Journal of Experimental Psychology:
Learning, Memory, and Cognition
2006, Vol. 32, No. 1, 58–69

Copyright 2006 by the American Psychological Association
0278-7393/06/$12.00 DOI: 10.1037/0278-7393.32.1.58

Scene and Position Specificity in Visual Memory for Objects

Andrew Hollingworth
University of Iowa

This study investigated whether and how visual representations of individual objects are bound in memory to scene context. Participants viewed a series of naturalistic scenes, and memory for the visual form of a target object in each scene was examined in a 2-alternative forced-choice test, with the distractor object either a different object token or the target object rotated in depth. In Experiments 1 and 2, object memory performance was more accurate when the test object alternatives were displayed within the original scene than when they were displayed in isolation, demonstrating object-to-scene binding. Experiment 3 tested the hypothesis that episodic scene representations are formed through the binding of object representations to scene locations. Consistent with this hypothesis, memory performance was more accurate when the test alternatives were displayed within the scene at the same position originally occupied by the target than when they were displayed at a different position.

Keywords: visual memory, scene perception, context effects, object recognition

Humans spend most of their lives within complex visual environments, yet relatively little is known about how natural scenes are visually represented in the brain. One of the central issues in the study of scene perception and memory is how visual information from discrete objects and events is bound together to form an episodic representation of a particular environment. During scene viewing, the eyes and attention are oriented serially to individual objects of interest (for a review, see Henderson & Hollingworth, 1998). For example, while viewing an office scene a participant might direct attention and the eyes to a coffee cup, then to a pen, then to a notepad. In each case, focal attention supports the formation of a coherent perceptual object representation (Treisman, 1988) and the consolidation of that object information into visual memory (Averbach & Coriell, 1961; Hollingworth & Henderson, 2002; Irwin, 1992; Schmidt, Vogel, Woodman, & Luck, 2002; Sperling, 1960). Visual representations of objects are retained robustly in memory both during online scene viewing (Hollingworth, 2004b; Hollingworth & Henderson, 2002) and across significant delays as long as 24 hr (Hollingworth, 2004b, 2005b). To form representation of a scene as a whole, however, visual object representations must be episodically linked to the scene context.

Although there has been considerable research examining object perception and memory within scenes (see Henderson & Hollingworth, 1999a, 2003b; Hollingworth, in press; Simons & Levin, 1997, for reviews), current evidence is insufficient to answer the question of whether object representations are episodically linked in memory to scene context. This is quite an extraordinary knowledge gap in the field of visual cognition, especially considering that work on scene perception and memory often assumes the existence of scene-level representations (e.g., Hollingworth & Henderson, 2002). If object representations were not linked to the scene in which they appeared, then the study of scene perception and memory would in key respects become equivalent to the study of object memory and object recognition; memory for discrete objects in a scene would be no more than a collection of unrelated object representations.

A necessary condition for examining the binding of objects to scenes is that object representations can be reliably retained in memory. Such evidence comes from a series of studies conducted by Hollingworth and Henderson (Hollingworth, 2003a, 2004, 2005b; Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001). These studies examined object memory in scenes, both during the online viewing of a scene and after delays as long as 24 hr. The basic method (also used in the present study) was to present an image of a 3-D rendered scene containing a number of discrete objects. At some point during or after viewing, participants completed a change detection or forced-choice recognition task. In the change detection task, a single target object in the scene either remained the same or changed. When changed, the target was either replaced by a different object from the same basic-level category (token change) or rotated 90° in depth (orientation change). In the forced-choice test, two versions of the target were shown sequentially in the scene. One was the same as the original target object, and the other was either a different-token or different-orientation distractor. Both tasks required memory for the visual form of a single object in a scene.

Recent theories in the scene perception and change blindness literatures hold that performing these object memory tasks should be difficult if the target object is not the focus of attention when tested, as coherent visual object representations are proposed to disintegrate either immediately upon the withdrawal of attention from an object (Rensink, 2000; Rensink, O'Regan, & Clark, 1997) or as soon as an object is purged from visual short-term memory (VSTM; Becker & Pashler, 2002; Irwin & Andrews, 1996). Yet memory performance in the Hollingworth and Henderson studies

---

This research was supported by National Institute of Mental Health Grant R03 MH65456.

Correspondence concerning this article should be addressed to Andrew Hollingworth, Department of Psychology, University of Iowa, 11 Seashore Hall East, Iowa City, IA 52242-1407. E-mail: andrew-hollingworth@uiowa.edu

reflected robust retention of visual object representations, easily exceeding the capacity of VSTM (Hollingworth, 2004b, 2005b; Hollingworth & Henderson, 2002). Memory for the visual form of objects in scenes remained well above chance even when more than 400 objects, on average, intervened between target object fixation and test (Hollingworth, 2004b) and after a delay of 24 hr (Hollingworth, 2005b).

Given that visual object representations can be retained reliably from natural scenes, are they episodically organized within a scene-level representation? There is surprisingly little evidence bearing on this issue. A handful of studies have examined memory for objects in scenes as a function of the semantic consistency between the object and the scene in which it appeared (Brewer & Treyens, 1981; Friedman, 1979; Hollingworth & Henderson, 2000, 2003; Pedzek, Whetstone, Reynolds, Askari, & Dougherty, 1989). The typical result has been better memory for objects that are inconsistent with a scene (e.g., a coffeemaker in a farmyard) than those that are consistent (e.g., a chicken in a farmyard). However, inconsistent objects in these studies were clearly anomalous during initial viewing, and the memory advantage for inconsistent objects could therefore reflect differences in initial encoding rather than differences in the organization of memory (Friedman, 1979; Gordon, 2004; Henderson, Weeks, & Hollingworth, 1999; Pedzek et al., 1989). Mandler and colleagues examined memory for the spatial position and visual form of objects in scenes as a function of whether the scene was coherently organized (objects displayed in plausible spatial relationships) or not coherently organized (objects displayed in implausible spatial relationships) (Mandler & Johnson, 1976; Mandler & Parker, 1976; Mandler & Ritchey, 1977). They found that long-term memory for object position was improved by coherent scene organization but that memory for the visual form of objects was independent of scene organization. The work by Mandler and colleagues provides no evidence to suggest that visual object representations (coding visual form) are episodically bound to scene context. However, their stimuli were highly abstract, consisting of five to six individual line drawings of objects. Contextual information was minimal, in contrast to the more naturalistic images used in the present study.

Outside the scene perception and memory literature, there is at least some indication that complex visual stimuli (containing multiple discrete objects) are episodically organized. In the VSTM literature, Jiang, Olson, and Chun (2000) manipulated contextual information in a change detection task. A memory array of colored squares was presented for 400 ms, followed by a 900-ms interstimulus interval and a test array. In the test array, a single target square either changed color or stayed the same color. In addition, the contextual objects either remained the same or were changed. In one condition, the positions of the contextual objects were scrambled at test. In another, the contextual objects were deleted at test. Color change detection was impaired with both types of change in background context, suggesting that object color was not stored independently of memory for the other objects in the array.

