The Relationship between Online Visual Representation of a Scene and Long-term Scene Memory

Andrew Hollingworth

Department of Psychology, The University of Iowa

Four experiments investigated the relationship between the online visual representation of previously attended objects in natural scenes and long-term visual memory. A change detection task was used, in which a target object either was changed or remained the same from an initial image of a natural scene to a test image. Two types of changes were possible: rotation in depth, or replacement by another object from the same basic-level category. To measure online visual memory, the test scene appeared immediately after the initial scene. To measure long-term visual memory, the test scene was delayed either one trial (Experiment 1 and 2) or until after all scenes had been viewed (Experiments 3 and 4). There was no decrement in change detection performance from the immediate test condition to the test delayed one trial, demonstrating that the online representation of previously attended objects is of similar specificity to object representations maintained in LTM. There was only a small decrement in change detection performance with longer delays in Experiments 3 and 4, demonstrating remarkably robust LTM for visual objects.

How do people construct visual representations of the environments around them? Natural scenes, such as an office or playground, are highly complex, often containing scores of individual, discrete objects. All of the visual detail within a scene cannot be perceived in a single glance, so the visual system sequentially selects individual objects by movements of the eyes and attention (see Henderson & Hollingworth, 1998). As a result, visual scene perception is extended over time and space as the eyes and attention are oriented from object-to-object. If a visual representation of the scene as a whole is to be constructed, information from previously attended objects must be retained in memory and integrated with information from subsequently attended objects.

The role of visual memory in the representation of complex scenes has received a great deal of attention over the last 10 years. Evidence from the phenomenon of change blindness has been interpreted as suggesting that visual memory may be fleeting and thus that scene representations are visually impoverished (O’Regan, 1992; O’Regan & Nöe, in press; O’Regan, Rensink, & Clark, 1999; Rensink, 2000; Rensink, O’Regan, & Clark, 1997; Simons, 1996; Simons & Levin, 1997; Wolfe, 1999). In change blindness demonstrations, a change is introduced into a scene during some form of visual disruption, such as a saccadic eye movement (e.g., Grimes, 1996; Henderson & Hollingworth, in press-b) or brief ISI (e.g., Rensink et al., 1997). As a result, the change cannot be detected directly; change detection depends on the retention of information from the initial image and comparison of that representation to perceptual information in the changed image. Participants’ surprisingly poor change detection performance in these paradigms has been taken to indicate that the visual memory representation of a scene is impoverished, limited to the currently attended object (Rensink, 2000) or to the small number of objects (three or four) that can be maintained in visual short-term memory (VSTM) (Irwin & Andrews, 1996; Irwin & Zelinsky, 2002). According to these visual transience hypotheses, scene representation are visually impoverished because visual object representations disintegrate almost immediately after the withdrawal of attention from an object.

The hypothesis that visual memory is transient is difficult to reconcile with evidence from research on LTM for pictorial stimuli. A series of studies in 1960s and 1970s demonstrated that people can remember multiple thousands of different pictures (Nickerson, 1965; Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). In the most ambitious of these studies, Standing (1973) presented each participant with 10,000 different pictures of various subject matter for 5 s each with a 660 ms ISI (which required more that 15 hours of study over the
course of 5 days). On a two-alternative, forced-choice recognition test conducted after the fifth study session, mean percentage correct was 83%. Accounting for guessing, 83% correct suggests the retention of approximately 6,600 pictures. This and similar demonstrations of capacity show that some form of representation is retained from a picture that is sufficient to discriminate it from a picture that was not studied. The distractors used in these experiments, however, were typically chosen to be maximally different from studied images, making it difficult to identify the type of information supporting such feats of memory. Participants may have remembered studied pictures by maintaining visual representations (coding visual properties such as shape, color, texture, and so on), by maintaining conceptual representations of picture identity, or by maintaining verbal descriptions of picture content.

Evidence that picture memory depends, at least in part, on specifically visual memory comes from a study by Standing et al. (1970). After viewing 120 slides of complex images for 1 s each followed by a 30 min delay, participants could correctly identify the left-right orientation of studied pictures at a rate of approximately 86% correct. Because mirror reversal does not change the identity of a picture, such performance could not have been supported by a conceptual coding of picture identity. In addition, it is very unlikely that a brief verbal description would have been sufficient to code picture orientation, especially since participants had only 1 s to view the scene and were not aware at study that picture orientation would be tested. Although the Standing et al. study provides evidence for some form of robust visual memory, it does not provide direct evidence that the visual properties of individual objects in a picture are retained robustly; global orientation could have been remembered as an abstract coding of the configuration of individual elements in the picture (Simons, 1996). Evidence of robust visual memory for individual objects comes from a study by Goldstein and Chance (1970) and from the literature on visual object recognition. Goldstein and Chance found above-chance recognition performance for highly homogenous sets of novel stimuli that could not be easily described verbally, such as snowflakes and inkblots. In addition, studies of visual object recognition demonstrate orientation specificity in LTM for novel objects (e.g., Tarr, Bülthoff, Zabinski, & Blanz, 1997).

