## Object-Position Binding in Visual Memory for Natural Scenes and Object Arrays

#### Andrew Hollingworth University of Iowa

Nine experiments examined the means by which visual memory for individual objects is structured into a larger representation of a scene. Participants viewed images of natural scenes or object arrays in a change detection task requiring memory for the visual form of a single target object. In the test image, 2 properties of the stimulus were independently manipulated: the position of the target object and the spatial properties of the larger scene or array context. Memory performance was higher when the target object position remained the same from study to test. This same-position advantage was reduced or eliminated following contextual changes that disrupted the relative spatial relationships among contextual objects (context deletion, scrambling, and binding change) but was preserved following contextual change that did not disrupt relative spatial relationships (translation). Thus, episodic scene representations are formed through the binding of objects to scene locations, and object position is defined relative to a larger spatial representation coding the relative locations of contextual objects.

Keywords: scene memory, visual memory, binding, spatial memory, change detection

One of the fundamental functions of human cognition is to construct internal representations of the complex visual environments people typically inhabit. Recent work on the visual representation of natural scenes has debated whether, and to what extent, people are able to construct internal visual representations of scenes. Evidence from the phenomenon of change blindness suggests that visual scene representations are sparse (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997, 1998), perhaps limited to the currently attended object (Rensink, 2000). However, research designed to examine the capabilities of visual memory has found robust memory for the visual form of hundreds of individual objects in scenes (Hollingworth, 2004, 2005b; Hollingworth & Henderson, 2002) and thousands of wholescene images (Standing, Conezio, & Haber, 1970) over retention intervals of up to 1 year (Nickerson, 1968). Given evidence of high-capacity robust visual memory for objects and scenes, the present study examined how visual representations of individual objects are episodically structured to form a larger scale representation of a natural environment.

The strongest evidence of robust visual memory for objects in natural scenes comes from two studies by Hollingworth (2004, 2005b). In Hollingworth (2004), participants viewed images of scenes and fixated a series of objects that were cued sequentially in each scene. After all scenes had been viewed, memory for the visual form of individual objects was tested; a scene was represented with a target object cued, and participants responded to indicate whether the object was the same as the one viewed originally or whether it had been replaced by a different object from the same basic-level category (token change). Approximately 400 objects intervened between study and test of a particular object. Yet, memory performance was well above chance (68% correct). Hollingworth (2005b) added an orientation change condition (the target object was either the same as the original or rotated 90° in depth) and delayed the test 24 hr. Token and orientation change detection still remained well above chance. For example, in an office scene, participants saw many individual objects, one of which was a pen. After viewing all 48 scenes and hundreds of individual objects (any one of which could have been tested), and after 24 hr of delay, participants could still detect that the orientation of the pen had changed.

If participants can robustly retain visual representations from individual objects, are those representations episodically structured within a larger representation of the environment, and if so, what is the nature of that structure? Surprisingly, very little research exists on the question of object-to-scene binding. Hollingworth (2006) examined the fundamental issue of whether object memory is stored as part of a larger scene representation or independently of scene context. After scene viewing, memory for the visual properties of an object was tested (in a token or orientation discrimination task), with the target object displayed either within the original scene background or in isolation. Memory performance was consistently superior when the target object was presented within the scene background, demonstrating that object memory was stored as part of a more comprehensive scene representation. Jiang, Olson, and Chun (2000) and J. W. Tanaka and Farah (1993) have found similar context effects in visual shortterm memory (VSTM) for simple features and in long-term memory (LTM) for faces, respectively.

Hollingworth and Henderson (2002) proposed that larger scale scene representations are formed through the binding of local

Parts of this study were reported at the 44th Annual Meeting of the Psychonomic Society, Vancouver, British Columbia, Canada, November 2003. 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

object representations to positions in a spatial representation of the scene. In this view, as the eyes and attention are oriented within a scene, higher level visual representations are formed for attended objects and are bound to scene positions (Henderson, 1994; Hollingworth, 2005a; Irwin & Zelinsky, 2002; Kahneman, Treisman, & Gibbs, 1992; Zelinsky & Loschky, 2005). Visual object representations are likely maintained in infero-temporal (IT) 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 hippocampalparahippocampal place codes and IT object representations, similar to models of landmark-position binding in the rodent navigation literature (Gallistel, 1990; McNaughton et al., 1996; Redish & Touretzky, 1997). Initial support for the spatial binding hypothesis comes from the fact that participants can reliably remember the binding of object identity and position within real-world scenes (Diwadkar & McNamara, 1997; Hollingworth, 2005a; Irwin & Zelinsky, 2002; Rieser, 1989; Shelton & McNamara, 2001; Zelinsky & Loschky, 2005).

As a direct test of the spatial binding hypothesis, Hollingworth (2006) examined memory for individual objects in scenes, manipulating object position at test. Participants saw a scene image for 20 s, followed by a brief mask and two test scene alternatives. In one of the two alternatives, the target object was the same as the object originally viewed, and in the other alternative, the target object was either a different token or the original object rotated in depth. In addition, the target object alternatives were presented either in the same position in the scene as the target object had originally appeared or in a different position. If object property information is bound in memory to the spatial location of the object in the scene, then retrieval of object information should be less efficient when the target object changes position than when it retains the same position, leading to impaired memory performance when the target position is changed (e.g., Kahneman et al., 1992). Indeed, discrimination performance was more accurate in the same-position condition than in the different-position condition, a same-position advantage.

The binding of visual object representations to spatial positions is also found for nonscene stimuli in studies examining VSTM. Kahneman et al. (1992) demonstrated that STM for letter identity is spatially mediated, and Henderson and Siefert (2001) extended that finding to memory for the visual form of common objects. In addition, Irwin and Andrews (1996) found evidence that VSTM across saccades can maintain the binding of letter identity, color, and position for 3-4 objects. Finally, Jiang et al. (2000) found that the detection of color changes in a VSTM task was impaired when the spatial configuration of array elements changed from study to test. Jiang et al. argued that the spatial configuration of objects constitutes the principal episodic structure in which memory for the visual properties of individual objects is embedded. This configuration hypothesis also finds support from the contextual cuing phenomenon (Chun & Jiang, 1998), in which memory for the location of a target within a search array facilitates search even when the identities of the objects occupying each array location are randomly varied on each repeated search trial, implying that contextual spatial structure is abstracted away from the properties of the objects occupying each location in the configuration.<sup>1</sup> In conjunction with the results of Hollingworth (2006), the literature on

spatial binding in VSTM suggests that VSTM for abstract arrays of objects might engage the same episodic representational mechanisms as memory for complex, natural scenes.

#### The Present Study

Although there is strong evidence to suggest that object representations are structured according to spatial location, little is known about the properties of the spatial context constituting that structure. The following experiments investigated the nature of the contextual information serving to define object position, which is central to understanding how complex visual stimuli, such as natural scenes, are represented in the brain.

In each experiment, participants viewed an image of a scene or object array, followed by a brief interstimulus interval (ISI) and a test image. Figure 1 shows the basic method. In the test image, a target object was either the same as it appeared during initial viewing or changed; the task was change detection. In the changed condition, the target object was left–right mirror reflected (except in Experiment 3, in which the change was token replacement). In addition, the target either appeared in the same position within the scene or array or in a different position (Hollingworth, 2006); the same-position advantage was observed in all experiments.

To examine the properties of the contextual representation serving to define target position, I manipulated spatial contextual information available in the test image, and I observed the effect of contextual manipulations on the magnitude of the same-position advantage. If a particular contextual manipulation disrupts properties of the spatial context serving to define target position, then the same-position advantage should be reduced or eliminated when the context is changed. For example, if the same-position advantage were eliminated when the scene context was deleted at test, this would indicate that object position was defined relative to the particular scene context in which the object was originally viewed. If, however, a contextual manipulation does not influence the same-position advantage, one can infer that the manipulated property does not play a central role in defining object position. Thus, the magnitude of the same-position advantage can be used to identify the contextual factors serving to structure object information into spatially organized scene representations.

