlights of the experimental procedure.

tion. In particular, to make certain that the subjects were indeed looking at this cross, the study scene would not appear until the gaze deviated from the center of the cross by less than 1° at the time of the buttonpress. This precaution eliminated the need to discard trials caused by anticipatory saccades to the fixed display locations. Thus, trial data were deleted from analysis in Version 2 of the experiment only if the eyetracker lost track of the eye during scene, blank, or probe presentation. The proportion of deleted trials in Version 2 did not vary significantly across

Table 1
Viewing Time and Number of Objects Foveated as a Function of Fixation Condition

Fixation Condition	Viewing Time (msec)	Objects Foveated
1	174	0.00
3	739	1.25
5	1,360	2.68
9	2,810	3.47
15	4,927	5.10

conditions; 5.4% of the trials were deleted from the 3-fixation condition, 10.9% from the 9-fixation condition, and 9.2% from the 15-fixation condition.

Scene Viewing Time and Number of Objects Foveated

Because the duration of the display on each trial was determined by fixation condition, scene viewing time increased as the number of fixations allowed on the scene increased (see Table 1). The mean viewing times ranged from 174 msec in the 1-fixation condition to 4,927 msec in the 15-fixation condition. As one might expect, the number of objects foveated in the scene also increased with the number of fixations allowed on the scene. In the 1-fixation condition, the scene was erased as soon as the eyes left the fixation box at the bottom of the scene, so no objects were foveated in this condition. At the other extreme, an average of 5.1 objects were foveated in the 15-fixation condition (see Table 1). The number of objects foveated in each condition was less than the number of fixations, because some fixations fell on no object and some objects were foveated more than once. Mean fixation duration tended to increase as the number of fixations on the scene increased (see Figure 3). The first 2 fixations on the scene were quite short (approximately 150 msec in duration), whereas the duration of subsequent fixations was captured reasonably well ($r^2 = .77$) by the following equation: fixation duration = 237.6 + 5.5(fixation number).

Given these differences in viewing time and in the number of objects foveated, it would be reasonable to expect that overall accuracy in reporting the probed object would increase across fixation conditions. This is discussed next.

Overall Accuracy

In Version 1 of the experiment, mean accuracy was 36.0% in the 1-fixation condition, 65.8% in the 3-fixation condition, and 67.2% in the 5-fixation condition. The effect of fixation condition was significant [$F(2,10) = 41.9$, $MS_e = 44.6$, $p < .001$]. A planned comparison based on the error term of the analysis of variance (ANOVA) showed that differences of 7.8% were significant at the .05 level; thus, all pairwise comparisons were significant, except for the difference between the 3-fixation condition and the 5-fixation condition. In Version 2 of the experiment, mean accuracy was 54.5% in the 3-fixation condition, 70.7% in the 9-fixation condition, and 77.7% in the 15-fixation condition. The effect of fixation condition was significant [$F(2,10) = 28.8$, $MS_e = 29.4$, $p < .001$]. A planned comparison based on the error term of the ANOVA showed that differences of 6.3% were significant at the .05 level; thus, all pairwise comparisons were significant. A *t* test comparing the two 3-fixation conditions was not significant [$t(10) = 2.1$, $p > .05$]. Mean accuracy at reporting the probed object as a function of the number of fixations on the scene is shown in Figure 4 (the data from the 3-fixation conditions were combined for this figure).

A number of secondary analyses were done to investigate whether practice or learning had any effect on accuracy. Because the same objects appeared in different positions on every trial, it seemed possible that the memory of where an object appeared on trial $n - 1$ might influence recall of where that object had appeared on trial n (e.g., because of proactive interference). There was no evidence that accuracy declined over trials, however. In Version 1 of the experiment, mean accuracy during the first, second, and last third of the experiment was 52.7%,

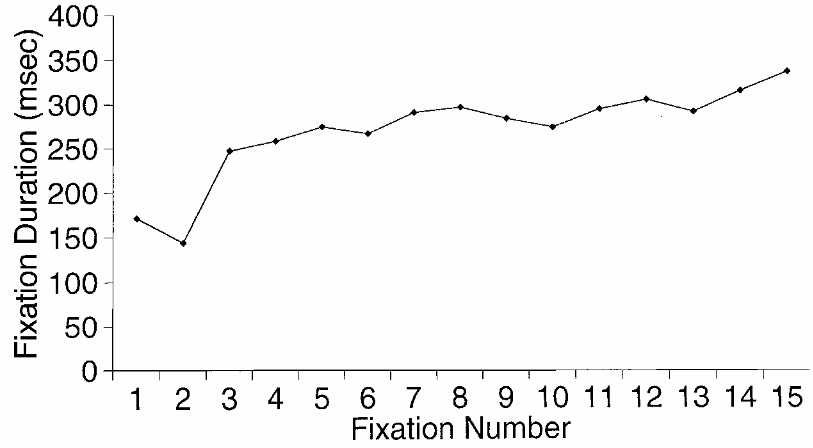

Figure 3. Mean fixation duration as a function of number of fixations on the scene.