To investigate the apparent discrepancy between claims of visual transience and evidence of robust long-term visual memory, Hollingworth & Henderson (2002) tested whether visual object representations in natural scenes are retained and accumulated as the eyes and attention are oriented from object-to-object within the scene. Eye movements were monitored, and the computer waited until the subject had fixated a target object in the scene. During a subsequent saccadic eye movement to a different object in the scene, the target either was changed or remained the same. The change was either a 90° rotation of the target in depth or the replacement of the target with another object from the same basic-level category (token change). Because attention precedes the eyes to the goal of an eye movement (e.g., Hoffman & Subramaniam, 1995), the target object was not attended when the change occurred. Yet, participants detected these subtle changes to previously attended objects at a rate significantly above the false alarm rate, demonstrating that visual representations do not necessarily disintegrate after the withdrawal of attention (see also Hollingworth, Williams, & Henderson, 2001). In addition, for scenes in which the target was not changed during initial viewing, a delayed, two-alternative forced-choice test was administered after all scenes had been viewed. Participants were surprisingly accurate on this delayed test, performing above 80% correct both when discriminating the target from a rotated distractor and when discriminating the target for a different token distractor. These data demonstrate that visual object representations are robust not only during the online perception of a scene but also over the longer time-scales tested in the picture memory literature. Clearly, visual memory is not transient.1

These results lead to the issue of central interest in the present study: What is the relationship between the visual scene representation constructed during online viewing and the scene representation maintained in LTM? Hollingworth and Henderson (2002, Experiment 3) examined online memory for target objects as a function of the number of fixations intervening between fixation on the target and a forced-choice test. Discrimination performance remained accurate even when nine or more fixations on other objects intervened. Capacity constraints in VSTM make it unlikely that VSTM supported this ability. Hollingworth and Henderson therefore proposed that LTM may play a significant role in the online visual representation of natural scenes. According to this visual memory theory of scene
representation, as the eyes and attention are oriented from object-to-object within a scene, higher-level visual representations (abstracted away from precise sensory information) are formed for attended objects. The higher-level visual representation of an object is initially maintained in VSTM and is bound to a position within a spatial representation of the scene, forming an object file (Kahneman, Treisman, & Gibbs, 1992). This representation is then consolidated into LTM. As new objects are attended and fixated, the object representation in VSTM is replaced by new object representations. However, the LTM representation is retained robustly and accumulates with visual representations from other previously attended objects. Thus, over the time-course of scene viewing, LTM supports the construction of a relatively detailed visual representation of the scene.

If the online representation of previously attended objects is primarily supported by LTM, then memory for previously attended objects during online scene viewing should have similar characteristics to object memory under conditions that unambiguously require LTM. Hollingworth & Henderson (2002) provided initial evidence that online and LTM scene representations may retain object information of similar specificity. Forced-choice discrimination performance was only moderately higher when the test was administered during online scene viewing compared with when the test was administered after all scenes had been viewed, when performance must have been based on LTM (Hollingworth & Henderson, 2002, Experiments 1 and 2). However, there were numerous methodological differences between these experiments, making any direct comparison difficult. In particular, participants had almost twice as long to study the scene for the test delayed until the end of the session compared with the test administered online. Thus, these results can only be taken as suggestive of similar specificity in online object representation and long-term object memory.

**The Present Study**

The present study sought to investigate the relationship between online and LTM scene representations by directly comparing memory for previously attended objects during online scene viewing with object memory under conditions that unambiguously require LTM. Evidence of similar specificity in online and LTM representations would provide support for the visual memory theory claim that the online representation of previously attended objects is primarily supported by LTM.

A change detection method was used. The basic paradigm was based on that of Hollingworth (in press). The sequence of events in a trial is illustrated in Figure 1. An initial image of a 3-D scene was presented for 20 s. The initial scene presentation was chosen to be long enough to ensure that the target would be fixated and attended before the change on the vast majority of trials (in Hollingworth & Henderson, 2002, using a similar set of stimuli, the target object was fixated within 20 s of viewing on 98% of trials). The initial scene presentation was followed by an abrupt dot onset for 150 ms, initial scene again for 200 ms, pattern mask for 200 ms, and finally a test scene, which remained visible until response.

<< Insert Figure 1 about here >>

The principal manipulation in the present study involved when the test scene was displayed. To investigate memory for previously attended objects during the online perception of the scene, in the immediate test condition, the test scene was presented immediately after the pattern mask, as illustrated in Figure 1. Although a short delay is introduced in this condition (the 200 ms mask), this delay is only the duration of a typical blink. The phenomenology of this condition is of a single presentation of the scene, interrupted briefly by the mask. To investigate object memory under conditions that unambiguously require