In the face recognition literature, Tanaka and Farah (1993; see also Tanaka & Sengco, 1997) examined memory for features of faces (nose, eyes, and mouth) and houses (windows and door), manipulating the presence of the original context at test. Testing face features within the face context led to higher recognition performance compared with testing the features in isolation. But there was no such contextual advantage for the recognition of house features. Tanaka and Farah argued that face features are remembered as part of a holistic face representation, containing information from all features of the face and, further, that faces are unique in this respect. This contrast between faces and houses has supported the view that face recognition is functionally different from other forms of visual pattern recognition (Farah, 1995). However, Donnelly and Davidoff (1999) failed to replicate the Tanaka and Farah result with houses, finding a reliable whole-context advantage for the recognition of house features. With respect to the present question of episodic binding in scene representations, the results are ambiguous, with one study showing a contextual advantage for house features consistent with episodic binding of objects within a scene (Donnelly & Davidoff, 1999) and two showing no such advantage (Tanaka & Farah, 1993; Tanaka & Sengco, 1997). In addition, it is not clear from these studies whether the doors and windows of a house are to be considered parts of a single object (house) or discrete objects within a scene.

As is evident from this review, the fairly small body of relevant research provides no clear indication that visual object representations are bound in memory to the larger scene context. A primary goal of the present study was to provide such evidence. In Experiments 1 and 2, participants viewed 3-D-rendered images of complex, natural scenes for 20 s each. After viewing, memory for the visual form of a single object in the scene was tested in a two-alternative forced-choice test. In each experiment, participants had to discriminate the original target object from a different token or different orientation distractor, as illustrated in Figure 1. To examine the binding of objects to scenes in memory, at test the test object alternatives were presented either within the original scene context (background present condition) or in isolation (background absent condition), similar to the background presence manipulations of Jiang et al. (2000) and Tanaka and Farah (1993). If visual object representations are episodically bound to scene context, memory performance should be higher when that context is reinstantiated at test. This prediction follows from the encoding specificity framework of episodic memory (Tulving & Thompson, 1973). If memory for a particular object is linked to other scene elements in memory, then the presentation of those other elements at test should provide multiple cues for target retrieval. In the background absent condition, however, contextual cues would not be available at test, leading to impaired retrieval of target properties. If visual representations of individual objects are stored independently of the scene in which they appeared, as suggested by the work of Mandler and colleagues and Tanaka and Farah (1993), then the scene context cannot act as a retrieval cue at test, leading to the prediction of no difference in memory performance between background present and background absent conditions.

As a preview of the results of Experiments 1 and 2, both experiments found superior object recognition performance when the test alternatives were presented within the original scene context. Experiment 3 examined the manner by which object representations are bound to scene context, testing the hypothesis that object representations are bound to scene locations (Hollingworth & Henderson, 2002; Zelinksy & Loschky, in press). In this final experiment, the test object alternatives were always presented within the original scene context. In the same position condition, the test alternatives occupied the same position as had been occupied by the target object at study. In the different position condition, the test alternatives were presented at a different position

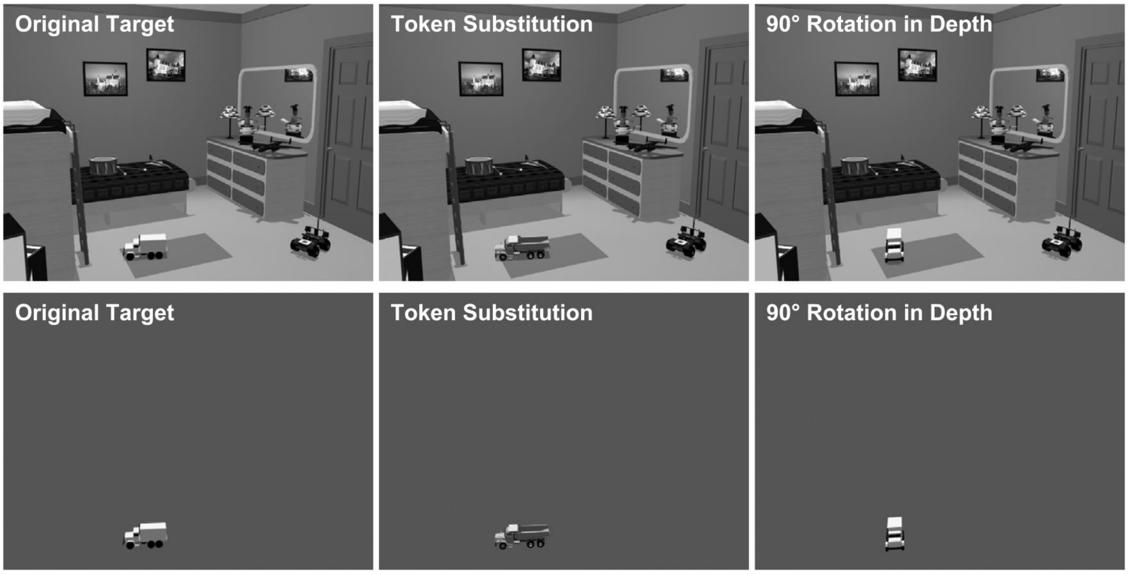

Figure 1. A sample scene illustrating object manipulations in the present study. The top row shows the background present stimuli. The bottom row shows the background absent stimuli. In the experiments, stimuli were presented in color.

within the scene. If object representations are bound to specific scene locations in memory, recognition performance should be more accurate when tested at the original object location than when tested at a different scene location (Kahneman, Treisman, & Gibbs, 1992). This predicted result was obtained in Experiment 3.

Experiment 1: Scene Specificity in Object Memory

Experiment 1 examined whether visual memory for objects during online scene viewing is episodically bound in memory to the scene context. The events in a trial are illustrated in Figure 2. Participants viewed a 3-D-rendered image of a scene, followed by a two-alternative, forced-choice recognition test. One of the options was the same as the target object initially viewed in the scene. The distractor object was either a different object from the same basic-level category (token discrimination condition) or the same object rotated 90° in depth (orientation discrimination condition). Token discrimination required object memory more specific than basic-level identity and typically required memory for visual form, as the majority of token pairs were identical at the subordinate level (see Figures 1 and 6). Rotation discrimination required memory for visual form, as the identity of the object did not change when rotated. The test options were displayed either within the scene context (background present condition) or in isolation (background absent condition). If visual object information is bound to the larger scene context, then object retrieval should be more efficient when the background is available at test, leading to higher recognition performance in the background present condition than in the background absent condition. However, if visual object representations are stored independently of scene context, then no effect of background presence at test should be observed.

Method

Participants. Twenty-four participants from the Yale University community completed the experiment. They either received course credit or were paid. All participants reported normal or corrected-to-normal vision.

Stimuli. Forty scene images were rendered from 3-D models of 40 different real-world environments. Scenes contained at least 7 discrete objects (conservatively defined as fully visible, movable objects), with the average number of objects in a scene approximately 11. In each model, a single target object was chosen. Target objects varied in size and location from scene to scene, with some targets in the foreground (see Figure 1) and some targets in the background (see Figure 6). To produce the target rotation images, the scene was rerendered after the target object had been rotated 90° in depth. To produce the token change images, the scene was rerendered after the target object had been replaced by another object token. Targets and token replacements were equated at the basic level, and the large majority were further equated at the subordinate level of categorization (e.g., the target and token replacement objects are both toy trucks in Figure 1 and are both sailboats in Figure 6). The objects for token changes were chosen to be approximately the same size as the initial target object. The background absent images were created by applying a transparent texture to all objects except the target object. Although not visible in the rendered image, background objects still reflected light and cast shadows within the model, ensuring that the target object appearance was identical to that in the standard scenes. The background was set to a uniform olive green (red/green/blue [RGB]: 90, 110, 20), chosen because none of the target objects contained this color, and thus they would not blend into the background.