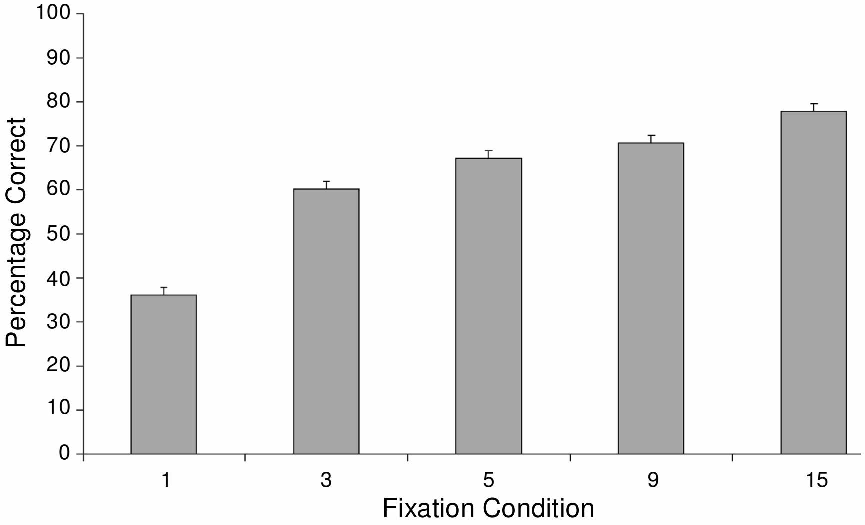

Figure 4. Mean accuracy as a function of the number of fixations on the scene (error bars represent the standard error of the mean).

55.1%, and 58.9%. In Version 2 of the experiment, the corresponding accuracies were 68.3%, 67.1%, and 68.3%. There was also no evidence that probing the same target object (e.g., the duck) in different positions on successive trials influenced performance; accuracy was 62.6% when the target object on trial *n* had also been the target object (but in a different position) on trial *n* − 1, and accuracy was 62.1% when different target objects were probed on trial *n* and trial *n* − 1 (there were no trials in which the same target object appeared in the same probed position on successive trials). In sum, there was no evidence that learning or repetition effects interfered with memory recall.

The analyses that follow were designed to illuminate what factors did affect overall accuracy.

Accuracy × Probe Position

Previous partial-report experiments have found that accuracy depends strongly on an item's position in a display (e.g., Mewhort, Campbell, Marchetti, & Campbell,

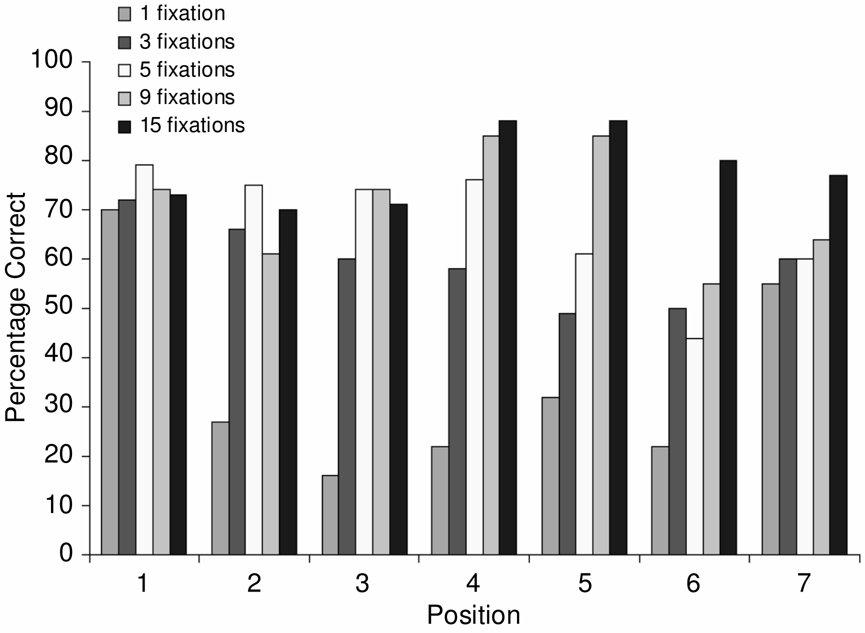

Figure 5. Mean accuracy as a function of probe position for each fixation condition.

1981). Figure 5 shows accuracy as a function of which position in the scene was probed for each fixation condition; Position 1 refers to the leftmost object in the scene, Position 7 to the rightmost, and Position 4 to the object directly above the fixation point. Accuracy in the one-fixation condition showed a U-shaped function with position, so that objects at the sides of the display were remembered better than objects at other array positions. There are probably several factors contributing to this result: There is less lateral masking for the two end items, previous researchers have found that both retinal acuity and visual attention gradients are not circular but, rather, are elongated in the horizontal dimension (e.g., Pan & Eriksen, 1993), and subjects may have shifted their attention preferentially to the end items as part of a high-level scanning strategy.

Of more interest is what happened to the accuracy × probe position function when more fixations were made on the display. Recall, for example, that overall accuracy increased as the number of fixations increased from 1 to 3 fixations on the scene; Figure 5 shows that not all array positions benefited equally, however. Rather, accuracy for the center-left items in the display increased more than accuracy for the center-right items in the display. Increasing the number of fixations from 3 to 5 had similar effects. In contrast, accuracy for the rightmost items in the scene increased, whereas accuracy for the leftmost items in the scene decreased, when the number of fixations on the scene increased from 5 to 9 and from 9 to 15. In general, Figure 5 shows that peak accuracy for positions in the array occurs early for positions on the left side of the array and late for positions on the right side of the array.