---

2 Visual memory appears to be composed of four different memory stores: visible persistence, informational persistence, VSTM, and VLTM (see Irwin, 1992, for an excellent review). Visible and informational persistence (Coltheart, 1980), often grouped together as “iconic memory” (Neisser, 1967) or “sensory persistence”, preserve a precise, high-capacity, point-by-point, sensory trace that decays very quickly and is maskable. Visible persistence, as the name suggests, is phenomenologically available and decays within approximately 130 ms after stimulus onset (Di Lollo, 1980). Informational persistence is a non-visible trace that persists for approximately 150-300 ms after stimulus offset (Irwin & Yeomans, 1986). These forms of sensory persistence do not survive an eye movement and thus could not be used to accumulate visual information as the eyes and attention are oriented from object-to-object within a scene (Henderson & Hollingworth, in press-b; Irwin, 1991; Irwin, Yantis, & Jonides, 1983). They will not be considered further. VSTM maintains visual representations abstracted away from precise sensory information. It has a limited capacity of 3-4 objects (e.g., Luck & Vogel, 1997) and less spatial precision than point-by-point sensory persistence (Phillips, 1974). However, VSTM is not disrupted by masking and can be maintained over fairly long durations (at least 5 s) (Phillips, 1974) and across saccades (see Irwin, 1992). VLTM appears to maintain visual representations similar to those maintained in VSTM (see General Discussion) but with remarkably large capacity and robust storage, as demonstrated in Experiments 3 and 4. I use the term “higher-level visual representation” to refer to the type of abstracted visual representation maintained in VSTM and VLTM.
LTM, the test scene was delayed, either by one trial (Experiments 1 and 2) or until after all scene items had been displayed (Experiments 3 and 4).

In the test scene, the target was either the same as the original version or changed in one of two ways: a rotation, in which the target was rotated 90° in depth; or a token change, in which the target was replaced by another object from the same basic-level category. The target object was specified in the test scene by an arrow cue to ensure that participants could limit retrieval and comparison processes to the target (see Hollingworth, in press). The dot onset and offset served to draw attention away from the target object immediately before the change (Hollingworth, in press), on the assumption that an abrupt object onset will capture attention (e.g., Jonides, 1981; Yantis & Jonides, 1984). Thus, the change detection test in the immediate test condition probed memory for an object that was not focaly attended when the change occurred.

To preview the results of these experiments, change detection performance with a one-trial delay was unreduced from that found in the immediate test condition, and performance was only slightly reduced when the test was delayed until the end of the session. These results demonstrate that the online representation of previously attended objects is of similar specificity to long-term object memory. In addition, they confirm the findings of the picture memory literature suggesting that visual memory has quite remarkably large capacity.

This study also investigated a secondary issue: the extent to which robust visual object memory is supported by strategic encoding of object properties. To examine potential strategic verbal encoding, Experiment 2 added a verbal working memory load to the basic paradigm. Consistent with prior findings (Hollingworth, in press), verbal encoding appears to play little or no role in object memory in this paradigm. To investigate differences in visual memory as a function of task demands and participant strategy, Experiment 4 either maintained the same change detection task from immediate to delayed test or switched tasks. Task switching produced a reliable decrement in change detection performance, demonstrating that the contents of visual memory are sensitive to task demands and viewer goals.

**Experiment 1**

As an initial comparison of online and long-term scene representation, the change detection test was administered either directly after scene viewing (immediate test condition), or it was administered after a delay of one trial (delayed test condition). A delay of one trial introduced the following events between scene study and test of that scene: 1) change detection test for the previous scene and 2) presentation of a different, intervening scene for 20 s plus dot onset, offset, and mask. These intervening events between study and test in the delayed condition are sufficient to ensure that participants were no longer maintaining the target object in VSTM, given current VSTM capacity estimates of three or four objects and the fact that the each scene stimulus contained many more than four individual objects. Thus, this experiment compares the specificity of online visual representations of previously attended objects (in the immediate test condition) with the specificity of visual object representations maintained in LTM (in the delayed test condition). If, as held by visual memory theory, the online visual representation of previously attended objects is supported to a significant degree by LTM, then there may be little difference in the specificity of object representations maintained online and in LTM and therefore little decrement in change detection performance from that in the immediate test condition to that in the delayed test condition.

**Method**

**Participants.** Twenty-four participants from the University of Iowa community completed the experiment. They either received course credit in introductory psychology or were paid. All participants reported normal or corrected-to-normal vision.

**Stimuli.** Forty-two scene images were created from 3-D models of real-world environments. A target object was chosen within each model. To produce the rotation and token-change images, the target object was either rotated 90° in depth or replaced by another object from the same basic-level category. The objects for token changes were chosen to be approximately the same size as the initial target object. Scene images subtended 16.9° x 22.8° visual angle at a viewing distance of 80 cm. Target objects subtended 3.33° on average along the longest dimension in the picture plane. The onset dot was a neon-green disk with a diameter of 1.15°. The post cue was a neon-green arrow subtending 2.22° in length, and it pointed unambiguously to the target in all test images. The mask was made up of a patchwork of small colored shapes, and was the same size as the scene stimuli.

**Apparatus.** The stimuli were displayed at a resolution of 800 by 600 pixels by 24-bit color on a 17-in. Sony Trinitron monitor at a refresh rate of 100
Hz. The initiation of image presentation was synchronized to the monitor’s vertical refresh. Responses were collected using a serial button box. The presentation of stimuli and collection of responses was controlled by E-Prime software running on a Pentium IV-based computer. Viewing distance was maintained at 80 cm by a forehead rest. The room was dimly illuminated by a low-intensity light source.

Procedure. Participants were tested individually. Each participant was given a written description of the experiment along with a set of instructions. Participants were informed that they would view a series of scene images and would have to determine whether a change had been introduced for each. The nature of the possible changes was described. In addition, participants were informed that the change detection test would be administered either immediately after the scene presentation or after a delay of one trial.