Scene stimuli subtended 16.9° × 22.8° of visual angle at a viewing distance of 80 cm, maintained by a forehead rest. Target objects subtended 3.3° on average along the longest dimension in the picture plane. The mask was a patchwork of small colored shapes and was the same size as the scene stimuli. The onset dot was a neon green disk (RGB: 0, 255, 0), with a diameter of 1.2°. It appeared in a position within each scene unoccupied by any object that could plausibly be considered a target. The dot onset was a carryover from experiments seeking to ensure that the target was not

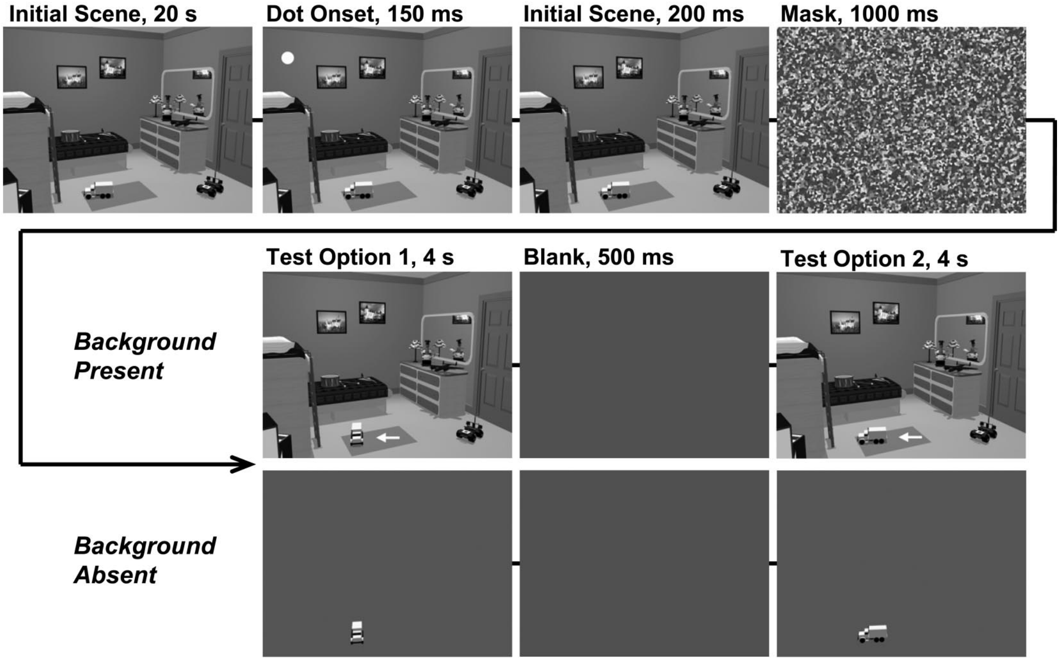

Figure 2. Sequence of events for sample trials in the background present and background absent conditions of Experiment 1. Trial events were identical for the background present and background absent conditions except for the test option displays. The figure shows orientation discrimination trials in which the correct target appears second in the test.

currently attended when it was tested (Hollingworth, 2003a), on the assumption that the dot would capture attention immediately before the test. Subsequent work has demonstrated that the presence or absence of the dot onset produces no observable influence on object memory (Hollingworth, 2004a). The postcue arrow was also neon green, subtended 2.2° in length, and pointed unambiguously to the target object in the test scene. The postcue was necessary in the background present condition to ensure that decision processes were limited to a single object, as in the background absent condition.

Apparatus. The stimuli were displayed at a resolution of 800 by 600 pixels by 24-bit color on a 17-in. video monitor with a refresh rate of 100 Hz. The initiation of image presentation was synchronized to the monitor's vertical refresh. Responses were collected using a serial button box. The presentation of stimuli and collection of responses was controlled by E-Prime software running on a Pentium IV–based computer. The room was dimly illuminated by a low-intensity light source.

Procedure. Participants were tested individually. Each participant was given a written description of the experiment along with a set of instructions. Participants were informed that they would view a series of scene images. After viewing each scene, they would have to decide between two object options, one of which was the same as an object that had appeared in the original scene. The nature of the possible distractors was described, as was the background presence manipulation.

Participants pressed a pacing button to initiate each trial. Then, a white fixation cross on a gray field was displayed for 1,000 ms. This was followed by the initial scene presentation for 20 s, dot onset within the scene for 150 ms, initial scene again for 200 ms, pattern mask for 1,000 ms, Test Option 1 for 4 s, gray field for 500 ms, Test Option 2 for 4 s, and finally a screen asking participants to respond whether Option 1 or 2 was the same as the original target object. Participants were instructed to respond as accurately as possible; response speed was not mentioned. They either pressed a button on the serial box labeled *first* or a button labeled *second*. Button response terminated the trial.

Participants first completed a practice session of four trials, one in each of the conditions created by a 2 (background present, background absent) × 2 (orientation discrimination, token discrimination) factorial combination. The scene items used for the practice trials were not used in the experimental session. In the experimental session, participants viewed each of the 40 scene items once, five scenes in each of the eight conditions created by the full 2 (background present, background absent) × 2 (orientation discrimination, token discrimination) × 2 (correct option first, second) factorial design. Across participants, condition–item assignments were counterbalanced by Latin square so that each scene item appeared in each condition an equal number of times. Trial order was determined randomly. The entire session lasted approximately 45 min.

Results and Discussion

Mean percent correct performance on the two-alternative forced-choice task is displayed in Figure 3 as a function of background presence and discrimination type (token and orientation). In this and in subsequent analyses, two analyses of variance (ANOVAs) were conducted, one treating participant as a random effect (F_1) and one treating scene item as a random effect (F_2). Reported means were derived from analyses treating participant as a random effect. There was a reliable main effect of background presence, with higher performance in the background present condition (88.3%) than in the background absent condition (79.8%), $F_1(1, 23) = 14.52$, $p < .001$; $F_2(1, 39) = 13.45$, $p < .001$. There was also a reliable main effect of discrimination condition by subjects and a marginal effect by items, with higher performance for token discrimination (87.3%) than for orientation discrimination (80.8%), $F_1(1, 23) = 9.50$, $p < .01$; $F_2(1, 39) = 3.95$, $p = .05$. These two factors did not interact (Fs < 1).

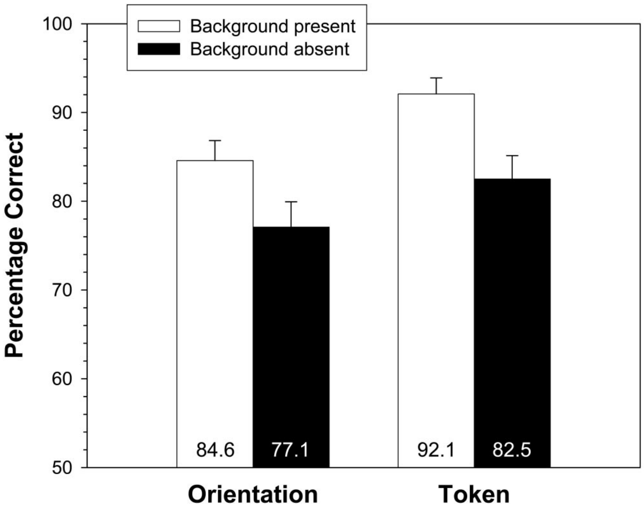

Figure 3. Experiment 1: Mean percentage correct as a function of background presence and discrimination condition (orientation and token). Error bars are standard errors of the means.