The first goal of this study was to replicate the Hollingworth (2006) same-position advantage for object memory in scenes and to test the generality of that finding to objects in random arrays (Experiments 1–3). The second, and primary, goal of the study was to identify contextual factors serving to structure memory for objects (Experiments 4–9). Each of these experiments examined the effect of contextual manipulations on the magnitude of the same-position advantage. In Experiments 4 and 5, the presence of the scene or array context at test was manipulated to establish whether object position is defined relative to the specific scene or array context in which an object was viewed. Experiments 6 and 7 addressed whether object position is defined relative to the positions of other visible objects by scrambling the locations of contextual objects at test. In Experiment 8, I examined whether spatial

<sup>&</sup>lt;sup>1</sup> More recent work on the contextual cuing phenomenon has demonstrated sensitivity to the identities of the objects comprising the configuration under some conditions (Jiang & Song, 2005).

#### **OBJECT-POSITION BINDING**





contextual representations maintain the bindings of contextual objects to locations by changing those bindings at test (preserving abstract spatial configuration). Experiment 9 addressed whether object position is defined allocentrically (relative to the configuration of contextual objects) by translating contextual objects at test. The final goal of this study was to build an empirical bridge between research examining contextual structure in VSTM (Jiang et al., 2000), which has typically used simple objects in random configurations and brief display durations, and research examining memory for objects in natural scenes, which has used coherent scene stimuli and longer study durations, often implicating LTM (Hollingworth, 2004, 2005b, 2006).

### Experiments 1–3: Position Specificity in Visual Object Memory

Before examining properties of scene context that serve to organize object memory, I conducted Experiments 1-3 to replicate and extend the same-position advantage found by Hollingworth (2006). In that study, object memory was superior when the target object was presented at the same position in the scene as it had been viewed originally. In a two-alternative forced-choice task, this advantage held both for memory for the in-depth orientation of the target object and memory for the token version of the target. Experiments 1–3 used a change detection task, with l (left) – r (right) mirror reflection (Experiments 1 and 2) or token replacement (Experiment 3) as the change. The use of change detection tasks is standard in the visual memory literature, and Experiments 1-3 sought to ensure that position-specificity effects would generalize to that method. The l - r mirror reflection change was used because there is no ambiguity about which object in the initial scene is being tested in the test scene. Because mirror reversal changes only the orientation of the object-preserving shape, surface features, and identity-the mapping of initial object to test object is maintained despite changes in position. Experiment 3

used a token change detection task to establish the generality of position effects in object memory.

Experiment 1 used natural scene stimuli (see Figure 1A). Participants saw an initial image of a scene for 20 s. After 20 s, a small dot was briefly onset and offset in a nontarget location,<sup>2</sup> followed by a 1,000-ms masked ISI. After the mask, a test scene was displayed. Figure 2 shows a sample initial scene and test scenes for the same-position and different-position conditions. The different position was on the opposite side of the screen at the same height as the original position and at the same distance from scene center. The target object in the test scene was presented within an olive green disk to ensure that position effects were not due to differences in the intersection of local contours between target and scene. The disk was surrounded by a neon green ring to provide a salient target postcue so that participants could limit decision processes to the target (Hollingworth, 2003). Participants responded to indicate whether the target object was in the same orientation or had been mirror reversed. In all experiments, verbal encoding was suppressed by means of digit repetition. And in all experiments, participants were informed that target position was irrelevant to the change detection task, providing no information about whether the target was the same or had been mirror reversed.

Experiments 2 and 3 used random arrays of individual objects. Figure 2 displays sample stimuli and manipulations for each experiment. The use of random arrays is important for three reasons. First, random arrays allowed examination of whether spatial mechanisms of episodic structure are natural scene specific or a more

<sup>&</sup>lt;sup>2</sup> The dot onset was carried over from experiments seeking to ensure that the target was not currently attended when it was tested (Hollingworth, 2003), 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, 2002).



*Figure 2.* Sample stimuli, manipulations, and accuracy data for Experiments 1–3. The left column shows the studied image. The center column shows sample test stimuli in the same- and different-position conditions. All test stimuli depict trials in which the target object has changed orientation (Experiments 1 and 2) or token (Experiment 3). The right column shows mean A' data for the same- and different-position conditions (error bars are standard errors of the means).

general property of the representation of complex visual stimuli. Second, the use of random arrays allows direct connection with the large literature on visual memory, which typically uses abstract arrays of objects, not scenes. Third, random arrays support contextual manipulations that are not feasible using natural scene stimuli (see Experiments 6–9). In addition to these general considerations, random arrays eliminate a specific problem in the Experiment 1 method. In the different-position condition of Experiment 1, the target object occluded portions of the scene that were previously visible at the different location (see Figure 2). With object arrays, the different-position target can be presented at a previously unoccupied location, so that changes in target position do not affect contextual information at test.

On each trial of the random array method (see Figure 1B), participants viewed an initial array of 6 objects for 4 s, followed by a 900-ms blank ISI, followed by a test array. Pilot testing indicated that the 4-s initial array duration produced intermediate levels of change detection performance for a set size of 6. The 6 objects were displayed in 6 of 12 possible locations (randomly chosen) in a  $3 \times 4$  virtual grid. All objects were full color, complex images derived from detailed 3-D models. In the test array, the target object was postcued by a green box. Participants responded to indicate whether the target in the test image was the same or changed. For the different-position condition, the new target position was chosen randomly from the positions unoccupied in the initial array.

#### Method

*Participants.* Participants in all of the experiments were recruited from the University of Iowa community. Each participant completed only one experiment. All participants were naive with regard to the hypotheses under investigation. All participants reported normal or corrected-to-normal vision. They either received course credit or were paid \$8. Twenty-four participants completed Experiment 1, 16 completed Experiment 2, and 16 completed Experiment 3.

*Apparatus.* Stimuli were presented on a 17-in. (43.18-cm) computer monitor with a refresh rate of 100 Hz. Stimulus presentation was synchronized to the refresh cycle. Responses were collected by a serial button box. Stimulus presentation and response collection were controlled by a PC-compatible computer running E-prime software (Schneider, Eschman, & Zuccolotto, 2002). Viewing distance was maintained at 80 cm by a forehead rest.

Stimuli: Experiment 1. Forty-eight scene images were rendered from 3-D models of real-world environments. In each model, a single target object was chosen. The test images were produced by rendering the target object within a uniform olive green (red–green–blue: 90, 110, 20) disk that was slightly larger than the target. Surrounding the disk was an eight-pixel-wide  $(0.23^{\circ})$  neon green ring (red–green–blue: 0, 255, 0). The target disk was presented either in the same position as the target had appeared at study or in a different position. The different position was at the same height in the scene but on the opposite side of the screen, the same distance from scene center as the original target position.

Scene stimuli subtended  $16.9^{\circ} \times 22.8^{\circ}$  visual angle. Target objects subtended  $3.21^{\circ}$  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 (red–green–blue: 0, 255, 0), with a diameter of  $1.2^{\circ}$ . It appeared in a position within each scene unoccupied by any object that could plausibly be considered a target.

Stimuli: Experiments 2 and 3. The stimulus set for Experiment 2 consisted of 40 images of common objects rendered from detailed 3-D

models. For each object, 2 mirror image versions were created. Each object was chosen to be asymmetrical so that mirror reversal introduced a significant change. Figure 2 shows a mirror reversal of the cell phone object.

The stimuli for Experiment 3 consisted of a different set of 40 common objects, approximately half of which were used in the Experiment 2 set. For each object, 2 token versions were created. Tokens in each pair were chosen to be approximately the same size and were rendered in the same orientation. The color of major parts was also made equivalent. All token pairs were equivalent at the basic level of categorization. Some token pairs differed at the subordinate category level (e.g., a mountain and street bicycle), whereas others were equivalent at the subordinate level (e.g., two differently shaped watering cans; see Figure 2).

In both experiments, objects were sized so that each object fit within a  $4.92^{\circ} \times 4.92^{\circ}$  square. The object array stimuli were created by dividing the screen into a virtual  $3 \times 4$  grid. Six objects were randomly assigned to 6 of the 12 cells. The target object item was determined on the basis of the condition item assignments described below. The other 5 objects were selected randomly without replacement from the remaining 39 objects in the set. The orientation (Experiment 2) or token version (Experiment 3) of each distractor object was also determined randomly. The array background was set to a neutral gray. For the test images, the target object was postcued by a neon green box. For the different-position conditions, the target was moved to a different, unoccupied grid cell. The new position was selected randomly from the 6 empty locations, with the constraint that the new target cell could not be immediately adjacent to the old target cell, ensuring a significant change in location. The entire array subtended 16.9°  $\times 22.8^{\circ}$ .