Participants pressed a pacing button to initiate each trial. Then, a white fixation cross on a gray field was displayed for 1000 ms. This was followed by the scene presentation for 20 s, dot onset within the scene for 150 ms, initial scene again for 200 ms, and a pattern mask for 200 ms. In the immediate test condition, the pattern mask was followed immediately by the test image, which remained visible until response. Participants pressed one of two buttons to indicate that the object specified by the arrow either had changed or was the same. In the delayed test condition, the mask was followed by a gray screen with the message “prepare for test of previous scene” displayed for 1500 ms. This was followed by the test image for the scene item displayed one trial earlier. Participants responded in the same manner as in the immediate test condition.

Participants completed two blocks of trials, an immediate test block and a delayed test block. Block order was manipulated between participant groups. Each block began with six practice trials, two in each of the three change conditions. The practice scenes were not used in the experimental session. The practice trials were followed by 21 experimental trials, seven in each of the three change conditions (same, rotation, token change). In the delayed test block, an extra scene presentation was added to the end of the block so that the test of the final experimental item could be delayed one trial. Thus, participants viewed a total of 55 different scenes: 12 practice items, 42 experimental items, and one dummy item at the end of the delayed test block. Across participants, each of the 42 experimental items appeared in each condition an equal number of times. The entire experiment lasted approximately 50 minutes.

Results

Percentage correct data were used to calculate \( A' \), a nonparametric signal detection measure with a functional range of \(.5 \) (chance) to \(1.0\) (perfect sensitivity) (Grier, 1971). For each participant in each change condition, \( A' \) was calculated using the mean hit rate when the target changed and the mean false alarm rate when it did not. For example, \( A' \) for token changes in the immediate test condition was calculated using the participant’s hit rate in the immediate token-change condition and the false alarm rate in the immediate same condition.\(^3\) Because \( A' \) corrects for potential differences in response bias in the percentage correct data, it forms the primary data for interpreting these experiments. Raw percentage correct data are reported in the Appendix.

Because block order did not produce a reliable main effect, nor did it interact with the change type or test delay variables, the following analysis collapsed across block order. Mean \( A' \) in each of the test delay and change conditions is displayed in Figure 2. The test delay factor did not produce a reliable main effect, \( F(1, 23) = 1.04, p = .32 \). Mean \( A' \) was .868 in the immediate test condition and .897 in the delayed test condition. The absence of a difference between immediate and delayed tests is unlikely to have been caused by a ceiling effect, because overall percentage correct performance was well below ceiling (81.0%). The main effect of change type was not reliable \((F < 1)\) nor was the interaction between test delay and change type \((F < 1)\).

Discussion

Overall, change detection performance was very good, consistent with previous experiments in which the target object was specified in the test scene

\[^3\text{For above-chance performance, } A' \text{ was calculated as specified by Grier (1971): } \]

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

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

\[
A' = \frac{1}{2} \frac{(x-y)(1+x-y)}{4x(1-y)}
\]
(Hollingworth, in press). Because the target object was no longer attended when the change occurred in the immediate test condition, these results provide further evidence that visual object representations do not necessarily disintegrate immediately after the withdrawal of attention. In addition, because change detection performance remained accurate with a one-trial delay, these data demonstrate that visual scene memory is not limited to the contents of VSTM.

The principal issue in this experiment was the relationship between online scene representation and long-term scene memory. Change detection performance was no worse when the test was delayed one trial compared with when the test was administered immediately after scene viewing. In fact, the numerical trend was in the direction of higher change detection performance on the delayed test. Thus, Experiment 1 demonstrates that no token- or orientation-specific object information is lost from the representation of a scene maintained during viewing to the representation maintained after one has subsequently inspected an entirely different scene for 20 s. Consistent with the visual memory theory claim that online scene representation is supported by LTM, long-term object representations appear to be just a specific as online representations of previously attended objects, at least with respect to the sorts of visual information relevant to detecting changes of orientation and token.

Performance in this experiment was almost certainly based on the retention of visual object representations in memory, as opposed to conceptual representations or verbal descriptions. First, token substitution does not alter the basic-level conceptual identity of the target, and rotations do not change the identity of the target at all. Thus, conceptual representations of object identity could not support accurate change detection in this paradigm. To examine the potential contribution of verbal description to change detection performance, Hollingworth (in press) added a four-digit verbal working memory load and articulatory suppression to the change detection task used in the immediate test condition of Experiment 1. Performance was not reduced compared to the standard, no verbal load condition, demonstrating that change detection performance was based on visual, not verbal, memory. However, it is possible that verbal encoding may play a larger role in scene memory across longer delays, such as those introduced by the delayed test condition in Experiment 1. Thus, Experiment 2 replicated the design of Experiment 1 but with the addition of a four-digit verbal working memory load and articulatory suppression.

**Experiment 2**

**Method**

**Participants.** Twenty-four new participants from the University of Iowa community completed the experiment. They either received course credit in introductory psychology or were paid. All participants reported normal or corrected-to-normal vision.