In addition to these principal effects, there was an effect of option order in the forced-choice test, with higher performance when the correct target was the first option (90.8%) than when the correct target was the second option (77.3%), $F_1(1, 23) = 32.40$, $p < .001$; $F_2(1, 39) = 28.63$, $p < .001$. Participants were biased to respond *first* in the two-alternative test. In addition, option order interacted with background presence, $F_1(1, 23) = 9.37$, $p < .01$; $F_2(1, 39) = 11.59$, $p < .005$. The advantage for first-option performance was larger in the background absent condition (21.3%) than in the background present condition (5.8%). Presumably, poorer recognition in the background absent condition yielded more trials that could be influenced by the first-option bias. The first-option bias does not influence interpretation of the background present advantage, as option order was counterbalanced across the main conditions of interest.

Discrimination performance on the object recognition test was reliably more accurate when the test objects were presented within the original scene context versus in isolation. This scene specificity effect demonstrates that memory for the visual form of individual objects in natural scenes is stored as part of a more comprehensive scene representation.

Experiment 2: Scene Specificity in Visual Long-Term Memory (VLTM)

Experiment 1 tested object memory immediately after viewing each scene, examining memory formed during online scene viewing. Hollingworth (2004b; see also Zelinksy & Loschky, in press) demonstrated that online scene memory is composed of both a VSTM component (for approximately the last two objects fixated) and a VLTM component (for objects attended earlier). In the Experiment 1 method, contextual effects on object memory could have depended on VSTM if the target object happened to be attended late in viewing or VLTM if the target object was attended earlier in viewing, raising the question of whether episodic binding is a property of VSTM, VLTM, or both. As reviewed above, Jiang et al. (2000) have already demonstrated contextual sensitivity for object memory in a VSTM paradigm. To assess episodic binding in VLTM, Experiment 2 replicated Experiment 1 but delayed each object test one trial after scene viewing, so that the viewing of scene *n* was followed by the test for scene *n* – 1, and so on (Hollingworth, 2005b). Given severely limited capacity in VSTM, the one-trial delay ensured that scene representation must have been entirely dependent on VLTM.

Method

Participants. Twenty-four new participants (19 from the Yale University community and 5 from the University of Iowa community) completed the experiment. They either received course credit or were paid. All participants reported normal or corrected-to-normal vision.

Stimuli and apparatus. The stimuli and apparatus were the same as in Experiment 1.

Procedure. Each trial consisted of the viewing of scene *n* followed by the forced-choice test for scene *n* – 1. Participants pressed a pacing button to initiate each trial. Then a white fixation cross on a gray field was displayed for 1,000 ms, followed by the initial scene presentation for 20 s. The scene was followed by a gray screen with the message "Prepare for previous scene test" displayed for 3 s. This was followed by scene *n* – 1 Test Option 1 for 4 s, gray field for 500 ms, scene *n* – 1 Test Option 2 for 4 s, and response screen. Participants responded as in Experiment 1. The dot onset, offset, and scene mask used in Experiment 1 were eliminated from Experiment 2. Intervening between the viewing and test of a particular scene was the test of the previous scene item and the viewing of the subsequent scene item. Given current VSTM capacity estimates of no more than three or four objects (Pashler, 1988; Vogel, Woodman, & Luck, 2001) and the fact that each scene item contained many more than four individual objects, these intervening events ensured that the target object was no longer being maintained in VSTM when it was tested; performance must have depended on VLTM. The mean temporal delay between the end of scene viewing and the start of the forced-choice test of that item was 38.5 s.

Participants first completed four practice trials, one in each of the conditions created by a 2 (background present, background absent) × 2 (orientation discrimination, token discrimination) factorial combination. The scene items used for the practice trials were not used in the experimental session. The experimental trials followed the practice trials without interruption: Viewing of the first experimental item was followed by test of the last practice item. Participants viewed all 40 scene items once, five items in each of the eight conditions created by the full 2 (background present, background absent) × 2 (orientation discrimination, token discrimination) × 2 (correct target option first, second) factorial design. Across participants, each scene item appeared in each condition an equal number of times. The last trial presented a dummy item for viewing so that the test of the final experimental item could be delayed one trial. The dummy item was not tested. Otherwise, trial order was determined randomly. The entire session lasted approximately 45 min.

Results and Discussion

Mean percent correct performance on the two-alternative forced-choice task is displayed in Figure 4 as a function of background presence and discrimination type (token and orientation). There was a reliable main effect of background presence, with higher performance in the background present condition (89.4%) than in the background absent condition (84.2%), $F_1(1, 23) = 6.18$, $p < .05$; $F_2(1, 39) = 4.62$, $p < .05$. There was a trend toward an effect of discrimination type, with higher performance for token discrimination (88.8%) than for orientation discrimination (84.8%), $F_1(1, 23) = 2.34$, $p = .14$; $F_2(1, 39) = 2.16$, $p = .15$.

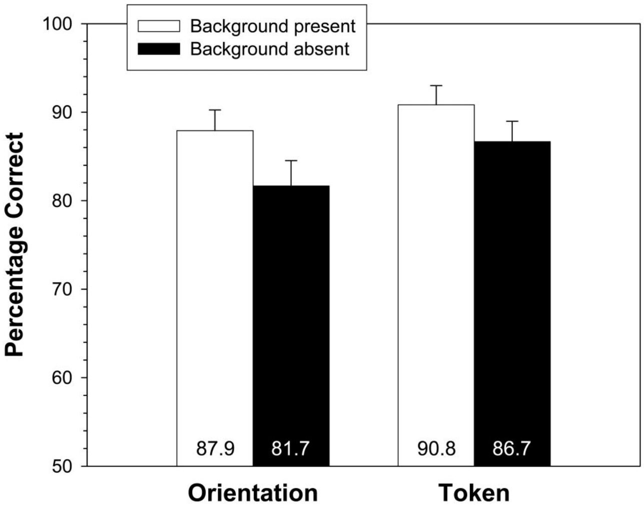

Figure 4. Experiment 2: Mean percentage correct as a function of background presence and discrimination condition (orientation and token). Error bars are standard errors of the means.

These two factors did not interact (*F*s < 1). As in Experiment 1, there was also an effect of option order in the forced-choice test, with higher performance when the correct target was the first option (90.8%) than when the correct target was the second option (82.7%), $F_1(1, 23) = 9.87, p < .005$; $F_2(1, 39) = 10.18, p < .005$. The interaction between option order and background presence was not reliable, $F_1(1, 23) = 1.40, p = .25$; $F_2(1, 39) = 2.27, p = .14$, although the numerical trend was in the same direction as in Experiment 1.