*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 or object arrays. After viewing each stimulus, they would have to decide if a single object was the same or changed (the nature of the possible change was described). They were instructed that the target object might appear in the same location as it had originally or in a different location, but position had no influence on whether the target was the same or changed.

Each trial began with a screen instructing participants to *Press a button to begin next trial.* This screen also contained four randomly chosen digits. Participants began repeating the digits aloud at a rate of at least two digits per second and continued repetition throughout the trial. Digit repetition was used for verbal suppression. In all experiments, participants pressed a button to begin the trial. Then, a white fixation cross on a gray field was displayed for 1,000 ms.

In Experiment 1 (see Figure 1A), the study scene was then presented for 20 s. A green dot appeared abruptly within the scene for 150 ms and was removed for 200 ms. Next, a pattern mask was displayed for 1,000 ms. Finally, the test scene was displayed until response. In Experiments 2 and 3 (see Figure 1B), the study array was then presented for 4 s, followed by a 900-ms blank (gray) ISI and the test array.

When the test scene or array appeared, participants pressed one button to indicate that the target object was the same or a different button to indicate that it had changed. They were instructed to take as long as necessary to respond accurately but no longer. Thus, accuracy was the principal dependent measure, with response time (RT) providing converging evidence. Button response terminated the trial. There was a 2-s delay between trials. In the object array experiments, correct–incorrect feedback was provided on every trial.

Experiment 1 was a 2 (same target position, different target position)  $\times$  2 (target same, changed) factorial design. Participants first completed a practice session of eight trials, two in each of the four conditions. The scene items used for the practice trials were not used in the experimental session. In the experiment session of Experiment 1, participants viewed all 48 scene items, 12 in each of the four conditions. Trial order was determined randomly. For each participant, each of the 48 scene items was viewed only once; there was no scene repetition. Across participants,

condition-item assignments were counterbalanced by Latin square so that each scene item appeared in each condition an equal number of times.

Experiments 2 and 3 were a 2 (same target position, different target position)  $\times$  2 (target same, changed)  $\times$  2 (initial target version) factorial design. Participants first completed a practice session of 8 trials, 1 in each of the 8 conditions. In the experiment session, participants completed 160 trials, 20 in each of the 8 conditions. Each of the 40 object items was the target on 4 trials. Across participants, each object item was the target in each condition an equal number of times. Trial order was determined randomly. All experiments required approximately 45 min to complete.

#### Results

Percentage correct data were used to calculate A', a signal detection measure with a functional range of .5 (chance) to 1.0 (perfect sensitivity). For each participant in each target position condition, A' was calculated using the mean hit rate when the target changed and the mean false alarm rate when it did not.<sup>3</sup> Because A' corrects for potential differences in response bias in the percentage correct data, it forms the primary data for interpreting these experiments. In all experiments, percentage correct and d' measures produced the same pattern of results as A'. Raw percentage correct data and RT data for all experiments are reported in the Appendix. Mean A' performance in each of the target position conditions is displayed in Figure 2.

Experiment 1: Natural scenes, l - r mirror reversal change detection. Change detection accuracy was reliably higher in the same-position condition than in the different-position condition, F(1, 23) = 7.67, p = .011,  $\eta^2 = .25$ . Consistent with the A' data, RT for all responses was reliably faster in the same-position condition (1,864 ms) than in the different-position condition (2,192 ms), F(1, 23) = 39.8, p < .001,  $\eta^2 = .63$ . In this and in all subsequent experiments, RT for correct responses only produced the same pattern of results as RT for all responses.<sup>4</sup>

One possible concern with the Experiment 1 results is that for some of the scene items, the different position was not a plausible location for the target. The scene depicted in Figure 2 is an example of such a scene; the patio umbrella could not plausibly appear at the different position, as it would have been partially occluded by the walls of the room. It is possible that the sameposition advantage might have been generated because the original position was always plausible, whereas the different position was often implausible. To assess this possibility, I conducted a secondary analysis limited to the 24 scene items for which the different position was also a plausible location for the target. The sameposition advantage was actually larger for this subset of scenes (same position, A' = .88; different position, A' = .79), F(1, 23) =8.00, p = .010, than for the full data set, eliminating the possibility that differences in position plausibility caused the same-position advantage.

Experiment 2: Object arrays, l - r mirror reversal change detection. Change detection accuracy was reliably higher in the same-position condition than in the different-position condition, F(1, 15) = 14.2, p = .002,  $\eta^2 = .49$ . Consistent with the *A'* data, RT was reliably faster in the same-position condition (1,247 ms) than in the different-position condition (1,384 ms), F(1, 15) = 18.5, p < .001,  $\eta^2 = .55$ .

*Experiment 3: Object arrays, token change detection.* Change detection accuracy was reliably higher in the same-position condition than in the different-position condition, F(1, 15) = 6.01,

p = .027,  $\eta^2 = .29$ . Consistent with the *A'* data, RT was reliably faster in the same-position condition (1,290 ms) than in the different-position condition (1,372 ms), F(1, 15) = 24.7, p < .001,  $\eta^2 = .62$ .

#### Discussion

The results from Experiments 1–3 provide direct support for the Hollingworth and Henderson (2002) claim that scene representations are constructed by binding visual object representations to scene locations. Second, they suggest that IT representations coding object form are not stored independently of the scene in which an object appears; it is likely that associative connections exist between medial temporal regions coding spatial properties of the environment and IT representations of object form, supporting episodic scene representations (Hollingworth, 2006).

The magnitude and generality of the position effects are noteworthy. Many studies demonstrating spatially mediated object memory have produced effects only on RT, with RT differences on the order of 20-30 ms (e.g., Kahneman et al., 1992). In Experiments 1-3, position effects were observed both on accuracy and on RT (with the RT effects an order of magnitude larger than in previous studies). Changes in spatial position influenced not only the efficiency of visual memory retrieval and comparison but also the ultimate success or failure of those operations. In addition, the same-position advantage has now been observed both in the present change detection task and in a forced-choice task (Hollingworth, 2006), both using natural scenes and object arrays, and in methods probing three different types of object information (l - r)orientation, in-depth orientation, and object token). Position specificity is clearly a fundamental property of episodic object memory.

#### Experiments 4 and 5: Context Deletion

The remaining experiments in this article addressed the central issue of this study: What contextual factors serve to organize object memory into spatially structured scene representations? Each experiment examined the effects of contextual manipulations on the magnitude of the same-position advantage. Experiments 4

$$A' = \frac{1}{2} + \frac{(y-x)(1+y-x)}{4y(1-x)}$$

where y is the hit rate and x is the false alarm rate. In the one case that a participant performed below chance in a particular condition, A' was calculated using the below-chance equation developed by Aaronson and Watts (1987):

$$A' = \frac{1}{2} - \frac{(x - y)(1 + x - y)}{4x(1 - y)}.$$

<sup>4</sup> For the experiments using a discrete set of natural scene stimuli (Experiments 1 and 4), analyses treating scene item as a random effect yielded the same pattern of statistical significance observed in the main analyses treating participant as a random effect. Item analyses examining accuracy were conducted over percentage correct data, because A' requires the estimation of individual participant decision criteria.

<sup>&</sup>lt;sup>3</sup> For above-chance performance, A' was calculated as specified by Grier (1971):

and 5 examined whether object position is defined relative to the particular scene or array context in which the object was viewed.

In Experiment 4, the presence of the scene context in the test image was manipulated. The background-present condition was identical to Experiment 1. The background-absent condition was identical to Experiment 1, except the scene background was eliminated in the test image: The test image consisted of the target object (displayed within the olive green disk) in an otherwise empty (neutral gray) field, as illustrated in Figure 3. The absolute positions of the target in the background-absent condition were the same as those in the background-present condition. The following pattern of results is predicted by the hypothesis that object visual memory is bound to a location within the particular scene context in which the object was viewed (Hollingworth & Henderson, 2002). First, in the background-present condition, the sameposition advantage should be observed, as in Experiments 1–3. Second, in the background-absent condition, the same-position advantage should be reduced or eliminated, as the context serving to define target position has been eliminated.