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

**Procedure.** The procedure was identical to Experiment 1 except for the addition of a four-digit working memory load. On each trial, the initial screen instructing participants to “press a button to being the next trial” also contained four randomly chosen digits. Before initiating the trial, the participant began repeating the series of digits aloud and continued digit repetition until the appearance of the test scene. Participants were instructed to repeat the digits without interruption or pause, and the experimenter monitored digit repetition to ensure that participants complied.

**Results and Discussion**

**A’ Analysis.** Mean A’ in each of the test delay and change conditions is displayed in Figure 3. Block order did not produce a reliable main effect, nor did it interact with change type, but it did interact with test delay, $F(1, 23) = 12.18, p < .005$. This interaction was produced by the fact that participants performed more accurately in the second block of trials ($A’ = .879$) compared with the first ($A’ = .821$). The same trend was present in the Experiment 1 data, but the effect was not reliable. In all other respects, however, the pattern of results was very similar to that in Experiment 1. The test delay manipulation did not produce a reliable main effect ($F < 1$), with $A’$ of .840 in the immediate test condition and .860 in the delayed test condition. The main effect of change type was not reliable ($F < 1$) nor was the interaction between test delay and change type ($F < 1$). To examine performance with and without verbal load, the data from Experiments 1 and 2 were combined in an omnibus analysis, with experiment treated as a between-subjects factor. The effect of verbal load was not reliable, $F(1, 44) = 1.57, p = .21$, nor was the effect of test delay, $F(1, 44) = 2.23, p = .14$.

<< Insert Figure 3 about here >>

Although there was a slight numerical drop in overall performance from Experiment 1 to Experiment 2, this difference was not statistically reliable. More importantly, Experiment 2 replicated the pattern of results found in Experiment 1, with no
decline in change sensitivity with a test delay of one trial. Thus, the Experiment 2 data demonstrate that the robust retention effect is supported by visual, not verbal, memory.

**Experiment 3**

In Experiment 3, to provide a more demanding test of visual memory for objects in scenes, the delay between scene viewing and test was increased substantially. Participants viewed all 42 scene stimuli initially. For half of these, the test image was displayed immediately after viewing (immediate test condition). For the other half, the test was delayed until after all 42 stimuli had been viewed (delayed test condition). In the latter condition, successful change detection depends on retaining object information over multiple minutes of delay and numerous intervening scenes, as in the literature on picture memory (e.g., Standing, 1973; Standing et al., 1970). Because the presence of a verbal working memory load had no observable effect on change detection performance between Experiments 1 and 2, Experiment 3 used the standard, no load method, as in Experiment 1.

**Method**

Participants. Twenty-four new participants from the University of Iowa community completed the experiment. They either received course credit in introductory psychology or were paid. All participants reported normal or corrected-to-normal vision.

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

Procedure. Participants were instructed that they would first view a set of images of natural scenes. For half of these, they would be asked to determine whether an object had changed immediately after scene viewing; for the other half, this test would be delayed until after all the scenes had been viewed.

There were two sessions in this experiment. In the initial session, all 42 scenes were displayed, and half of these were tested immediately after viewing. The initial session was followed by a delayed test session, in which the change detection test was administered for the scenes not tested in the initial session. Before beginning the initial session, participants completed six practice trials. All practice trials presented the test scene immediately after scene viewing, with two trials in each of the three change conditions. Practice scenes were not used in the experimental sessions. Participants then completed 42 experimental trials in the initial session, 21 in the immediate test condition and 21 in the delayed test condition, randomly intermixed. The sequence of events in a trial in the immediate test condition was identical to that in Experiment 1. For the delayed test condition in the initial session, the trial ended after the mask (i.e., the test scene was not displayed). After all 42 scenes had been presented in this manner, participants completed the delayed test session, in which each of the 21 test scenes for the delayed condition were displayed, and participants made a same-changed response to each. Thus, participants saw the same set of stimuli in the two test-delay conditions, but in the delayed test condition, the final test image was not displayed until after all scenes had been viewed initially. Approximately 30 s elapsed between the end of the initial session and the beginning of the delayed test session.

The order of scene presentation in the initial session and the delayed test session were determined randomly, with the condition that the last item in the initial session could not be the first item in the delayed test session. The random ordering of trials produced a range of delays for the delayed test. The mean temporal delay between scene viewing and test was 9.2 min, with a range of 32 s to 20 min. Delay can also be expressed as the number of scenes intervening between initial viewing and test. Considering all intervening scenes (including those in the delayed test session itself), the mean number of intervening scenes was 30.5, with a range of 2 to 61.

**Results**

Mean \( A' \) in each of the test delay and change conditions is displayed in Figure 4. The test delay manipulation did not produce a reliable main effect, \( F(1, 23) = 1.41, p = .25 \). Mean \( A' \) was .853 in the immediate test condition and .823 in the delayed test condition. The main effect of change type was not reliable \( (F < 1) \), but there was a reliable interaction between test delay and change type, \( F(1, 23) = 4.53, p < .05 \). Examining each of the contrasts, there was no effect of test delay in the token change condition \((F < 1)\), but there was a reliable advantage for the immediate test in the rotation condition, \( F(1, 23) = 4.44, p < .05 \).