It seemed possible that the changes in accuracy across probe position with increasing number of fixations on the scene might be due to the subjects' viewing strategies. Figure 6 shows the percentage of fixations on different object positions in the scene as a function of fixation number (e.g., 33% of all second fixations fell on Position 1, 3% of all second fixations fell on Position 2, and so on).[1] In general, the pattern of fixations looks very similar to the pattern of accuracies shown in Figure 5: The subjects tended to fixate the leftmost items in the array early in scene viewing but shifted to rightward positions later in scene viewing. There was a great deal of variability between and even within subjects in terms of their viewing strategies, however. Two subjects (1 in Version 1 of the experiment and 1 in Version 2) fixated the objects from left to right on most trials; 1 subject fixated the objects from right to left on most trials; 1 subject moved from left to right on half the trials and right to left on the other half; 1 usually fixated Position 1, then Position 4, then Position 7, skipping over the other positions; 1 usually fixated Position 4 first and then moved left; 1 usually fixated Position 4 first and then moved right; and the remainder used a mixture of these strategies, changing from one to another during the course of the experiment. There was little difference in accuracy between the 3 subjects who consistently used a left-to-right or a right-to-left strategy (59.1%) and the 9 who did not (62.8%).

The similarity between the pattern of accuracies in Figure 5 and the pattern of eye fixations in Figure 6 suggests that accuracy was affected by whether (and perhaps when) the probed item was foveated during scene presentation. This was examined directly in the next two analyses.

Accuracy for Foveated Versus Nonfoveated Objects

On some trials, the subjects foveated the object that happened to be probed, whereas on other trials they did not. Figure 7 shows accuracy for foveated versus non-

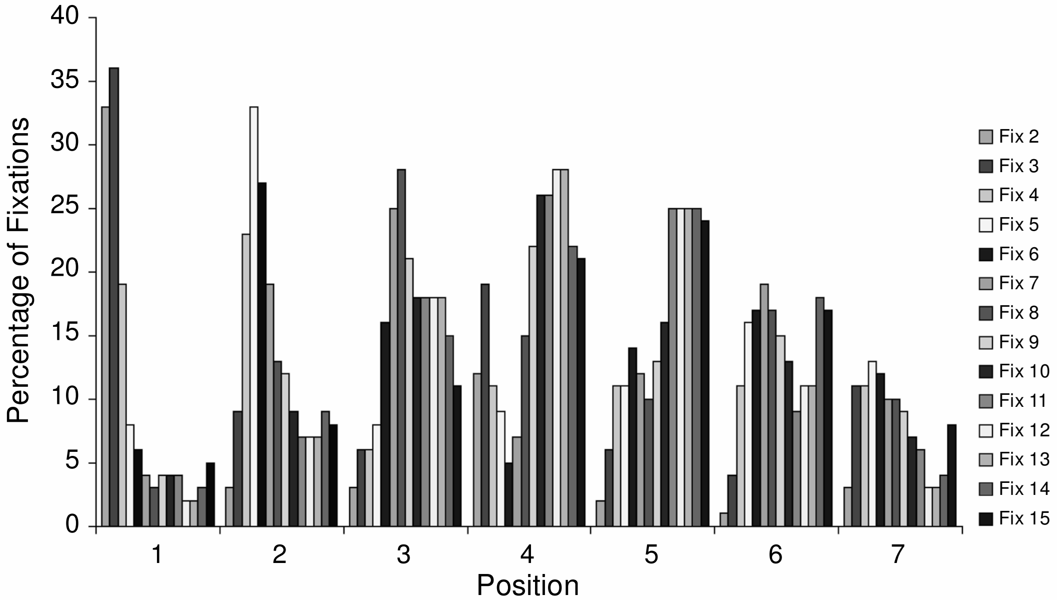

Figure 6. The percentage of fixations on each object position as a function of fixation number.

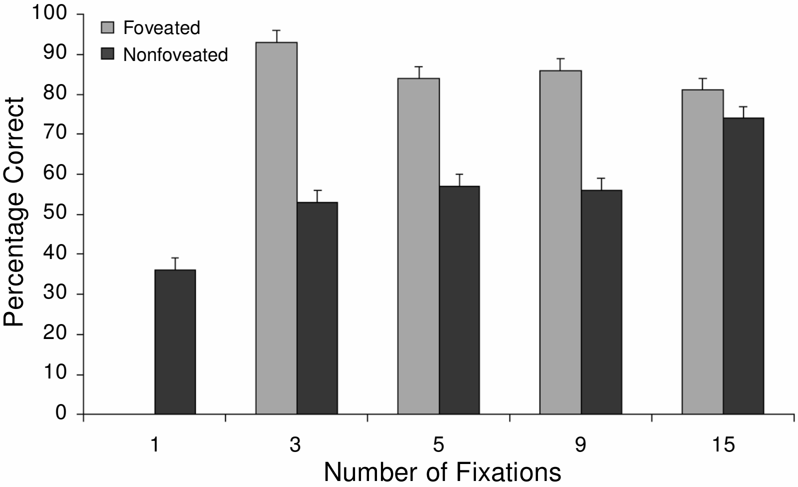

Figure 7. Mean accuracy for foveated versus nonfoveated targets as a function of the number of fixations on the scene (an error bar represents the standard error of the mean).