In Experiment 5, the background-presence manipulation was extended to random object arrays. The background-present condition was identical to Experiment 2. In the background-absent condition, the target object was presented in isolation.

#### Method

# *Participants.* Forty-eight participants completed Experiment 4, and 16 completed Experiment 5.

*Stimuli and apparatus.* In Experiment 4, the scene stimuli were the same as in Experiment 1, with the exception of the background-absent test images. For the background-absent images, the target (within the blank disk and surrounded by the neon green ring) was displayed against a neutral gray background in the same locations as those used in the background-present conditions.



*Figure 3.* Sample stimuli, manipulations, and accuracy data for Experiments 4 and 5. The left column shows the studied image. The center column shows sample test stimuli in each of the principal conditions. All test stimuli depict trials in which the target object has changed orientation. The right column shows mean A' data (error bars are standard errors of the means). Pos = position; Diff = different.

In Experiment 5, the set of 40 common objects used in Experiment 2 was increased to 80. For the background-absent test images, the target object was displayed in isolation. In all other respects, the stimuli were identical to those in Experiment 2.

*Procedure.* For Experiment 4, the sequence of events in a trial was the same as in Experiment 1. The background-presence manipulation was blocked. Each block began with 4 practice trials. Practice scenes were not used in the experimental session. Across the 2 blocks, participants viewed all 48 scene items, 6 in each of the 8 conditions created by the 2 (background present, absent)  $\times$  2 (target same position, different position)  $\times$  2 (target same, changed) factorial design. Block order was counterbalanced across participant groups. Within each block, trial order was determined randomly. For each participant, each scene item was viewed once. Across participants, condition-item assignments were counterbalanced by Latin square so that each scene item appeared in each condition an equal number of times.

For Experiment 5, the sequence of events in a trial was the same as in Experiment 2. The background-presence manipulation was blocked. Each block began with 8 practice trials. Across the 2 blocks, participants completed 240 experimental trials, 15 in each of the 16 conditions created by the 2 (background present, absent)  $\times$  2 (same target position, different target position)  $\times$  2 (target same, changed)  $\times$  2 (initial target version) factorial design. Each of the 80 object items was the target on 3 trials. Across participants, each object item was the target in each condition an equal number of times. Block order was counterbalanced across participant groups. Within each block, trial order was determined randomly.

#### Results

Mean A' performance in each of the background-presence and target-position conditions is displayed in Figure 3. In all of the experiments manipulating context and target position (4–9), the critical analysis was the interaction between these two factors, as the same-position advantage was predicted to be reduced if the contextual manipulation disrupted relevant spatial context. The main effect analyses were secondary and are reported in parentheses following report of the interaction.

*Experiment 4: Natural scenes.* There was a marginally reliable interaction between background presence and target position, F(1,47) = 2.99, p = .090. (Main effect of background presence, F < .0901; main effect of target position, p = .06.) To provide a stronger assessment of the interaction between background presence and target position, I conducted an additional analysis of variance using log-transformed percentage correct data.<sup>5</sup> Over logtransformed percentage correct, the interaction between background presence and target position was statistically reliable, F(1,47) = 4.05, p = .049. A' contrasts revealed a reliable sameposition advantage in the background-present condition, F(1,47) = 8.24, p = .006,  $\eta^2 = .15$ , but no effect of target position in the background-absent condition (F < 1,  $\eta^2 = 0.00$ ). In addition, when target position was the same, there was a reliable advantage for the background-present condition over the background-absent condition, F(1, 47) = 4.48, p = .040.

The RT data complemented the accuracy data. The sameposition advantage (different position RT – same position RT) was larger in the background-present condition (592 ms,  $\eta^2 = .39$ ) than in the background-absent condition (52 ms,  $\eta^2 = .02$ ), F(1, 47) =20.6, p < .001.

*Experiment 5: Random arrays.* There was a near-significant interaction between background presence and target position, F(1, 15) = 4.50, p = .051. (Main effect of background presence, F < 150)

1; main effect of target position, p = .002.) The interaction was only marginal over log-transformed percentage correct data, F(1, 15) = 2.86, p = .112. Planned contrasts revealed a reliable same-position advantage in the background-present condition, F(1, 15) = 18.42, p < .001,  $\eta^2 = .55$ , but no effect of target position in the background-absent condition, F(1, 15) = 1.62, p = .22,  $\eta^2 = .10$ . Unlike Experiment 4, when target position was the same, there was not a reliable advantage for the background-present condition over the background-absent condition (F < 1), although the numerical trend was in the same direction.

The RT data complemented the accuracy data. Although the interaction between background presence and target position was only near reliable over accuracy data, that interaction was reliable over RT data, F(1, 15) = 10.2, p = .006; the same-position advantage was larger in the background-present condition (206 ms,  $\eta^2 = .63$ ) than in the background-absent condition (51 ms,  $\eta^2 = .27$ ).

#### Discussion

In Experiments 4 and 5, the same-position advantage was observed when the scene or array context was present at test but was essentially eliminated when the scene or array context was absent at test, demonstrating that target position was coded relative to the particular scene or array context in which the target object was viewed. Note again that the size of the position effects in these experiments was remarkably large when compared with effects of position specificity found in previous studies. Reliable differences in accuracy were observed, and RT effects were on the order of 200–600 ms.

#### Experiments 6-8: Background Position Manipulations

Experiments 4 and 5 demonstrated that object position was coded relative to the particular scene or array context in which the object was viewed, but they did not isolate spatial properties of the context, because deletion of the entire background eliminates more than just spatial information. In Experiment 6, spatial properties of the context were isolated from nonspatial properties (such as object or scene identity) by retaining all of the original objects but scrambling their locations at test. That is, on half the trials, the contextual objects retained their original positions. On the other half of trials, the contextual objects were all moved to new, previously unoccupied locations (i.e., scrambled). Target object position was manipulated independently of background scrambling. Figure 4 illustrates the design. If target position is defined relative to the positions of contextual objects, the advantage for same target position should be reduced or eliminated when those background positions are scrambled. This experiment used object arrays. Scrambling locations is not possible in 3-D scenes without changing the visual appearance of the objects. However, given the tight correspondence between the results with scenes and arrays thus far, object arrays appear to engage similar representational mechanisms as natural scenes.

<sup>&</sup>lt;sup>5</sup> Because sensitivity measures such as A' are not scaled linearly, log-transformed percentage correct (which is linearly scaled) is preferable for testing interaction effects.

#### **OBJECT-POSITION BINDING**



*Figure 4.* Sample stimuli, manipulations, and accuracy data for Experiments 6-8. The left column shows the studied image. The center column shows sample test stimuli in each of the principal conditions. All test stimuli depict trials in which the target object has changed orientation. The right column shows mean A' data (error bars are standard errors of the means). Pos = position; Diff = different.

Experiment 7 was identical to Experiment 6, except the number of array objects was reduced from six to four, and the initial array duration was decreased from 4 s to 400 ms. The context scrambling manipulation and 400-ms display duration of Experiment 7 generate a paradigm that is similar to the paradigm used in Jiang et al. (2000) and is able to test whether the binding of object representations to the larger spatial context extends to brief presentation conditions tapping VSTM. In Experiment 3 of Jiang et al. (2000), participants viewed an array of color patches or simple shapes for 400 ms. After a 900-ms ISI, a test array appeared, with a target object same or changed. In the test image, the positions of all objects were the same, the positions of all objects (including the target) were changed in a manner that preserved the basic configuration of elements (systematic expansion), or the position of all objects was changed in a manner that altered the spatial configuration. Change detection was impaired in the last condition relative to the first two, and performance was not different in the two conditions that retained the original configuration. However, Jiang et al. did not independently manipulate contextual positions and target position (when the configuration changed, the target position changed as well) and therefore could not isolate the contextual factors serving to define object position. In the present experiment, target position was manipulated independently of contextual positions, isolating contextual factors in visual memory, as indexed by changes in the same-position advantage.