<< Insert Figure 4 about here >>

The variable delay in the delayed test condition allowed examination of percentage correct change detection performance as a function of the number of scenes intervening between initial viewing and test. The number of intervening scenes (including those in the delayed test session) was regressed against the dichotomous change detection variable, yielding a point-biserial correlation coefficient. Each trial was treated as an observation. Since each
participant contributed more than one sample to the analysis, variation caused by differences in participant means was removed by including participant as a categorical factor (implemented as dummy variables) in the model. Separate regression analyses were conducted for token change and rotation. Same trials were included in both analyses to minimize the possibility that changes in response bias would yield spurious correlations. The results of these analyses are displayed in Figure 5. Neither analysis produced a reliable correlation [token: $r_{pb} = - .08$, $t(479) = - 1.40$, $p = .16$; rotation: $r_{pb} = - .04$, $t(479) = -.78$, $p = .43$], although the trends indicated a very slight negative relationship.

Discussion

In Experiment 3, robust visual memory was observed over a much longer retention interval and across multiple different intervening scenes. These data demonstrate a quite remarkable memory capacity for visual stimuli, consistent with the literature on picture memory (e.g., Standing, 1973; Standing et al., 1970). In addition, the present paradigm provided a much more stringent and targeted test of memory compared with that earlier literature. The stimuli were a set of 3-D rendered images depicting fairly similar environments (e.g., different rooms within a house), and the change detection test required retention of the visual attributes of a single object in each scene.

Experiment 3 did provide the first evidence of forgetting, however, with a significant drop in performance from immediate to delayed test in the rotation condition. Token change detection performance was not reduced at all with test delay. These data suggest that information specific to orientation may be slightly less robust than information specific to object token, perhaps because differences between object tokens can be coded on more dimensions (e.g., differences in shape, color, texture, and so on), and thus are less susceptible to interference from subsequent objects and scenes. However, performance for delayed rotations remained well above chance.

In addition, Experiment 3 provided another opportunity to compare change detection performance without a verbal memory load to performance with a verbal memory load. A’ performance in the immediate conditions of Experiment 3, which did not employ a verbal load, was essentially equivalent to that in Experiment 2, which did. Thus, Experiment 3 provides further evidence that verbal encoding plays little or no role in this paradigm.

Experiment 4

Experiments 1-3 demonstrate that visual object memory is highly robust when participants intend to maximize memory performance for token and orientation change detection. In Experiment 4, I sought to investigate the extent to which visual object memory in the present paradigm is under strategic control. The basic method from Experiment 3 was used, with the following modifications. As in Experiment 3, half the experimental items were tested immediately after viewing, but participants completed only one type of change detection on these trials (either rotation or token change). Again, as in Experiment 3, the other half of the items were tested in a delayed test session after all stimuli had been viewed. In the delayed test session, participants either continued the change detection task from the initial session (same test condition) or switched to the other form of change detection (different test condition). Specifically, there were four groups of participants, two in the same test condition (rotation initial, rotation delayed; token initial, token delayed) and two in the different test condition (token initial, rotation delayed; rotation initial, token delayed). Participants were not informed of the delayed test until after completing the initial session, so when participants encoded scene information in the initial session, they did not know what information would be relevant on the delayed test (or even that there would be a delayed test). Thus, performance on the delayed test in the different test condition provides a measure of object memory for a type of test that was not anticipated at encoding. In contrast, for the same test participants, the type of information relevant during the initial session remained relevant in the delayed session.

One theoretical possibility is that simply attending to an object leads to a visual object representation that is maximally specific (to the informational limits on visual memory representation). This hypothesis could be generated from evidence that short-term visual memory capacity for objects is limited not by the number of features encoded for each object but by the total number of objects (Luck & Vogel, 1997). If the encoding of additional object information above that necessary for the current task does not consume additional memory capacity, then all available information may be encoded, regardless of task or strategy. If so, then performance on the delayed test should be independent of the type of change detection task performed during the initial session.

Alternatively, participants may be able to
selectively encode object information relevant to the immediate task, excluding irrelevant information from the memory representation. If so, performance on the delayed test should be lower if the test is different from that performed during the initial session. It is unlikely, however, that performance in the delayed session for the different test condition would fall to chance. The present stimuli were not designed to perfectly isolate token and orientation information. For example, in-depth rotation alters the 2-D shape properties of the object, and differences in shape are also relevant for detecting changes in object token. Thus, even if participants were able to encode only those object attributes necessary to detect rotations, they would likely retain at least some capability to detect token changes, and vice versa.

Method

Participants. Thirty-two participants from the Yale University community completed the experiment, eight in each participant group. They either received course credit in introductory psychology or were paid. All participants reported normal or corrected-to-normal vision.

Stimuli and Apparatus. The stimuli and apparatus were the same as in Experiment 1, except that two of the 42 experimental items were eliminated to enable efficient counterbalancing across four participant groups.

Procedure. Participants were instructed that they would first view a set of images of natural scenes. For half of these, they would be asked to determine whether an object had changed immediately after scene viewing; for the other half, the trial would simply end after scene viewing. Participants first completed four practice trials. Each practice trial presented the test scene immediately after viewing, with two trials in each of the two change conditions. The practice scenes were not used in the experimental session. Participants then completed 40 trials in the initial session, 20 in the immediate test condition and 20 in the delayed test condition, randomly intermixed. For the immediate test trials, two of the four participant groups completed rotation change detection, and the other two completed token change detection. Trials were evenly divided between same and changed conditions. The immediate test trials were identical to those in Experiments 1-3.