foveated targets as a function of the number of fixations on the scene. No targets were foveated in the 1-fixation condition, because the scene disappeared during the first saccade. In the 3-, 5-, and 9-fixation conditions, however, accuracy was much higher if the position that was probed had been foveated, as compared with when it had not been foveated, during scene presentation [92.9% vs. 53.0% in the 3-fixation case, $t(11) = 7.6$, $p < .001$; 84.2% vs. 56.5% in the 5-fixation case, $t(5) = 2.6$, $p < .05$; 85.8% vs. 55.9% in the 9-fixation case, $t(5) = 5.3$, $p < .005$]. The benefit of foveation (80.8% vs. 74.4%) was not significant when 15 fixations had been made on the scene, however [$t(5) = 1.0$, $p > .35$]. An inspection of Figure 7 suggests that this occurred for two reasons. First, accuracy for nonfoveated objects increased substantially (from 53.0% to 74.4%) as the number of fixations on the scene increased from 3 to 15, presumably because, with an increasing number of fixations, the eyes were landing near the nonfoveated objects even if they were not landing on them; second, accuracy for foveated objects decreased as the number of fixations increased. This suggests that recency of fixation on the target object might also be important (i.e., the 15-fixation condition includes fixations on the target further back in time than does the 3-fixation condition). This was examined in the next analysis.

Accuracy as a Function of Fixation Recency

Figure 8 shows accuracy for foveated targets only as a function of when, during the trial, the object had been

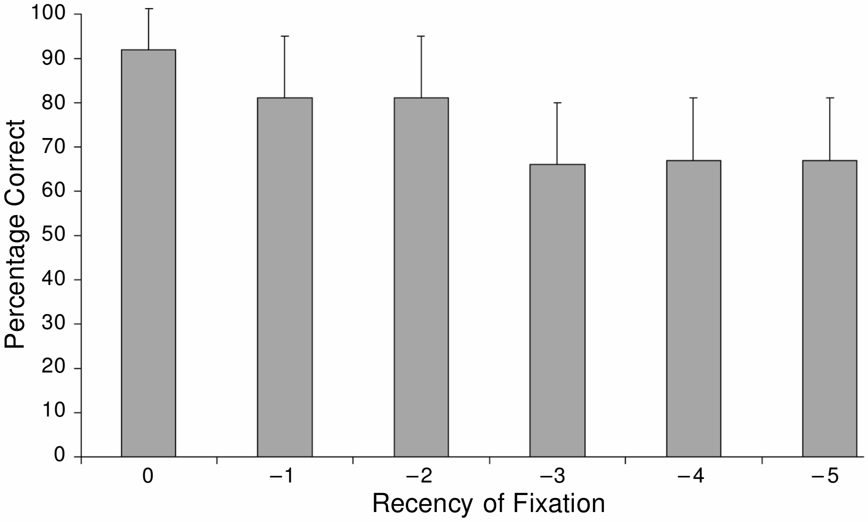

Figure 8. Mean accuracy for foveated targets as a function only of when during the trial the target had been foveated (an error bar represents the standard error of the mean).

foveated. If the object that had been foveated just before the critical saccade (i.e., the saccade that terminated the scene) happened to be probed for report (denoted as fixation recency 0 in the figure), the subjects reported it correctly on over 90% of the trials. If the object probed for report had been foveated 1 or 2 fixations earlier (corresponding to fixation recencies −1 and −2 in the figure), accuracy was slightly lower (approximately 80%). Probing an object that had been foveated 3 or more fixations before the critical saccade resulted in accuracies that were considerably lower (approximately 65%) and not much different from the accuracy obtained when a nonfoveated object had been probed for report (59%). The effect of fixation recency (using six levels of recency) was significant [$F(5,25) = 2.85$, $MS_e = 243.2$, $p < .05$] in a one-way repeated measures ANOVA that was conducted on the data from Version 2 of the experiment (this version allowed 3, 9, or 15 fixations on the display and thus provided a wide range of fixation recencies). A planned comparison based on the error term of the ANOVA showed that differences of 14.5% were significant at the .05 level; thus, accuracy for the three objects foveated just before the critical saccade (i.e., with fixation recencies of 0, −1, and −2) was significantly higher than accuracy for objects foveated 3 or more fixations back. These results suggest that fixation recency does influence accuracy; in particular, it appears that people remember the last three objects foveated in the scene much better than they remember objects foveated earlier in time (see Zelinsky & Loschky, 1998, for similar results). There was no evidence of a primacy effect; if the object probed for report was the first object foveated (and had been foveated 3 or more fixations before the critical saccade), accuracy was 66%.