Experiment 8 tested the Jiang et al. (2000) hypothesis that spatial configuration is the primary organizing structure in visual memory. Specifically, Jiang et al. argued that the visual representation of an individual object is associated with a position within an abstract spatial representation coding the relative locations (the configuration) of all of the items in the array. This contextual representation is proposed to be purely spatial, coding occupied locations but not necessarily information about the objects occupying those locations (although there may be independent contextual contributions from memory for object form). This view predicts that alterations to the configuration of array elements will impair change detection performance, but manipulations that do not alter configuration should not impair change detection performance.

In Experiment 8, the binding of contextual objects to locations was manipulated, as illustrated in Figure 4. In the backgroundsame condition, the background objects retained their original positions at test. In the background-binding-change condition, each of the contextual objects randomly traded positions. In the latter condition, the positions occupied by contextual objects (the spatial configuration) did not change; only the binding of contextual objects to locations was disrupted. Target position was manipulated independently of background binding.

The spatial configuration hypothesis (Jiang et al., 2000) predicts that target change detection should not be impaired when the binding of contextual objects to locations changes, as this manipulation does not alter abstract spatial configuration. The spatial configuration hypothesis also predicts that the same-position advantage should be observed in both the background-same condition and in the background-binding-change condition, with no interaction between background binding and target position. In contrast, if object position is defined relative to a contextual representation preserving not only the contextual locations but also information about the objects bound to each location (e.g., as a set of object files; Hollingworth & Henderson, 2002), change detection performance should be impaired when the contextual bindings change, and the same-position advantage should be reduced or eliminated when the contextual bindings change.

#### Method

*Participants.* Sixteen participants completed Experiment 6, 32 completed Experiment 7, and 16 completed Experiment 8.

Stimuli and apparatus. Experiments 6–8 used the same set of 80 objects as used in Experiment 5. In Experiments 6 and 8, each array contained 6 objects, as in previous object array experiments. In Experiment 7, each array contained 4 objects. For the background scrambled test images in Experiments 6 and 7, the contextual objects were randomly assigned to cells that had not contained an object in the study image. For the background-binding-change test images in Experiment 8, the contextual objects randomly traded positions, preserving the original configuration. As in previous experiments, in the different-position condition, the target was moved to a previously unoccupied cell.

*Procedure.* The sequence of events in a trial was the same as in the previous object array experiments, except that Experiment 7 presented the study array for 400 ms instead of 4 s.

The background manipulations were blocked. Each block began with 8 practice trials. Across the 2 blocks, participants completed 240 experimental trials, 15 in each of the 16 conditions created by the 2 (background same, changed)  $\times$  2 (target same position, different position)  $\times$  2 (target same, changed)  $\times$  2 (initial target version) factorial design. Across participants, each object item was the target in each condition an equal number of times. Block order was counterbalanced across participant groups. Within each block, trial order was determined randomly.

#### Results

Mean A' performance in each of the background conditions and target-position conditions is displayed in Figure 4.

*Experiment 6: Background scrambling.* There was a reliable interaction between background scrambling and target position, F(1, 15) = 6.82, p = .020. (Main effect of background scrambling, p = .005; main effect of target position, p = .003.) This interaction was also reliable over log-transformed percentage correct data, F(1, 15) = 7.81, p = .014. Planned contrasts revealed a reliable advantage for the same target position in the background-same condition, F(1, 15) = 22.8, p < .001,  $\eta^2 = .60$ , but no effect of target position in the background-scrambled condition (F < 1,  $\eta^2 = .01$ ). When target position was the same, there was a reliable advantage for the background-same condition over the background-scrambled condition over the background-scrambled condition F(1, 15) = 16.9,  $p < .001.^6$ 

The RT data were consistent with the accuracy data. The sameposition advantage was larger in the background-same condition (138 ms,  $\eta^2 = .81$ ) than in the background-scrambled condition (58 ms,  $\eta^2 = .24$ ), F(1, 15) = 5.95, p = .028.

*Experiment 7: Background scrambling, four-object array,* 400-ms study. There was a reliable interaction between background scrambling and target position, F(1, 31) = 5.64, p = .024.

<sup>&</sup>lt;sup>6</sup> The effect of background scrambling was replicated in an additional experiment (N = 12) that included only the background-scrambling manipulation (target position was always the same). Accuracy was reliably higher in the background-same condition (A' = .841) than in the background-scrambled condition (A' = .770), F(1, 11) = 8.31, p = .015. In addition, RT was reliably faster in the background-same condition (1,397 ms) than in the background-scrambled condition (1,515 ms), F(1, 11) = 10.4, p = .008.

(Main effect of background scrambling, p = .003; main effect of target position, p < .001.) This interaction was also reliable for log-transformed percentage correct data, F(1, 31) = 5.45, p = .026. Planned contrasts revealed a reliable same-position advantage in the background-same condition, F(1, 31) = 26.5, p < .001,  $\eta^2 = .46$ , but no effect of target position in the background-scrambled condition, F(1, 31) = 1.89, p = .18,  $\eta^2 = .06$ . When target position was the same, there was a reliable advantage for the background-scrambled condition over the background-scrambled condition, F(1, 31) = 30.2, p < .001.

The RT data were consistent with the accuracy data. The sameposition advantage was larger in the background-same condition (128 ms,  $\eta^2 = .37$ ) than in the background-scrambled condition (44 ms,  $\eta^2 = .16$ ), F(1, 15) = 10.65, p = .003.

*Experiment 8: Background binding change.* There was a reliable interaction between background binding and target position, F(1, 15) = 10.0, p = .007. (Main effect of background binding, F < 1; main effect of target position, p = .001.) This interaction was also reliable for log-transformed percentage correct data, F(1, 15) = 7.57, p = .015. Planned contrasts revealed a reliable same-position advantage in the background-same condition, F(1, 15) = 23.1, p < .001,  $\eta^2 = .61$ , but no effect of target position in the background-binding-change condition (F < 1,  $\eta^2 = .05$ ). When target position was the same, there was a reliable advantage for the background-same condition over the background-binding-change condition, F(1, 15) = 6.96,  $p = .019.^7$ 

The RT data were consistent with the accuracy data. The sameposition advantage was larger in the background-same condition (258 ms,  $\eta^2 = .64$ ) than in the background-binding-change condition (106 ms,  $\eta^2 = .57$ ), F(1, 15) = 8.23, p = .012.

#### Discussion

In Experiments 6 and 7, an advantage was observed for the background-same condition over the background-scrambled condition when keeping target position the same. In addition, the same-position advantage was eliminated when the background was scrambled, demonstrating that target position is defined relative to the positions of other contextual objects.

The findings of Experiments 6 and 7 are consistent with the configuration hypothesis of Jiang et al. (2000), as changes in array configuration disrupted the same-position advantage. However, the results of Experiment 8 do not support the hypothesis that configural representations maintain only the abstract spatial layout of occupied locations (Jiang et al., 2000). Contextual binding change, which changed the binding of contextual objects to locations but did not disrupt abstract spatial configuration, produced a significant decline in change detection performance and significantly reduced the same-position advantage. This result supports the hypothesis that object position is defined relative to a contextual representation that preserves not only spatial positions of contextual objects but also information about the individual objects bound to those locations (Hollingworth & Henderson, 2002).

In addition, the findings of position specificity across scene stimuli, array stimuli under relative long encoding conditions (4 s), and array stimuli under brief encoding conditions (400 ms) suggest that similar contextual mechanisms are at work in all of these cases. For scene stimuli viewed for 20 s, visual representation would have been largely dependent on LTM, as found by Holling-

worth (2004). In Experiment 7, with a 400-ms presentation of four objects, array representation was likely to have been dependent on VSTM. For the same-background/same-position condition in Experiment 7, Pashler's (1988) formula for computing the number of objects retained in memory yielded an estimate of 1.9 objects. This estimate is consistent with independent estimates of VSTM capacity for complex objects in Hollingworth (2004) and in Alvarez and Cavanagh (2004). Although there do not exist means to ensure that methods designed to isolate VSTM are entirely free of LTM influences, the memory capacity observed in Experiment 7 indicates little or no contribution from LTM. In contrast, the capacity estimate in Experiment 6 (4-s study duration) was four objects, which easily exceeds estimates of VSTM capacity for complex objects (Alvarez & Cavanagh, 2004; Hollingworth, 2004). Similar effects of position and context were observed across all of these different methods, suggesting that both VSTM and LTM depend on functionally similar mechanisms of spatial contextual structure.