In Experiment 2, the scene specificity effect was observed under conditions that required VLTM, demonstrating that the long-term visual representation of an object is linked to memory for the scene context in which the object was originally viewed. A notable feature of the Experiment 2 data is that overall performance was no lower than that in Experiment 1, despite the one-trial delay. Similarly, Hollingworth (2005b) found, in a within-subject design, that object memory performance was unreduced by the introduction of a one-trial delay from the level of performance when the test was administered immediately after viewing of the scene. Visual object memory is highly robust.

Discussion of Experiments 1 and 2

The results of Experiments 1 and 2 demonstrate that memory for object form is stored as part of a larger scene representation. These results raise the question of which scene properties serve to facilitate object memory. To this point, we have been considering the scene context as any information in the scene not directly part of the target object. Scene information might include large-scale geometric structures (such as the walls, floor, and ceiling of a room), other discrete objects in the scene, or the local contextual information where the target object contours intersect with the scene (such as the local relationship between the toy truck and the rug in Figure 1).

Considering the last possibility first, it is unlikely that local contextual information was driving the background present advantage in Experiments 1 and 2. Relevant evidence comes from an experiment conducted by Hollingworth (2003b). Similar to Experiments 1 and 2, on each trial participants viewed an image of real-world scene for 20 s, followed by a 1,000-ms scene mask, followed by a single test image. The task was a left–right mirror-reflection change detection: The target object in the test was either the same as in the studied scene or mirror reflected, and participants responded "same" or "changed." In one condition of this experiment, the test object was presented within the scene background, and in another condition it was presented in isolation, similar to the background presence manipulation in Experiments 1 and 2. However, in both conditions, the test object was presented within a blank disk so that local contextual information was eliminated. Figure 5 shows a sample scene. If the background present advantage in the present study was driven by memory for local contextual information, that advantage should have been reduced or eliminated under these conditions. Yet a reliable background present advantage was still observed (background present, 82% correct; background absent, 74% correct) and was of similar magnitude to the background present advantage found in Experiments 1 and 2 of this study. Thus, local contextual information is not critical for obtaining the background present advantage.

The results of Hollingworth (2003b) suggest that the background present advantage is likely driven by memory for more global scene information, such as memory for large-scale scene structures and other discrete objects. Earlier work using highly simplified scenes has dichotomized scene information into large-scale geometric structures (e.g., horizon lines, walls of a room) and

Figure 5. Sample stimuli from Hollingworth (2003b). The left panel shows the studied scene image. The middle panel shows the test object in the background present condition. The right panel shows the test object in the background absent condition. In the experiment, stimuli were presented in color.

individual objects (Mandler & Ritchey, 1977). But for naturalistic scene stimuli, such as the scenes used in the present study, there is no clear division between these classes of scene information. A desk serves as a large-scale surface for objects, but it is also a discrete object itself. Similarly, a refrigerator is a discrete, bounded object, but it is also a large, fixed (i.e., generally nonmovable) element within a kitchen that could easily be considered part of the large-scale scene structure. In the present stimulus set, all scenes contained elements that could be considered both discrete objects and large-scale contextual elements (e.g., the dresser and bed in the Figure 1 scene; the train and benches in the Figure 5 scene). Future research with scenes designed to isolate large-scale scene structures and discrete objects will be necessary to examine the contributions of these two sources of scene contextual information to object memory.

Experiment 3: Position Specificity

Experiments 1 and 2 established the basic scene specificity effect. Experiment 3 investigated the *means* by which object representations are structured within a larger scene representation. Hollingworth and Henderson (2002) proposed a possible mechanism for episodic binding within scenes. In this view, individual object representations are bound to positions within a spatial representation of the scene. Specifically, as the eyes and attention are oriented within a scene, higher level visual representations are formed for attended objects and are activated initially in VSTM. The higher level object representation is bound to a position within a spatial representation of the scene (Henderson, 1994; Hollingworth, 2005a; Irwin, 1992; Irwin & Zelinsky, 2002; Kahneman et al., 1992; Zelinsky & Loschky, in press), which is consolidated into VLTM. During scene viewing, VSTM representations are replaced as attention and the eyes select subsequent objects. However, the VLTM representation is retained robustly and accumulates with visual representations from other previously attended objects. In addition, the retrieval of object information is mediated by spatial position: Attending back to the original location of an object facilitates retrieval of the object information bound to that location (Kahneman et al., 1992; Sacks & Hollingworth, 2005).

Experiment 3 tested this spatial binding hypothesis of scene contextual structure. The method was similar to that in Experiment 1 (immediate two-alternative object test after scene viewing), except that the principal manipulation was the position of the test object alternatives in the scene rather than the presence of the background scene at test. After viewing each scene, the two test objects were presented within the scene either at the same position as had been occupied by the target object at study or at a different position on the other side of the scene (i.e., the left–right mirror-reflected position). Figure 6 shows the stimulus manipulations for a sample scene item. If object representations are bound to scene spatial positions, then memory performance should be more accurate when the test alternatives are presented at the original object location (Kahneman et al., 1992).

In Experiment 3, the test object alternatives were always presented within a blank, olive-green disk surrounded by neon-green ring (see Figure 6). The neon green ring simply provided a salient target postcue. The olive-green disk ensured that position effects were not confounded with differences in the intersection of local contours between object and scene. If the test objects had been integrated within the scene, then the intersection between the scene and object contours would have changed in the different position condition but would have remained the same in the same position condition. Eliminating local contextual information in both conditions prevented this confound.

Method

Participants. Sixteen new participants from the University of Iowa community completed the experiment. They received course credit or pay for their participation. All participants reported normal or corrected-to-normal vision.

Stimuli and apparatus. The set of scene items was expanded from the 40 used in Experiments 1 and 2 to 56. This change reflected general expansion of the set of 3-D scenes and was not related to any experimental manipulation. In Experiment 3, the test objects were presented within an olive-green disk surround by a neon-green ring. The disk was large enough to enclose all versions of the target object (initial, token substitution, and rotation). The ring served to cue the relevant object at test, eliminating the need for a postcue arrow. The apparatus was the same as in Experiment 1.

Procedure. In Experiment 3, a four-digit verbal working memory load and articulatory suppression were added to the paradigm. Experiments 1 and 2 did not include such measures for suppression of verbal encoding, because previous work has demonstrated that verbal encoding plays little or no role in paradigms examining object memory in scenes (Hollingworth, 2003a, 2005b) or even in paradigms examining memory for easily nameable color patches (Vogel et al., 2001). A verbal working memory load and articulatory suppression produce a dramatic impairment in verbal encoding and memory (Vogel et al., 2001) but produce minimal effects on memory for objects and colors. The inclusion of a verbal working memory load and articulatory suppression in Experiment 3 was simply a conservative measure to ensure that contextual effects would still be observed when the opportunity for verbal encoding was minimized. At the beginning of each trial, the initial screen instructing participants *press a button to start the next trial* also contained four randomly chosen digits. Participants began repeating the four digits aloud before initiating the trial and continued to repeat the digits until the object test. Participants were instructed to repeat the digits without interruption or pause, at a rate of approximately two digits per second. The experimenter monitored digit repetition to ensure that participants complied.

Otherwise, the sequence of events in a trial was similar to that in Experiment 1. Participants pressed a pacing button to initiate each trial. Then, a white fixation cross on a gray field was displayed for 1,000 ms, followed by the initial scene presentation for 20 s, pattern mask for 500 ms, Test Option 1 for 4 s, blank (gray) interstimulus interval for 500 ms, Test Option 2 for 4 s, and finally a screen instructing participants to respond to indicate whether the first or second option was the same as the initial target. Button response terminated the trial.