Accuracy as a Function of the Number of Objects Foveated

Given the importance of foveation to scene memory that is revealed by the preceding analyses, one might expect that overall accuracy would depend on the number of unique objects that were foveated in the scene (recall that 7 objects appeared in the scene). This relationship is shown in Figure 9. Accuracy increased as 0–6 objects were foveated, but there was no additional increase in accuracy from 6 to 7 objects. Accuracy peaked at 80% even when 6 or 7 unique objects were foveated. An accuracy level of 80% corresponds to memory for 5.4 objects in the scene.[2] This result suggests that scene representations contain no more than about five object/position pairings, regardless of the number of fixations on the scene. This is not to say that 5 discrete objects in all their full detail are stored on every trial, however; rather, this estimate represents the average amount of information stored from what is most likely a stochastic process. Thus, it is probably most accurate to say that approximately 5 objects' "worth" of information is stored in the on-line scene representation.

Accuracy When Final Saccade Was Directed at the Target Versus Directed Elsewhere

Previous research has shown that movements of attention precede movements of the eyes to locations in space, thereby facilitating the processing of information that is present at the saccade target location (e.g., Deubel & Schneider, 1996; Henderson, 1993; Henderson, Pollatsek, & Rayner, 1989; Hoffman & Subramaniam, 1995; Irwin & Gordon, 1998; Klein, 1980; Klein & Pontefract, 1994; Kowler, Anderson, Dosher, & Blaser, 1995; Rayner, McConkie, & Ehrlich, 1978; Shepherd, Findlay,

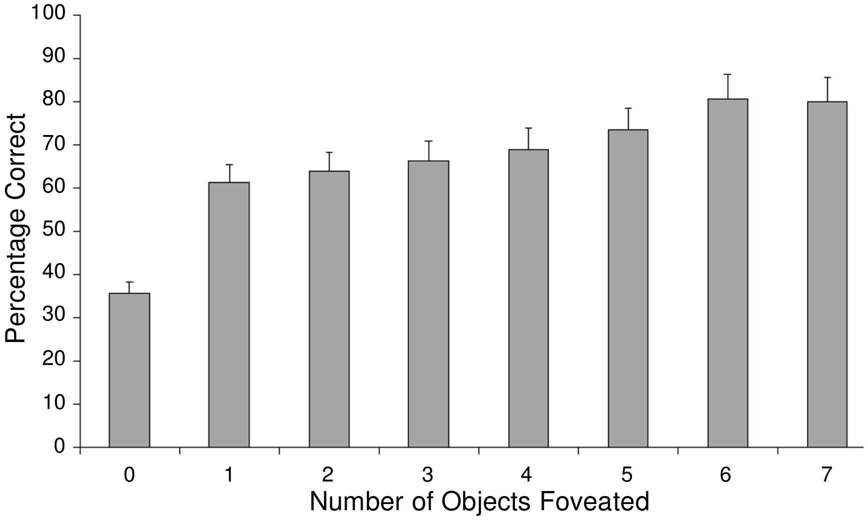

Figure 9. Mean accuracy as a function of the number of unique objects that were foveated during scene presentation (an error bar represents the standard error of the mean).

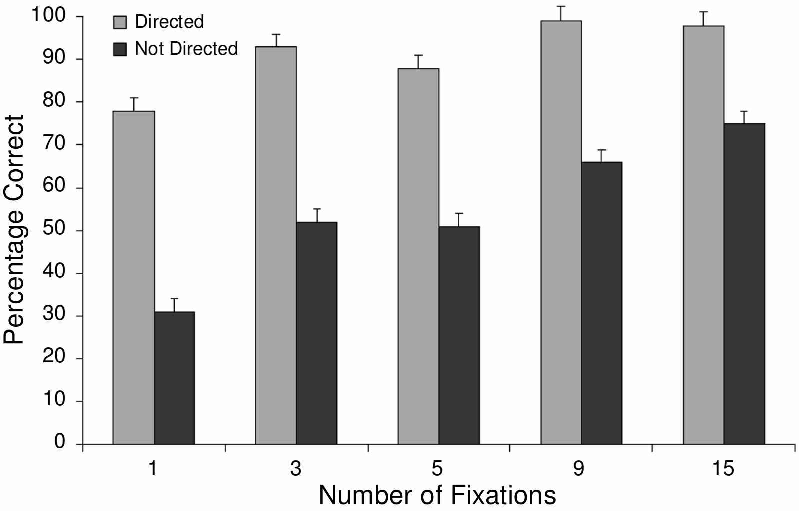

Figure 10. Mean accuracy when the final saccade was directed at the target versus not directed at the target as a function of the number of fixations on the scene (an error bar represents the standard error of the mean).

& Hockey, 1986). Thus, we expected that accuracy would be higher when the critical saccade was directed at the object that happened to be probed on any given trial, as compared with when the critical saccade was directed toward an unprobed object. To elaborate, consider a one-fixation trial in which the baby bottle was presented in the leftmost position of the display. Suppose that the subject moved his/her eyes from the central fixation box toward this position. Because this was a one-fixation trial, the scene disappeared as soon as the saccade was initiated to the bottle, so nothing was present on the screen when the eyes landed. A short time later the probe was presented at one of the display positions, and on some trials it just happened to be presented at the location to which the eyes were sent on the critical saccade; the question of interest is whether accuracy was higher when the probed item happened to be the item that the eyes were sent to on the critical saccade than when it was not.