#### **Experiment 9: Translation**

In all experiments thus far, the same target position has always been the same absolute screen position. Experiment 9 dissociated absolute and array-relative position by means of an array translation manipulation. Figure 5 shows the four principal conditions. In the all-same condition, all objects retained their original positions at test. In the target-shift-only condition, the background positions remained the same, but the target object was moved two object positions horizontally or vertically. In these first two conditions, a replication of the advantage for same target position was expected (all-same more accurate than target-shift-only). In the backgroundshift-only condition, all of the contextual objects were shifted together two object positions horizontally or vertically, but the target was not shifted. In this condition, the target's absolute position remained the same, but its position relative to the array changed. Finally, in the all-shift condition, all objects were shifted together two object positions horizontally or vertically (full array translation). In this condition, the absolute position of the target changed, but its position relative to the array remained the same.

If spatial contextual structure is established in an array-relative reference frame, then performance in the all-shift condition (target in same relative position, different absolute position) should be higher than performance in the background-shift-only condition (target in different relative position, same absolute position). However, if objects are bound to absolute screen locations, the reverse data pattern should be obtained.

#### Method

*Participants.* Forty-eight participants completed the experiment. *Stimuli.* The stimuli were displayed within cells of a  $7 \times 6$  virtual grid. A  $5 \times 4$  subsection of the grid was used to display the initial

<sup>&</sup>lt;sup>7</sup> The effect of background binding change was replicated in an additional experiment (N = 12) that included only the background-binding manipulation (target position was always the same). Accuracy was reliably higher in the background-same condition (A' = .844) than in the background-binding-change condition (A' = .773), F(1, 11) = 6.90, p = .024. In addition, RT was reliably faster in the background-same condition (1,315 ms) than in the background-binding-change condition. F(1, 11) = 9.60, p = .010.

#### HOLLINGWORTH



*Figure 5.* Sample stimuli, manipulations, and accuracy data for Experiment 9. The left column shows the studied image. The center column shows sample test stimuli in each of the principal conditions. All test stimuli depict trials in which the target object has changed orientation. The right column shows mean A' data (error bars are standard errors of the means).

object array on each trial. The  $5 \times 4$  grid was located in one of the four corners of the larger  $7 \times 6$  space (in Figure 5, the  $5 \times 4$  array space occupies the bottom right corner). The 6 objects in the initial array were randomly assigned to 6 of the 20 possible locations, with the condition that for each object, no other object could appear in a cell that was 2 cells away either vertically or horizontally. This was necessary because object shifts were always 2 cells horizontal or vertical. The test images were constructed as follows. In the all-same condition, all object positions remained the same. In the target-shift-only condition, the target object was shifted 2 cells either horizontally or vertically. The direction of the shift depended on the screen location of the initial array. For example, when the array occupied the bottom right corner, the shift was either 2 positions up or 2 positions to the left (with direction randomly chosen). Two possible shift directions ensured that participants could not predict the direction of the shift. In the backgroundshift-only condition, all objects except the target were shifted 2 positions in the same direction, again either horizontally or vertically; the target retained its original position. In the all-shift condition, all objects were shifted together 2 positions, horizontally or vertically (i.e., the entire array was translated). In each test image, the target object was postcued, and it was either the same as in the initial array or mirror reversed. Each cell in the  $7 \times 6$  grid subtended  $3.28^{\circ} \times 3.28^{\circ}$ . The set of 40 array objects was used, and these were sized so that each fit within a  $2.6^{\circ} \times 2.6^{\circ}$  square region at the center of each cell.

*Procedure.* The sequence of events in a trial was the same as in the previous object array experiments, with a 4-s initial array duration.

Participants completed 160 trials, 10 in each of the 16 conditions created by the 4 (all-same, target-shift-only, background-shift-only, all-shift)  $\times$  2 (target same, changed)  $\times$  2 (initial target version) factorial design. Each object appeared in each condition an equal number of times. The 4 major conditions were blocked, and block order counterbalanced across participants. Within a block, trial order was determined randomly.

#### Results

Mean A' performance in each of the background-shift and target-shift conditions is displayed in Figure 5. There was a reliable interaction between background shift and target shift, F(1,47) = 12.8, p < .001. This interaction was also reliable over log-transformed percentage correct data, F(1, 47) = 11.9, p =.001. In the all-same and target-shift-only conditions, the sameposition advantage was observed again, F(1, 47) = 11.0, p = .002,  $\eta^2 = .19$ . Of central interest, detection accuracy was higher in the all-shift condition (target same relative position, different absolute position) than in the background-shift-only condition (target different relative position, same absolute position), F(1, 47) = 5.04,  $p = .030, \eta^2 = .10$ . Translation of the array was not cost free, however. Performance in the all-same condition was higher than performance in the all-shift condition, F(1, 47) = 4.86, p < .05. This difference might arise from the need to shift attention to a new area of the computer screen, or it could indicate some binding of object memory to the original absolute locations. In either case, however, the preservation of the same-position advantage with array translation indicates that array-relative coding was the controlling factor in memory retrieval and change detection performance.

RT was faster in the all-same condition than in the target-shiftonly condition, F(1, 47) = 13.1, p < .001,  $\eta^2 = .22$ . However, RT did not differ between the background-shift-only and all-shift conditions (F < 1,  $\eta^2 = 0.00$ ).

As a converging analysis, the factorial design of Experiment 9 was restructured to compare the effect of change in relative target position with the effect of change in absolute target position: 2 (relative position: same, changed)  $\times$  2 (absolute position: same,

changed).<sup>8</sup> Accuracy was reliably higher when relative position was the same (A' = .889) than when the relative position changed (A' = .853), F(1, 47) = 12.8, p < .001. However, accuracy was no higher when absolute position was the same (A' = .872) than when absolute position changed (A' = .869; F < 1). These results provide further evidence that object-position binding occurs in array-relative coordinates.

In Experiment 9, the same-position advantage was governed by array-relative position, not by absolute position, demonstrating that object position is coded in array-relative coordinates. In addition, the translation manipulation in Experiment 9 shows that not all contextual transformations disrupt the same-position advantage. Only transformations that altered the spatial relationship between contextual objects (deletion, scrambling, background-bindingchange) eliminated the same-position advantage. Array translation, which preserved relative spatial relationships, did not.

#### General Discussion

The present experiments were designed to examine how visual object representations are structured to form a larger episodic representation of a natural scene or object array. Experiments 1–3 extended the finding of position-specific memory for the visual form of objects (Hollingworth, 2006). Changes to the visual properties of target objects were more accurately detected when the object retained its original position from study to test than when it occupied a different position at test, a same-position advantage.

Experiments 4–9 used the same-position advantage as a means to examine the properties of contextual information serving to define object position. Experiments 4 and 5 showed that target position is defined relative to the particular scene or array context in which the object appeared, as the deletion of the background context at test eliminated the same-position advantage. Experiments 6-9 examined the nature of the spatial representation serving to structure object memory. Scrambling the positions of the contextual objects at test led to a significant deficit in change detection performance, and the same-position advantage was eliminated when the configuration of contextual objects was disrupted by scrambling (Experiment 6 and 7), indicating that contextual representations maintain the spatial configuration of objects. Experiment 8 demonstrated that spatial contextual representations maintain not only the abstract configuration of contextual locations (Jiang et al., 2000) but also the binding of individual objects to each of those locations. Changes in the binding of contextual objects to locations (preserving the original configuration) significantly impaired change detection performance, and changes in background binding essentially eliminated the same-position advantage. Finally, target location appears to be coded in array-relative coordinates; the same-position advantage was governed by array-relative position, not by absolute position (Experiment 9). The results from this study are consistent with the proposal of Hollingworth and Henderson (2002) that scene representations are constructed by the binding visual object representations to particular locations within a spatial representation of the larger scene context.