In each two-alternative test, one object was the same as the original object presented in the scene. The other was either a different token distractor (token discrimination condition) or a different orientation distractor (orientation discrimination condition), as in Experiments 1 and 2. In the same position condition, the two test options were displayed in the same position as had been occupied by the target in the initial scene. In the different position condition, the two test options were displayed at the corresponding position on the other side of the screen (i.e., the left–right mirror-reflected position). Vertical position in the scene and distance from scene center were held constant. For approximately half of the scene items (27 of 56), it was possible to construct the scene so that the different position was a plausible location for the target object, as illustrated in Figure 6. For the remainder of the items, the different position was implausible (e.g., the object did not have a supporting surface). The results, reported below, did not differ for the two sets of items.

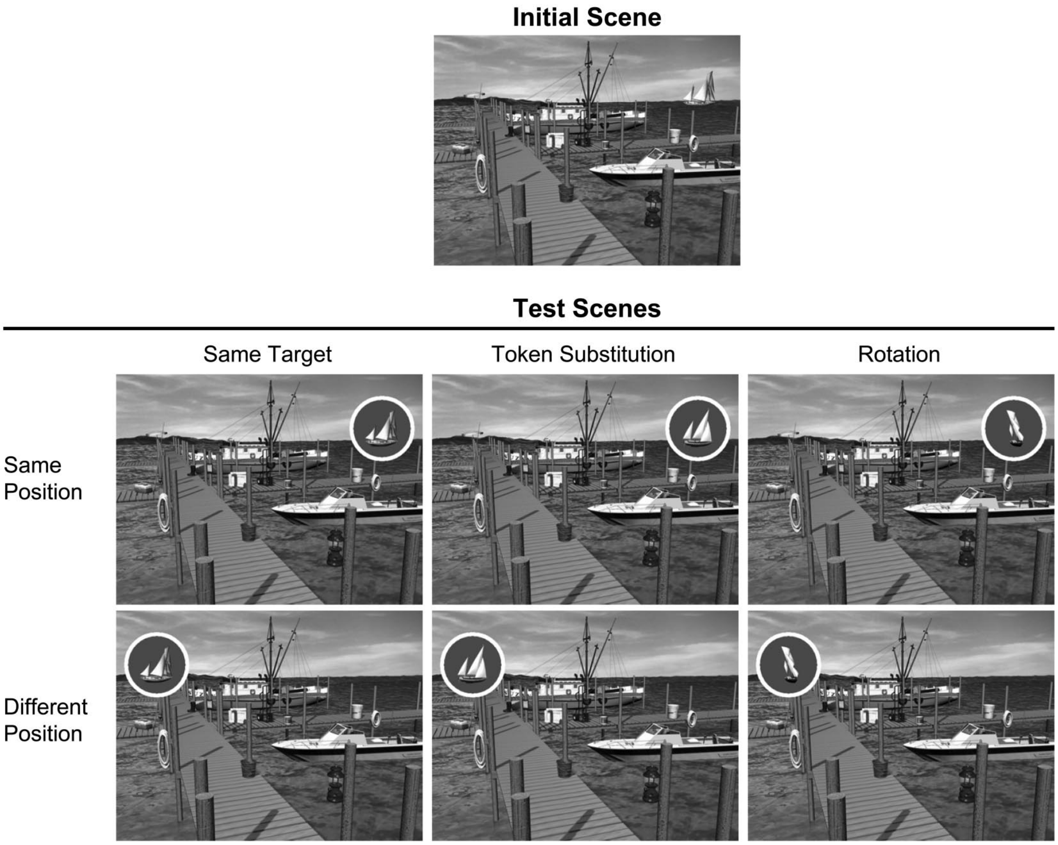

Figure 6. Sample scene stimuli illustrating the test object and position manipulations in Experiment 3. The initial, studied scene is displayed at the top of the figure. Test objects were presented either at the original target object location (same position) or at a different location equally distant from scene center (different position). In the two-alternative forced-choice test, one object option was the same as the original target (same target), and the other option was either a different token (token substitution) or the same object rotated 90° in depth (rotation). In the experiment, stimuli were presented in color.

Participants were instructed that the relevant object would always appear in the neon-green ring. In addition, they were told that the test options would appear either in the same position within the scene that the target object had initially occupied or in a different position. They were instructed that regardless of test object position, they should decide which of the two object options was the same as the object displayed initially in the scene.

Participants first completed eight practice trials, one in each of the conditions created by the 2 (same position, different position) × 2 (orientation discrimination, token discrimination) × 2 (correct target option first, second) factorial design. The scene items used for the practice trials were not used in the experimental session. In the experimental session, participants viewed each of the 56 scene items once, seven in each of the eight conditions. Across participants, condition–item assignments were counterbalanced by Latin square so that each scene item appeared in each condition an equal number of times. Trial order was determined randomly. The entire session lasted approximately 55 min.

Results and Discussion

Mean percent correct performance on the two-alternative forced-choice task is displayed in Figure 7 as a function of test object position and discrimination type (token and orientation). There was a reliable main effect of test object position, with higher

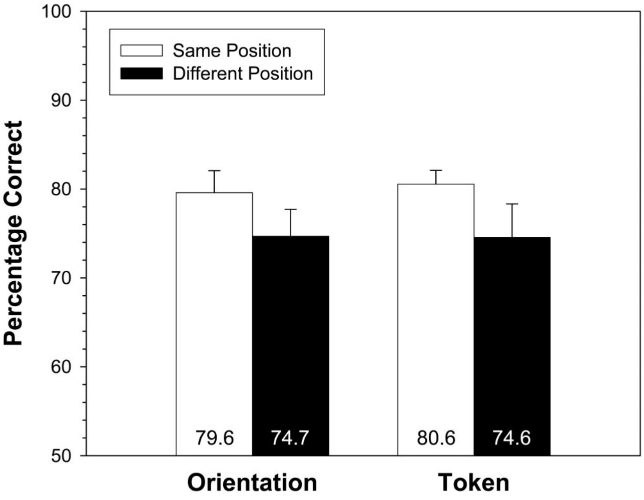

Figure 7. Experiment 3: Mean percentage correct as a function of test object position and discrimination condition (orientation and token). Error bars are standard errors of the means.

performance in the same position condition (80.1%) than in the different position condition (74.7%), $F_1(1, 15) = 4.79$, $p < .05$; $F_2(1, 55) = 4.13$, $p < .05$. Discrimination type did not produce a reliable main effect (*F*s < 1). There was no interaction between test object position and discrimination type (*F*s < 1). In addition to the effect of position, there was also an effect of option order in the forced-choice test, with higher performance when the correct target was the first option (87.9%) than when the correct target was the second option (66.9%), $F_1(1, 15) = 36.76$, $p < .001$; $F_2(1, 55) = 57.01$ ($p < .001$). Participants were biased to select the first option, but again, this effect does not influence interpretation of the position or discrimination type effects, as option order was counterbalanced across those conditions. The interaction between option order and background presence was not reliable, $F_1(1, 15) = 2.25$, $p = .15$; $F_2(1, 55) = 1.42$, $p = .24$.