Figure 10 shows accuracy for trials in which the probed item was or was not the target of the critical saccade in each fixation condition (by *target of the saccade*, we mean that the eyes landed within the virtual 2.4° square bounding box that defined object position). Only trials in which the probed item was not fixated at any time during scene presentation are included. In every fixation condition, accuracy was much higher when the probed item was the target of the final saccade than when it was not [77.8% vs. 30.9% in the 1-fixation case, $t(5) = 2.6$, $p < .05$; 92.9% vs. 51.5% in the 3-fixation case, $t(11) = 6.0$, $p < .001$; 87.9% vs. 51.3% in the 5-fixation case, $t(5) = 3.6$, $p < .02$; 99.8% vs. 65.8% in the 9-fixation case, $t(5) = 10.2$, $p < .001$; and 98.3% vs. 75.1% in the 15-fixation case, $t(5) = 4.9$, $p < .005$].

In sum, accuracy at reporting the probed item was very high if the probe happened to be presented at the location to which the eyes had been sent on the critical saccade, even though the object at that location had never been foveated during the trial; merely being the target of the final saccade increased its memorability considerably (see Henderson & Hollingworth, 1999b, for similar findings). This presumably occurred because attention preceded the eyes to the saccade target location. The effect of attention is to increase the likelihood that information at the saccade target location will be encoded into one's mental representation of the scene. In fact, it is possible that many of the beneficial effects of foveating an object discussed above are not due to foveation per se but, rather, to the attention shift that preceded it.

DISCUSSION

In the last 20 years, there has been an explosion of interest in the properties of on-line scene representation (e.g., Aginsky & Tarr, 2000; Blackmore et al., 1995; De Graef, 1992; Grimes, 1996; Hayhoe, 2000; Henderson & Hollingworth, 1999a, 1999b; Hollingworth & Henderson, 2000, 2002; Hollingworth et al., 2001; Intraub, 1997; Irwin, 1991, 1992; Irwin & Andrews, 1996; Irwin et al., 1988; Irwin et al., 1983; Irwin, Zacks, & Brown, 1990; Levin & Simons, 1997; McConkie & Currie, 1996; Mondy & Coltheart, 2000; O'Regan, 1992; O'Regan & Levy-Schoen, 1983; Pringle, Irwin, Kramer, & Atchley, 2001; Rayner & Pollatsek, 1983; Rensink, 2000; Rensink et al., 1997; Simons, 1996, 2000; Simons & Levin, 1998; Wallis & Bulthoff, 2000). One common finding is that scene representations appear to be sparse and undetailed. Consistent with previous research, the results of our experiment also indicate that scene representations are relatively sparse; even after 15 fixations on a scene, our subjects remembered the position/identity pairings of only about 78% of the objects in our seven-object displays, or the equivalent of about five objects' worth of information. But more important, our results provide new information about what factors influence the contents of the on-line scene representation and how this representation changes during the course of scene viewing. In particular, our analyses suggest that the last three items that are foveated and the item about to be foveated are actually remembered quite well, much better than other items in a scene. In other words, on-line scene representations are dynamic, so that the memory traces of recently fixated objects are much stronger than those of objects viewed several fixations back in time. Information about a scene does appear to accumulate over multiple fixations, but the capacity of the on-line scene representation appears to be limited to about five items. In most respects, the results of the present study are consistent with the results of other studies that have used the transsaccadic partial-report technique to investigate integration across saccades with nonscene displays (e.g., Irwin, 1992; Irwin & Andrews, 1996; Irwin & Gordon, 1998).

What is the nature of the memory representation underlying performance in our task? Given that the same seven objects appeared in fixed positions on each trial, one might imagine that subjects would simply adopt a left-to-right reading strategy and would store an ordered list of object names in verbal working memory. Then, when the probe was presented, they would scan their memorized list and report the appropriate item. Three aspects of our data seem inconsistent with this hypothesis, however. First, as was noted earlier, for the most part the subjects did not use a left-to-right reading strategy, nor did they even adopt any consistent strategy across trials. This variability would greatly complicate memory retrieval if the subjects had stored an ordered list of object names in verbal working memory, because there would be no consistent mapping between probe position and the position of an item in verbal working memory. That is, if the memory probe appeared at Position 6, say, this would correspond to the sixth position in the list if the subject scanned the array from left to right, the second position in the list if the subject scanned from right to left, the third position if the subject started in the middle and then moved right, and so on. The fact that the subjects did not adopt a fixed scanning order suggests to us that they were not storing an ordered list of object names in verbal working memory. Second, the fact that the subjects remembered only a maximum of five objects from the display also seems inconsistent with the verbal working memory hypothesis, because the capacity of verbal working memory is generally assumed to be between five and nine items; given 5 sec to study the display, we would think that the subjects would be able to remember more than five items if they were using verbal working memory. Finally, the absence of a primacy effect in our data also seems inconsistent with the verbal working memory hypothesis.