These results provide an opportunity to expand the theory of scene perception and memory originally described in Hollingworth and Henderson (2002). While one views a scene, the eyes and attention are oriented serially from object to object within the scene. During each fixation, low-level sensory representations are generated across the visual field, but these are fleeting. Visual sensory representations (i.e., iconic memory) decay within a few hundred milliseconds after a stimulus event (Averbach & Coriell, 1961; Di Lollo, 1980; Irwin & Yeomans, 1986; Sperling, 1960) and are not integrated across disruptions such as saccadic eye movements (Henderson & Hollingworth, 2003; Irwin, 1991). Although sensory representations are fleeting, the visual system constructs more durable, higher level object representations (abstracted away from precise sensory features) that are maintained in VSTM and LTM (Biederman & Cooper, 1991; Carlson-Radvansky & Irwin, 1995; Henderson, 1997; Irwin, 1991; Phillips, 1974; Pollatsek, Rayner, & Collins, 1984). Attention plays a critical role in the encoding of object representations during scene viewing. Attention is necessary to bind perceptual features into a higher level object representation capable of supporting object recognition (Treisman, 1988). Attention also supports the consolidation of higher level visual representations into memory (Averbach & Coriell, 1961; Hollingworth & Henderson, 2002; Irwin & Gordon, 1998; Schmidt, Vogel, Woodman, & Luck, 2002; Sperling, 1960). During scene viewing, object representations are maintained in VSTM for approximately the last two objects fixated and attended in the scene, with memory for objects attended earlier maintained in high-capacity, robust LTM (Hollingworth, 2004, 2005b; Zelinsky & Loschky, 2005). Thus, for attended objects, VSTM and LTM support the accumulation of object representations across shifts of the eyes and attention and enable the construction of relatively elaborate visual representations of scenes.

Once encoded into memory, higher level object representations are bound to locations within a spatial representation of the particular scene or array context in which the object appeared, generating an episodic scene representation. This contextual representation is scene specific, preserves the binding of contextual objects to locations in a configural representation, and codes object position in array-relative coordinates. The reduction (and in most cases elimination) of the same-position advantage with changes in contextual information was likely caused by the remapping of spatial representations following changes in context. If there is sufficient evidence to suggest that the scene or array context has changed (i.e., that a different environment is being depicted), then a new spatial contextual representation will be activated. Because object representations are associated with scene-specific spatial representations, if the active spatial representation is not the one to which the object information was bound, the same-position condition should not exhibit an advantage over the different-position condition. Changes in background presence (Experiments 4 and 5), spatial configuration (Experiments 6 and 7), and the binding of contextual objects to locations (Experiment 8) were all sufficient to reset spatial parameters, as the same-position advantage was essentially eliminated in all of these cases.9 Not all contextual manipulations led to spatial remapping, however. In Experiment 9, the same-position advantage was preserved after array translation.

<sup>&</sup>lt;sup>8</sup> I thank an anonymous reviewer for suggesting this analysis.

<sup>&</sup>lt;sup>9</sup> The elimination of the same-position advantage with context deletion is similar to contextual sensitivity effects found in the phenomenon of inhibition of return (Klein & MacInnes, 1999; Müller & Von Mühlenen, 2000; Takeda & Yagi, 2000). Slowed orienting to previously attended locations is maintained only as long as the original scene or array context is present. If the scene or array is deleted, inhibition of return is "reset."

The present results are broadly consistent with recent research on human spatial memory. In human spatial memory studies, the standard method requires participants to remember which objects appeared where within a physical array of objects and later, without vision, specify the direction of individual objects from different viewing perspectives (e.g., Rieser, 1989). This method does not directly assess the binding of visual object representations to locations, as object memory could be supported by nonvisual codes, such as conceptual or verbal labels. However, it is certainly plausible that similar binding mechanisms are used in these spatial memory tasks as in the present study. Recent research on human spatial memory (Mou & McNamara, 2002; Shelton & McNamara, 2001) indicates that object-to-object spatial relationships are defined not within an egocentric coordinate frame but rather within a reference frame intrinsic to the external environment (e.g., organized along the salient axes of an object array). The preservation of the same-position advantage after array translation provides initial evidence that the binding of visual object representations to locations also occurs in an external reference frame, relative to the spatial structure of the larger object array. Converging evidence regarding the use of array-relative references frames could be obtained by manipulating the viewpoint in the present change detection task.

A few qualifications are necessary. First, the present results demonstrate that spatial position serves to structure object memory into coherent scene representations, but this does not rule out the possibility that additional, nonspatial binding mechanisms support episodic scene representations, such as direct object-to-object association. Second, the present data do not speak to the possibility that stimuli from other perceptual and cognitive systems (e.g., auditory, semantic, or motor codes) could also be bound within a multimodal representation of an environment. Third, the spatial configuration of local objects serves to organize visual memory, but there might be other forms of spatial context that also serve that function. In particular, large-scale geometric structure (i.e., the structure of an environment defined by large-scale elements such as walls, floors, etc.) might also provide a spatial context within which to bind memory for individual objects (e.g., Cheng, 1986). Finally, although visual memory performance was poorer when position changed, participants were still quite accurate in the different-position conditions. Access to memory for the visual form of objects is clearly not limited to the case in which position consistency is maintained. Such flexibility would allow objects to be recognized in new locations.

This study cannot address directly the neural bases of episodic scene memory, but the literature on brain mechanisms of scene perception and memory point to a plausible neural model of the object-position binding observed in this study. A wealth of evidence suggests that memory for the visual properties of complex objects is maintained in IT brain regions (for reviews, see Logothetis & Sheinberg, 1996; K. Tanaka, 1996). And an even larger body of evidence suggests that spatial memory representations of visual scenes are constructed in medial temporal brain regions (Aguirre, Zarahn, & D'Esposito, 1998; Burgess, Maguire, & O'Keefe, 2002; Chun & Phelps, 1999; Epstein & Kanwisher, 1998; O'Keefe & Nadel, 1978).

In rats, spatially selective hippocampal cells coding environmental locations ("place cells") appear to represent the animal's physical location in the environment (O'Keefe & Nadel, 1978). In primates, however, hippocampal and parahippocampal place cells code environmental locations in terms of gaze position (for a review, see Rolls, 1999). These "spatial view cells" respond when a particular scene position is fixated, and their response generalizes over differences in viewing direction and the absolute position of the monkey in the environment. In addition to activating medial temporal position codes, fixation of an object at a particular location leads to selective activation of IT representations coding the visual form of that object (Rolls, Aggelopoulos, & Zheng, 2003; Sheinberg & Logothetis, 2001). Because object fixation leads to simultaneous activation of visual object and position codes, the binding of the two can be accomplished efficiently through autoassociative mechanisms that have been proposed to support episodic binding in the hippocampus (Marr, 1971; Treves & Rolls, 1994).<sup>10</sup>

Support for the general idea that medial temporal regions enable object-position binding in scenes comes from evidence that lesions to the hippocampus (Parkinson, Murray, & Mishkin, 1988; Pigott & Milner, 1993) and to hippocampal output pathways (Gaffan, 1994) lead to specific deficits in learning of and memory for the positions of objects in scenes. In addition, medial temporal damage impairs learning of the association between individual object locations and spatial context (Chun & Phelps, 1999). Direct support for medial temporal coding of the association between spatial view activity and object memory comes from evidence of a population of neurons that respond to particular combinations of fixated location and object (Rolls, Xiang, & Franco, 2005). In this study, monkeys were trained on a task that required learning both the visual form of objects (e.g., triangle, square) and their computer screen positions. Task-responsive neurons in the hippocampus, parahippocampal gyrus, and perirhinal cortex were approximately evenly divided among those responding to object independently of location, those responding to fixated location independently of object, and those responding to a particular combination of object and location. Neurons of the last kind encode precisely the type of information necessary to generate spatially organized representations of complex scenes.

Further, the pattern of contextual effects in the present study is in accordance with that found in the animal navigation literature. Hippocampal place fields tend to reorganize after contextual changes that alter the spatial relationships among distal cues, including manipulations in which salient cues trade positions, a binding change (Shapiro, Tanila, & Eichenbaum, 1997). Similar remapping of place fields after changes in contextual spatial relationships would account for the elimination of the same-position advantage in the present Experiments 4–8. However, contextual changes that preserve the relative spatial relationships among cues, such as global rotation, tend to generate systematic adaptation of place fields rather than complete reorganization (O'Keefe & Conway, 1978). Similar preservation of place field organization following array translation (which does not alter the spatial relation-

<sup>&</sup>lt;sup>10</sup> This view is consistent with the idea that gaze fixation serves as an indexing mechanism in vision (Ballard, Hayhoe, Pook, & Rao, 1997). It is also possible that covert attention can serve to index locations and bind selective activation associated with the attended object to the attended location (Kahneman et al., 1992).

ships among contextual objects) would account for the preservation of the same-position advantage in Experiment 9.