The delayed test trials in the initial session differed slightly from those in Experiment 3. The trial ended after the 20 s scene presentation, without presentation of the 150 ms dot onset, 200 ms representation of the scene, or 200 ms mask. If this difference were to have any effect, it would be to reduce performance on the delayed test, because of reduced encoding time, but the results of Experiment 4, discussed below, provide no cause for concern that delayed test performance was reduced by this difference. The initial session ended with a dummy item that was not tested initially and thus could be used to illustrate the type of change detection required in the delayed test session.

After completing the initial session, participants were informed that they would now be tested on the items that were not tested in the initial session. Two of the four groups completed the same type of change detection task as in the initial session (same test condition), and the other two groups switched to the other form of change detection (different test condition). For the latter condition, the nature of the new change detection task was described. In the delayed test session, participants first completed a practice change detection item that corresponded to the final, dummy item in the initial session. The practice item presented the changed version of the target object and was included primarily to provide an example of the new type of change for participants in the different test condition. This was followed by 20 experimental trials, corresponding to the 20 items not tested in the initial session. Each trial presented the test scene until response, and participants responded to indicate whether the object specified by the arrow had or had not changed from the version viewed initially. Ten items appeared in each of the two change conditions (same and changed). Item order was randomly determined. Approximately 1 min elapsed between the end of the initial session and the beginning of the delayed test session. Across participants, each of the 40 experimental items appeared in each condition an equal number of times. The entire experiment lasted approximately 50 minutes.

The order of scene presentation in the initial session and in the delayed test session was determined randomly. The random ordering of trials produced a range of delays. The mean temporal delay between scene viewing and test was 11.0 min, with a range of 2.2 min to 22.5 min. Considering all intervening scenes (including those in the delayed test session itself), the mean number of intervening scenes was 30.7, with a range of 2 to 58.

Results

Mean A’ performance for each participant group is displayed in Figure 6. To establish that there were no systematic differences between participant groups, performance on the immediate test (displayed
in the white bars of Figure 6) was analyzed in a 2 (same test, different test) x 2 (change type: rotation, token) between-subjects ANOVA. No main effect of same/different test was observed ($F < 1$), as should have been the case, since the same/different test factor was a manipulation of the delayed test and not the immediate test. In addition, there was no effect of change type ($F < 1$) and no interaction between same/different test and change type ($F < 1$). Thus, there were no observable differences in the baseline performance of the four participant groups.

The delayed test data provide the principal evidence to test the hypothesis that object memory is independent of task demands and observer strategy. These data (displayed in the black bars of Figure 6) were analyzed in a 2 (same test, different test) x 2 (change type: rotation, token) between-subjects ANOVA. Contrary to the prediction of the hypothesis in question, there was a reliable main effect of same/different test, $F(1, 28) = 4.53, p < .05$, with superior change sensitivity when the delayed test was preceded by the same type of test in the initial session ($A' = .880$) than when the delayed test was preceded by a different type of test in the initial session ($A' = .787$). The change type factor did not produce a reliable main effect ($F < 1$) nor did it interact with the same/different test factor, $F(1, 28) = 1.57, p = .22$.

A full analysis of the entire data set revealed a reliable interaction between same/different test and test delay, $F(1, 28) = 6.58, p < .05$. The drop in performance from immediate to delayed test was larger in the different test condition ($A' = .130$) than in the same test condition ($A' = .012$). Examining just the same test condition (left four bars of Figure 6), there was no effect of test delay ($F < 1$), no effect of change type, $F(1, 14) = 1.33, p = .27$, and no interaction ($F < 1$). Note the similarity of the results to those from Experiment 3, with a small numerical drop in performance with test delay for rotations and no drop in performance with delay for token changes. Examining just the different test condition (right four bars of Figure 6), there was a reliable main effect of test delay, $F(1, 14) = 11.38, p < .005$, with mean $A'$ of .917 on the immediate test and .787 on the delayed test. The effect of change type was not reliable ($F < 1$) nor was the interaction between change type and test delay, $F(1, 14) = 1.24, p = .28$.

As in Experiment 3, the variable delay in the delayed test condition allowed examination of change detection performance in the delayed test session as a function of the number of scenes intervening between initial viewing and test. Separate regression analyses were performed for each of the four participant groups. Both same and change trials were included in the analyses. The results of these analyses are displayed in Figure 7. Only rotation change detection in the same test condition produced a reliable correlation [rotation (following rotation): $r_{pb} = -.20$, $t(151) = -2.52, p < .05$; token (following token): $r_{pb} = -.07$, $t(151) = -.87, p = .38$; rotation (following token): $r_{pb} = -.08$, $t(151) = -.96, p = .34$; token (following rotation): $r_{pb} = -.07$, $t(151) = -.81, p = .42$]. As in Experiment 3, the trends indicated a very slight negative relationship.