Instead, we believe that our results are more consistent with the object-file theory of transsaccadic memory (Irwin, 1992, 1996; Irwin & Andrews, 1996). This theory is based on the theoretical framework for object perception proposed by Kahneman and Treisman (1984) and Kahneman, Treisman, and Gibbs (1992). It contains four levels of representation: feature maps, which register independently the presence of different sensory features in a scene, such as color and shape; a master map of locations, which registers where in a scene features are located; temporary object representations called *object files*, which are formed when features are conjoined into unitary wholes via attention and which thus represent what objects are located where in a scene; and an abstract, long-term recognition network that stores descriptions of objects along with their names. Object files may contain visual, verbal, and semantic information about objects, because they have access to long-term memory, as well as to perceptual input. According to the object-file theory of transsaccadic memory, relatively little information is preserved across eye movements. Rather, transsaccadic memory (and by extension, the on-line scene representation) consists of 3—5 object files that contain position and identity information for items that have been attended in the scene and of residual activation in the feature maps and in the long-term recognition network. The results of the present study are consistent with the object-file theory because, even after 15 fixations on a scene, memory for the scene was limited and consisted largely of the last few objects that had been attended in the scene.

Although the results of the present study are consistent with the object-file theory, they seem not entirely consistent with another theory of scene representation proposed recently by Rensink (2000). According to Rensink's *coherence theory*, in the absence of focused attention, volatile low-level proto-objects are continually being formed and replaced rapidly and in parallel across the visual field, reflecting whatever is present on the retina at any moment in time. Focused attention provides a *coherence field* that selects a small number of these proto-objects to form a stable individuated object that has spatiotemporal continuity. This stable object does not necessarily correspond to a single element in the visual field, however, but may represent a supra-object (cf. Pylyshyn's FINSTs; Pylyshyn & Storm, 1988). After focused attention is released, the object loses its coherence and dissolves back into its constituent proto-object parts. Once that occurs, there is little or no consequence of having been attended. According to coherence theory, visual short-term memory corresponds to the coherence field; it contains object tokens (i.e., specific instantiations at a particular place and time) that are currently in the focus of attention but that disappear when attention is with-drawn. The coherence theory does allow for the existence of a nonvisual short-term memory for previously attended items, but it posits that this memory contains information only about object types (general definitions abstracted away from specific spatiotemporal contexts).

The results of the present study seem to raise problems for the coherence theory, because object tokens were retained in memory for recently fixated objects that were no longer in the focus of attention. Consistent with coherence theory, we found that accuracy was very high when the object at the final saccade target location (hence, the object at the focus of attention) was probed for report. However, we also found that accuracy was quite high for the three objects that had been fixated just prior to this. Since attention movements precede eye movements in an obligatory fashion, the three objects viewed in prior fixations could not also be at the focus of attention. The fact that accuracy was not uniform for the four objects in question also indicates that they were not all simultaneously at the focus of attention. Rather, the objects viewed during the fixations just preceding the final saccade benefited from prior attentional processing. This is inconsistent with the claim that once focused attention is released, objects dissolve back into their constituent proto-objects, with no consequence of having been attended. Note also that object tokens, and not object types, were being remembered in our experiment; in order to respond correctly to the partial-report probe, the subjects had to remember the position and the identity of the probed item—that is, they had to remember an object token. Merely remembering that an object type had appeared somewhere in the display would not lead to an accurate response, especially since the same seven items appeared on every trial.

Although the present study provides important information about the factors that influence the contents of online scene representations, two limitations should be noted. One is that our study measured only short-term memory contributions to the on-line scene representation, and not potential long-term memory contributions. The same seven objects appeared in the same scene context on every trial, so top-down contributions to performance were essentially eliminated. In the real world, scene representations are bound to be influenced by knowledge about what kinds of objects are typically found in a particular scene context and where those objects are usually located. Indeed, recent research by Hollingworth and Henderson (2002; see, also, Hollingworth et al., 2001), in which a novel saccade-contingent change detection procedure was used, has demonstrated that long-term memory plays a crucial role in the construction and maintenance of on-line scene representations.

A second limitation of our study is that it involved only a test of explicit memory. Recently, several investigators have found evidence that explicit tests of scene memory may underestimate the amount of information in the representation (e.g., Fernandez-Duque & Thornton, 2000; Hollingworth et al., 2001; Williams & Simons, 2000). For example, Hollingworth et al. found that changing an object in a scene often affected fixation duration on that object even though subjects failed to report that a change had occurred. This suggests that, at some level, the perceptual system detected the change even though the subjects were not explicitly aware of it. It is important to note that the implicit effects that have been observed have often been small, however. Furthermore, even though implicit measures of performance (e.g., eye movements) appear to be more sensitive to change than are explicit measures, it does not follow from this that scene representations contain a *great deal* more information than has been measured by explicit tests. Explicit tests may underestimate the amount of information in the scene representation, but there is no evidence that they underestimate it by much. This is still an open question.

REFERENCES

Aginsky, V., & Tarr, M. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition*, **7**, 147-162.

Averbach, E., & Coriell, E. (1961). Short-term memory in vision. *Bell System Technical Journal*, **40**, 309-328.

Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception*, **24**, 1075-1081.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, **15**, 719-722.

Bridgeman, B., & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society*, **21**, 285-286.

Bridgeman, B., van der Heijden, A., & Velichkovsky, B. (1994). A theory of visual stability across saccadic eye movements. *Behavioral & Brain Sciences*, **17**, 247-292.

Busey, T., & Loftus, G. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, **101**, 446-469.

Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception & Psychophysics*, **62**, 673-683.

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

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, **36**, 1827-1837.

Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition*, **7**, 324-344.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception* (Vancouver Studies in Cognitive Science, Vol. 5, pp. 89-110). New York: Oxford University Press.

Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, **7**, 43-64.

Henderson, J. M. (1993). Visual attention and saccadic eye movements. In G. d'Ydewalle & J. Van Rensbergen (Eds.), *Perception and cognition: Advances in eye-movement research* (pp. 37-50). Amsterdam: North-Holland.

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

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

Henderson, J. M., & Hollingworth, A. (1999b). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, **10**, 438-443.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & Psychophysics*, **45**, 196-208.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. Sensory processes and perception* (pp. 22.1-22.64). New York: Wiley.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, **57**, 787-795.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, **7**, 213-235.

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

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

Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences*, **1**, 217-222.

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

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 307-317.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, **5**, 94-100.

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125-155). Cambridge, MA: MIT Press.

Irwin, D. E., Brown, J. S., & Sun, J.-S. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General*, **117**, 276-287.

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention, and transsaccadic memory. *Visual Cognition*, **5**, 127-155.

Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics*, **34**, 49-57.

Irwin, D. E., Zacks, J. L., & Brown, J. S. (1990). Visual memory and the perception of a stable visual environment. *Perception & Psychophysics*, **47**, 35-46.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29-61). Orlando, FL: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.

Klein, R. (1980). Does oculomotor readiness mediate cognitive control of visual attention? In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 259-276). Hillsdale, NJ: Erlbaum.

Klein, R., & Pontefract, A. (1994). Does oculomotor readiness mediate cognitive control of visual attention? Revisited! In C. Umiltà & M. Moskovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 333-350). Cambridge, MA: MIT Press, Bradford Books.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, **35**, 1897-1916.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, **4**, 501-506.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, **81**, 899-917.

McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 563-581.

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

Mewhort, D. J. K., Campbell, A. J., Marchetti, F. M., & Campbell, J. I. D. (1981). Identification, localization, and "iconic memory": An evaluation of the bar-probe task. *Memory & Cognition*, **9**, 50-67.

Mondy, S., & Coltheart, V. (2000). Detection and identification of change in naturalistic scenes. *Visual Cognition*, **7**, 281-296.

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

O'Regan, J. K., & Levy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, **23**, 765-768.

Pan, K., & Eriksen, C. W. (1993). Attentional distribution in the visual field during *same–different* judgments as assessed by response competition. *Perception & Psychophysics*, **53**, 134-144.

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

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

Pollatsek, A., Rayner, K., & Henderson, J. (1990). The role of spatial location in the integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 199-210.

Pringle, H. L., Irwin, D. E., Kramer, A. F., & Atchley, P. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin & Review*, **8**, 89-95.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179-197.

Rayner, K., McConkie, G. W., & Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 529-544.

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

Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception & Psychophysics*, **34**, 39-48.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, **7**, 17-42.

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

Shepherd, M., Findlay, J., & Hockey, R. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology*, **38A**, 475-491.

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

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, **7**, 1-15.

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

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, **5**, 644-649.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, **74**(11, Whole No. 498).

Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Wallis, G., & Bulthoff, H. (2000). What's scene and not seen: Influences of movement and task upon what we see. *Visual Cognition*, **7**, 175-190.

Williams, P., & Simons, D. J. (2000). Detecting changes in novel 3D objects: Effects of change magnitude, spatiotemporal continuity, and stimulus familiarity. *Visual Cognition*, **7**, 297-322.

Zelinsky, G. J. (1999). Precuing target location in a variable set size "nonsearch" task: Dissociating search-based and interference-based explanations for set size effects. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 875-903.

Zelinsky, G. J. (2001). Eye movements during change detection: Implications for search constraints, memory limitations, and scanning strategies. *Perception & Psychophysics*, **63**, 209-225.

Zelinsky, G. J., & Loschky, L. (1998). Toward a realistic assessment of visual working memory. *Investigative Ophthalmology & Visual Science*, **39**, S224.

Zelinsky, G. J., Rao, R., Hayhoe, M., & Ballard, D. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, **8**, 448-453.

NOTES

1. The percentage of fixations summed across positions in Figure 6 does not add up to 100%, because some fixations did not fall on any object position. This happened relatively rarely for Fixations 3–15, ranging from 3% to 9% of all the fixations. It happened quite often for Fixation 2 (the first real fixation on the scene, since Fixation 1 was on the fixation point), accounting for 45% of all the fixations. Approximately 67% of these fell near an object position, whereas the remainder were scattered randomly in the central area of the scene; these may have been "center of mass" fixations like those previously observed by Zelinsky, Rao, Hayhoe, and Ballard (1997).

2. In order to determine how many objects would have to be stored in memory to produce an accuracy level of 80%, we first, to correct for guessing (Busey & Loftus, 1994), applied the formula $p = (x - g)/(1 - g)$, where x is the raw proportion correct, g is the guessing probability, and p is the corrected proportion correct (note that $g = .143$, or 1/7, because there were always seven objects in the array). We then multiplied p by the number of objects in the display in order to estimate the number of objects remembered (Sperling, 1960).

(Manuscript received March 9, 2001;
revision accepted for publication November 28, 2001.)