Finally, the similar spatial effects observed when performance was likely to have been dependent on LTM (20-s study duration, Experiments 1 and 4) and when performance was likely to have been dependent on VSTM (400-ms study duration, Experiment 7) implies that STM and LTM systems depend on the same, or at least functionally similar, position-binding mechanisms. Recent work with hippocampal amnesic patients has found object-position binding deficits both over delays implicating LTM and over delays implicating STM (Hannula, Tranel, & Cohen, 2006; Olson, Page, Moore, Chatterjee, & Verfaellie, 2006), suggesting that spatial structure in VSTM and LTM both depend on medial temporal binding mechanisms.

#### Conclusions

The present study examined the manner in which memory for visual objects is structured to form a more comprehensive representation of a complex array or natural scene. The experiments focused on spatial structure in episodic visual memory and support the following conclusions: (a) Memory for the visual form of individual objects is stored as part of a more comprehensive representation of a scene or object array, (b) the visual memory representation of an object is associated with the remembered location where the object appeared, (c) object position memory is linked to the particular scene or array context in which the object was viewed, (d) object position is coded in a coordinate frame defined by the larger contextual layout of objects, (e) the spatial contextual representation encodes both the spatial configuration of objects and the binding of individual objects to locations, and (f) contextual changes that significantly disrupt object-to-object spatial relationships lead to a "resetting" of spatial parameters.

#### References

- Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A' and B" to below-chance performance. *Psychological Bulletin*, 102, 439–442.
- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). Neural components of topographical representation. *Proceedings of the National Academy of Sciences, USA*, 95, 839–846.
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15, 106–111.
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *The Bell System Technical Journal*, 40, 309–328.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral & Brain Sciences*, 20, 723–767.
- Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology*, 23, 393–419.
- Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35, 625–641.
- Carlson-Radvansky, L. A., & Irwin, D. E. (1995). Memory for structural information across eye movements. *Journal of Experimental Psychol*ogy: Learning, Memory, and Cognition, 21, 1441–1458.
- Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, 23, 149–178.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and

memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.

- Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic patients with hippocampal damage. *Nature Neuroscience*, 2, 844–847.
- Di Lollo, V. (1980). Temporal integration in visual memory. Journal of Experimental Psychology: General, 109, 75–97.
- Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, 8, 302–307.
- Epstein, R., & Kanwisher, N. (1998, April). A cortical representation of the local visual environment. *Nature*, 392, 598–601.
- Gaffan, D. (1994). Scene-specific memory for objects—A model of episodic memory impairment in monkeys with fornix transection. *Journal* of Cognitive Neuroscience, 6, 305–320.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424–429.
- Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The long and the short of it: Relational memory impairments in amnesia, even at short lags. *Journal of Neuroscience*, 26, 8352–8359.
- 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., & Hollingworth, A. (2003). Global transsaccadic change blindness during scene perception. *Psychological Science*, 14, 493–497.
- 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.
- Hollingworth, A. (2002). [Attention and scene memory]. Unpublished raw data.
- Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 388–403.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 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. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory,* and Cognition, 32, 58–69.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113–136.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23, 420–456.
- 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., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 343–360.
- Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene percep-

tion: 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.
- Jiang, Y., & Song, J.-H. (2005). Hyper-specificity in visual implicit learning: Learning of spatial layout is contingent on item identity. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1439–1448.
- 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.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. Annual Review of Neuroscience, 19, 577-621.
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London, Series B, 262, 23–81.*
- 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.
- Mou, W. M., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Mem*ory, and Cognition, 28, 162–170.
- Müller, H. J., & Von Mühlenen, A. (2000). Probing distractor inhibition in visual search: Inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1591–1605.
- Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science*, 11, 58.
- O'Keefe, J., & Conway, D. H. (1978). Hippocampal place units in the freely moving rat: Why they fire where they fire. *Experimental Brain Research*, *31*, 573–590.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* Oxford, England: Clarendon.
- Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., & Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *Journal of Neuroscience*, 26, 4596–4601.
- Parkinson, J. K., Murray, E. A., & Mishkin, M. (1988). A selective mnemonic role for the hippocampus in monkeys—Memory for the location of objects. *Journal of Neuroscience*, 8, 4159–4167.
- Pashler, H. (1988). Familiarity and the detection of change in visual displays. *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.
- Pigott, S., & Milner, B. (1993). Memory for different aspects of complex visual scenes after unilateral temporal- or frontal-lobe resection. *Neuropsychologia*, 31, 1–15.
- Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychol*ogy: General, 113, 426–442.
- Redish, A. D., & 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.
- Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory,* and Cognition, 15, 1157–1165.
- Rolls, E. T. (1999). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, 9, 467–480.
- Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. S. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience*, 23, 339–348.
- Rolls, E. T., Xiang, J. Z., & Franco, L. (2005). Object, space, and objectspace representations in the primate hippocampus. *Journal of Neurophysiology*, 94, 833–844.
- 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.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide* [Computer software manual]. Pittsburgh, PA: Psychology Software Tools.
- Shapiro, M. L., Tanila, H., & Eichenbaum, H. (1997). Cues that hippocampal place cells encode: Dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, 7, 624–642.
- Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience*, 21, 1340–1350.
- Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, 43, 274–310.
- 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).
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19, 73–74.
- Takeda, Y., & Yagi, A. (2000). Inhibitory tagging in visual search can be found if search stimuli remain visible. *Perception & Psychophysics*, 62, 927–934.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46(A), 225–245.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. Annual Review of Neuroscience, 19, 109–139.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 40(A), 201–237.
- Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4, 374–391.
- Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Perception & Psychophysics*, 67, 676–690.

## Appendix

	Percentage correct		Reaction time (ms)	
Condition	Target same	Target changed	Target same	Target changed
		Experiment 1		
Same target position	90.6	72.2	1,751	1,976
Different target position	78.5	73.6	2,129	2,255
		Experiment 2		
Same target position	90.5	55.0	1,194	1,299
Different target position	75.5	60.5	1,381	1,388
		Experiment 3		
Same target position	81.6	53.9	1,278	1,302
Different target position	71.6	59.2	1,352	1,393
		Experiment 4		
Background present				
Same target position	88.3	78.1	2,021	2,117
Different target position	80.2	72.9	2,616	2,705
Background absent				
Same target position	85.1 84.4	72.9 75 7	1,785 1,954	1,989 1 923
	01.1		1,551	1,723
		Experiment 5		
Background present				
Same target position	87.9	66.3	1,309	1,429
Different target position	77.3	62.1	1,593	1,557
Background absent				
Same target position	85.4	65.4	1,199	1,290
Different target position	80.6	65.2	1,269	1,323
		Experiment 6		
Background same				
Same target position	87.3	71.2	1,270	1,323
Different target position	76.2	63.8	1,426	1,444
Background scrambled				
Same target position	77.1	63.1	1,398	1,508
Different target position	74.6	62.5	1,514	1,508
		Experiment 7		
Background same				
Same target position	82.5	56.3	1,106	1,129
Different target position	68.5	60.2	1,233	1,258
Background scrambled	76.2	51 5	1 154	1 220
Same target position	/0.5	50.2	1,154	1,230
	08.8	59.2	1,229	1,244
		Experiment 8		
Background same				
Same target position	92.1	66.0	1,334	1,475
Different target position	77.5	64.6	1,613	1,713
Background binding change	07.0	<b>57</b> 0	1.120	
Same target position	87.9 81.7	57.9	1,420	1,555
		Experiment 0	1,000	
A 11	00.0		1 205	1 105
All same	90.8	77.5	1,385	1,435
Larget Shift Only Reakground shift only	84./	/3.1	1,525	1,580
All shift	04.9 88 6	71.0	1,304	1,007
All sillit	00.0	/1.9	1,329	1,342

## Mean Percentage Correct and Reaction Time Data for Experiments 1-9

Note. Reaction time data include all trials (correct and incorrect).

Received January 11, 2006 Revision received June 15, 2006 Accepted July 25, 2006