In Experiment 3, memory accuracy was higher when position consistency was maintained from study to test. The test object options were always presented within the original scene context. Thus, a general benefit for reinstantiating the original context could not account for the same position advantage. The same position advantage indicates that visual object representations are bound to scene locations in memory, as claimed by Hollingworth and Henderson (2002). Object-position binding is therefore a plausible candidate mechanism for the construction of episodic scene representations (Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; Zelinksy & Loschky, in press).

General Discussion

The present study asked a simple but important and heretofore unresolved question: Are visual object representations bound in memory to the scene context in which they were viewed? In Experiments 1 and 2, participants more accurately recognized object exemplars when the object was displayed at test within the original scene context versus in isolation. This is the first study to provide unequivocal evidence that objects in scenes are episodically bound to scene context in memory, forming a scene-level representation. Experiment 3 then tested the hypothesis that episodic scene representations are constructed by the binding of object representations to specific scene locations (Hollingworth & Henderson, 2002; Zelinsky & Loschky, in press). Supporting this spatial binding hypothesis, participants more accurately recognized object exemplars when the test alternatives were presented at the target's original location in the scene than when they were presented at a different scene location.

The idea that spatial position within a scene plays an important role in structuring object memory is supported by evidence from at least three sources. First, VSTM studies have found evidence of object-position binding (Henderson, 1994; Henderson & Anes, 1994; Irwin, 1992; Irwin & Gordon, 1998; Irwin & Zelinsky, 2002; Kahneman et al., 1992; Noles, Scholl, & Mitroff, 2005) and of contextual structure based on global spatial configuration (Jiang et al., 2000). Second, Hollingworth and Henderson (2002) observed that when the deletion of an object was not initially detected during online scene viewing, participants often detected the change later in viewing, but only after they had fixated the location where the object had originally appeared, suggesting that object memory was bound to spatial position and that attending to the original position facilitated object retrieval. Finally, three studies have found direct evidence that participants can successfully bind local object information to specific scene locations (Hollingworth, 2005a; Irwin & Zelinsky, 2002; Zelinsky & Loschky, in press). Visual object representations are likely maintained in inferotemporal brain regions (Logothetis & Pauls, 1995), and spatial scene representations, in medial temporal regions (Epstein & Kanwisher, 1998). Binding of objects to scene locations could be produced by simple associative links between scene-specific hippocampal or parahippocampal place codes and inferotemporal object representations, similar to models of landmark-position binding in the rodent navigation literature (Gallistel, 1990; McNaughton et al., 1996; Redish & Touretzky, 1997).

The spatial binding hypothesis can account for the basic background present advantage in Experiments 1 and 2 if we assume that object positions are defined relative to the particular scene spatial context in which the object was viewed (Hollingworth, 2003b; see Klein & MacInnes, 1999, for evidence that position memory during search within a scene is defined relative to the particular scene spatial context). When the scene background was presented at test, the spatial context serving to define object position was reinstantiated, allowing participants to attend the location in the scene where the target appeared. Attending to object location relative to the scene facilitated retrieval of the object representation associated with that scene location (Sacks & Hollingworth, 2005). When the scene context was not presented at test, participants could not efficiently reinstantiate the scene spatial context that served to define object location, participants could not attend to the scene-relative location where the target had originally appeared, and object retrieval was impaired.

The conclusion that scene spatial context supports episodic binding of objects to locations requires a pair of qualifications. First, spatial binding is not the only possible binding mechanism for the construction of episodic scene representations; it is merely a plausible one. For example, representations of objects in the same scene could be associated directly with each other rather than through scene spatial position. Although object-to-object association could certainly account for the basic background present advantage in Experiments 1 and 2, object-to-object association could not easily account for the same position advantage in Experiment 3, as the same set of contextual objects was visible in the same and different position conditions. Although Experiment 3 does not eliminate the possibility of object-to-object association, it does demonstrate that at least one mechanism of binding in scene memory is inherently spatial.

Second, although the present study found episodic structure in memory for objects in scenes, this cannot be taken as evidence that such binding is unique to visual scenes. Faces (Tanaka & Farah, 1993), individual objects (Donnelly & Davidoff, 1999; Gauthier & Tarr, 1997), and arrays of simple objects (Jiang et al., 2000) have shown similar contextual effects. In addition, the present data do not speak to the possibility that stimuli from other perceptual and cognitive systems (e.g., auditory information) could also be bound within a multimodal representation of an environment. Further research will be required to determine whether object-to-scene binding depends on scene-specific mechanisms or on domain-general binding mechanisms.

The present results demonstrated contextual effects in the exemplar-level recognition of objects in scenes. The token manipulation in Experiments 1 and 2 probed exemplar-level object recognition. The orientation manipulation probed subexemplar recognition of visual form. This raises the question of why contextual facilitation is observed in the present experiments but not in paradigms examining context effects on the perceptual categorization of objects at the entry level (Hollingworth & Henderson, 1998, 1999). The critical difference likely lies in the nature of object recognition in the two types of paradigm. In studies examining effects of scene consistency on the perceptual categorization of objects (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989; Davenport & Potter, 2004; Hollingworth & Henderson, 1998, 1999; Palmer, 1975), scene stimuli are presented very briefly, and the task is usually to detect the presence of a particular *type* of object at the basic, or entry, level. Under structural description approaches to object recognition, entry-level categorization depends on highly abstracted object models (Biederman, 1987). Contextual effects would not be expected, because stored category models simply do not contain contextual information. Under image-based approaches, entry-level categorization is proposed to depend on the combined activation of large numbers of exemplar representations (Perrett, Oram, & Ashbridge, 1998; Tarr & Gauthier, 1998). Again, contextual information should play little or no role in entry-level categorization, because even if we assume that object image representations retain contextual information, contextual features would be lost as activation from multiple exemplars is pooled. It is possible that semantic-level knowledge (e.g., that toasters are likely to appear in kitchens but not in bathrooms) could directly feed back into object recognition processes (Biederman et al., 1982) or that scene recognition could prime object category models (Friedman, 1979; Palmer, 1977), but neither class of object recognition theory has proposed such a mechanism, and the data suggest that when detection sensitivity is isolated from participant bias to report consistent objects, semantically consistent objects are detected no more accurately than inconsistent objects (Hollingworth & Henderson, 1998, 1999). The contextual independence of entry-level object identification supports the human ability to identify objects across different scene contexts (a bear should be identified as a bear whether it appears in the woods or on Main Street; Tarr & Vuong, 2002) and to do so just as efficiently for objects unexpected within a scene as for objects expected within a scene.

In contrast, object exemplar recognition, by its very nature, depends on memory for an individual object. In the present experiments, the test object alternatives were displayed for 4 s each. There were minimal demands placed on perceptual processing of the test objects, and the present results therefore do not address top-down effects on perception. Instead, contextual differences were likely attributable to differences in memory retrieval. Under image-based theories, factors that influence the efficiency or success of retrieving the appropriate exemplar image will influence recognition performance. Retrieval of stored exemplar representations has not typically been considered a limiting factor in exemplar recognition, but it certainly could be when attempting to retrieve a single object exemplar representation (e.g., to decide whether an object has changed token or orientation) from among many thousands of such representations stored in memory. Although exemplar recognition was significantly worse in the background absent and different position conditions, it was still fairly accurate. And indeed, an exemplar recognition mechanism that failed to identify individual objects in new contexts or in new locations would be suboptimal. Exemplar recognition appears to balance contextual specificity, as individual objects are often consistently found in specific locations in a scene, with the ability to generalize recognition to new scenes and new locations. Theories of exemplar-level object recognition, which have typically addressed object recognition in isolation, will need to account for effects of contextual specificity.

The present results also have implications for theorizing in the face recognition literature. The fact that faces appeared to be unique in showing contextual sensitivity for the recognition of local features was taken as evidence that faces are represented in a manner different from other visual stimuli, holistically rather than by part decomposition (Tanaka & Farah, 1993). Subsequent work, however, has demonstrated that recognition of house features also shows contextual sensitivity (Donnelly & Davidoff, 1999), as does the recognition of object parts under conditions of observer expertise (Gauthier & Tarr, 1997). The present results demonstrate that recognition of local objects in scenes shows contextual sensitivity, providing further evidence that faces are not unique in this respect. But in any case, contextual sensitivity cannot be taken as strong evidence of holistic representation. Contextual sensitivity could indeed be generated by holistic representation, but it could also be generated if discrete parts or objects, parsed from the larger stimulus, are bound together within a higher level episodic representation of the object, face, or scene.

Finally, the results from Experiment 2 extend our understanding of contextual structure in visual memory systems. Jiang et al. (2000) found strong evidence of contextual structure in VSTM. Experiment 2 of this study demonstrated that contextual sensitivity is also a property of VLTM. The relationship between VSTM and VLTM is not yet well understood. Evidence from Hollingworth (2004b) suggests that VSTM and VLTM are closely integrated to support the online visual representation of natural scenes. However, Olson and Jiang (2004) found that the existence of a VLTM representation of an array does not improve VSTM representation of the items in the array, suggesting a degree of independence. Regardless of the precise relationship between VSTM and VLTM, both memory systems appear to maintain visual representations of similar format. Visual representations maintained over the short term are sensitive to object token (Henderson & Hollingworth, 2003a; Henderson & Siefert, 2001; Pollatsek, Rayner, & Collins, 1984), orientation (Henderson & Hollingworth, 1999b, 2003a; Henderson & Siefert, 1999, 2001; Tarr, Bülthoff, Zabinski, & Blanz, 1997; Vogel et al., 2001), and object part structure (Carlson-Radvansky, 1999; Carlson-Radvansky & Irwin, 1995) but are insensitive to absolute size (Pollatsek et al., 1984) and precise object contours (Henderson, 1997; Henderson & Hollingworth, 2003c). Similarly, visual representations retained over the long term are sensitive to object token (Biederman & Cooper, 1991), orientation (Tarr, 1995; Tarr et al., 1997), and object part structure (Palmer, 1977) but are insensitive to absolute size (Biederman & Cooper, 1992) and precise object contours (Biederman & Cooper, 1991). The present results demonstrate a further commonality between VSTM and VLTM object representations: Both are stored as part of a larger contextual representation of the scene.

References

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

Becker, M. W., & Pashler, H. (2002). Volatile visual representations: Failing to detect changes in recently processed information. *Psychonomic Bulletin & Review, 9,* 744–750.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23,* 393–419.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 121–133.

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

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

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 207–230.

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

Carlson-Radvansky, L. A., & Irwin, D. E. (1995). Memory for structural information across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1441–1458.

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

Donnelly, N., & Davidoff, J. (1999). The mental representations of faces and houses: Issues concerning parts and wholes. *Visual Cognition, 6,* 319–343.

Epstein, R., & Kanwisher, N. (1998, April 9). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Farah, M. J. (1995). Dissociable systems for recognition: A cognitive neuropsychology approach. In S. M. Kosslyn & D. N. Osherson (Eds.), *An invitation to cognitive science: Vol. 2. Visual cognition* (pp. 101–119). Cambridge, MA: MIT Press.

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

Gallistel, C. R. (1990). *The organization of learning.* Cambridge, MA: MIT Press.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37,* 1673–1682.

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception & Performance, 30,* 760–777.

Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General, 123,* 410–426.

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

Henderson, J. M., & Anes, M. D. (1994). Effects of object-file review and type priming on visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 826–839.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–283). Oxford, England: Elsevier.

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., & Hollingworth, A. (2003a). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics, 65,* 58–71.

Henderson, J. M., & Hollingworth, A. (2003b). Eye movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 356–383). New York: Oxford University Press.

Henderson, J. M., & Hollingworth, A. (2003c). Global transsaccadic change blindness during scene perception. *Psychological Science, 14,* 493–497.

Henderson, J. M., & Siefert, A. B. (1999). The influence of enantiomorphic transformation on transsaccadic object integration. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 243–255.

Henderson, J. M., & Siefert, A. B. C. (2001). Types and tokens in transsaccadic object identification: Effects of spatial position and left–right orientation. *Psychonomic Bulletin & Review, 8,* 753–760.

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

Hollingworth, A. (2003a). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 388–403.

Hollingworth, A. (2003b, November) *The structure of scene representations.* Talk presented at the Annual Meeting of the Psychonomic Society, Vancouver, British Columbia, Canada.

Hollingworth, A. (2004a). [Change detection with and without onset cue.] Unpublished raw data.

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

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

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

Hollingworth, A. (in press). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition.*

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127,* 398–415.

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

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 and Performance, 28,* 113–136.

Hollingworth, A., & Henderson, J. M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. *Memory & Cognition, 31,* 930–940.

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.

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

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

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics, 64,* 882–895.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 683–702.

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. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science, 10,* 346–352.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex, 3,* 270–288.

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.

Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 38–48.

Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 386–396.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., et al. (1996). Deciphering the hippocampal polyglot: The hippocampus as a path integration system. *Journal of Experimental Biology, 199,* 173–185.

Noles, N., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception & Psychophysics, 67,* 324–334.

Olson, I. R., & Jiang, Y. (2004). Visual short-term memory is not improved by training. *Memory & Cognition, 32,* 1326–1332.

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

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9,* 441–474.

Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception & Psychophysics, 44,* 369–378.

Pedzek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 587–595.

Perrett, D. I., Oram, W. M., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalization of recognition without mental transformations. *Cognition, 67,* 111–145.

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

Redish, D. A., & Touretzky, D. S. (1997). Cognitive maps beyond the hippocampus. *Hippocampus, 7,* 15–35.

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.

Sacks, D. L., & Hollingworth, A. (2005, May). *Attending to original object location facilitates visual memory retrieval.* Paper presented at the annual meeting of the Vision Sciences Society, Sarasota, FL.

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics, 64,* 754–763.

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

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

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A,* 225–245.

Tanaka, J. W., & Sengco, J. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25,* 583–592.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2,* 55–82.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8,* 282–289.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition, 67,* 71–108.

Tarr, M. J., & Vuong, Q. C. (2002). Visual object recognition. In H. Pashler (Series Ed.) & S. Yantis (Ed.), *Stevens' handbook of experimental psychology: Vol. 1. Sensation and perception* (3rd ed., Vol. 1, pp. 287–314). New York: Wiley.

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

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Vogel, E. K., Woodman, G. E., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114.

Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Perception & Psychophysics, 67,* 676–690.

Received April 18, 2005
Revision received July 18, 2005
Accepted August 26, 2005 ■