To provide a more powerful analysis of the relationship between object memory and the number of scenes intervening between study and test, the delayed test data from Experiments 3 and 4 were combined into an omnibus regression analysis. The results are displayed in Figure 8. There was a reliable negative correlation between percentage correct change detection performance and number of intervening scenes, $r_{pb} = -.08, t(1255) = -2.97, p < .005$. The regression produced a slope estimate of $-0.27$ (percentage points per intervening scene item) and an intercept of $83.5\%$. The loss of target object information with more scenes viewed is clearly exceedingly gradual.

But this evidence creates a minor puzzle. If some target object information is lost with more intervening scenes, why was no drop in mean change detection performance found from the immediate to delayed tests in, particularly, the token change conditions of Experiments 3 and 4? To examine this issue, the delayed test data from the token change conditions of Experiment 3 and 4 (same test only) were combined. There was a trend toward a negative correlation between change detection performance and number of intervening scenes, $r_{pb} = -.08, t(463) = -1.65, p = .10$, a pattern very similar to that of the omnibus analysis. These data are displayed in Figure 9. Mean percentage correct for token change detection in the immediate test conditions of Experiments 3 and 4 is also plotted in Figure 9. The failure to find mean differences between immediate and delayed token tests, despite evidence of forgetting from the regression analyses, appears to have been caused by the fact that change detection performance was actually slightly improved, relative to the immediate test, for delayed trials with a small number of intervening scenes. This pattern is consistent with the results of Experiments 1 and 2, in which there was a numerical advantage for tests delayed one trial.
compared with immediate tests. These data are not sufficient to support any theoretical conclusions, but they do provide an explanation for the pattern of empirical results.

Discussion

In Experiment 4, delayed test performance was not independent of the type of test conducted during the initial session. Delayed test performance was reliably lower when the change detection task switched from initial session to delayed test session. These data demonstrate that participants can selectively encode visual features that are relevant to the immediate change detection task. Attending to an object does not automatically yield a maximally specific representation.

In addition, the same test conditions of Experiment 4 replicated the robust retention of object representations found in Experiment 3. Performance was no lower on the test delayed until the end of the session compared with the test administered immediately. However, regression analyses of delayed test performance in Experiments 3 and 4 revealed a reliable negative correlation between change detection performance and number of intervening scenes. There was some loss of visual object information with more intervening scenes inspected, but this forgetting was exceedingly gradual. With a small number of intervening scenes, however, there was a numerical trend toward better performance in the delayed test condition compared with the immediate test condition, consistent with the results of Experiments 1 and 2. Thus, there appears to be no loss of object information from the online scene representation to the LTM representation maintained a few scenes later, consistent with the hypothesis that online memory for previously attended objects is supported, to a significant degree, by LTM.

General Discussion

The present study investigated the relationship between the online representation of previously attended objects in a scene and long-term object memory. The basic paradigm presented an image of a natural scene for 20 s, a dot onset and offset, a mask, and a test scene in which a single target object was either the same as the original, rotated, or replaced by another token. The test scene either was presented immediately after scene viewing or was delayed, either for one trial (Experiments 1 and 2) or until all stimuli had been viewed (Experiments 3 & 4). A delay of one trial did not reduce change detection performance from that in the immediate test condition. When the test was delayed until the end of the session, there was a small drop in change detection performance for rotations but no evidence of a drop for token changes. The data from Experiments 1 and 2 demonstrate that there is no loss of token- or orientation-specific visual information from the online representation of previously attended objects to the LTM representation maintained one trial later. The data from Experiments 3 and 4 demonstrate that long-term visual object memory is remarkably robust, consistent with the literature on picture memory (e.g., Standing, 1973; Standing et al., 1970). Analyses of delayed test performance in Experiments 3 and 4 did find evidence of forgetting as more intervening scenes were viewed, but the rate of loss was exceedingly gradual.

These data provide support for the visual memory theory of scene representation (Hollingworth, in press; Hollingworth & Henderson, 2002). In this view, as the eyes and attention are oriented from object-to-object within a scene, higher-level visual representations of attended objects are activated, are maintained briefly in VSTM, and are consolidated into LTM. The VSTM representation is soon replaced as attention moves on to other objects. However, higher-level visual representations of previously attended objects accumulate in LTM, forming, over time, a robust and relatively detailed representation of the scene. This proposal originated in evidence that online visual object memory remained accurate even when many fixations on other objects intervened between target fixation and test, making an explanation in terms of VSTM retention unlikely (Hollingworth & Henderson, 2002). If the representation of previously attended objects is primarily supported by LTM, then one should find that memory for previously attended objects during online scene viewing is of similar specificity to object memory under conditions that unambiguously require LTM, such as delay of one trial or delay until the end of the session. The fact that online and long-term scene representations in the present study were indeed of similar specificity provides strong evidence that the online representation of previously attended objects is supported by the same form of representation supporting long-term object memory.

This conclusion is based on the assumption that similar levels of change detection performance provide evidence of representations with similar format and content. It is possible, however, that similar levels of change detection performance could be supported by different forms of representation if the content of those representations was equivalent.
As an example, a verbal description of an object’s visual properties could, in theory, carry the same informational content as visual representation of that object. However, it is more parsimonious to assume that similar levels of performance reflect similar forms of representation than to posit that different forms of representation happen to carry the same informational content. In addition, Experiment 2 demonstrated that verbal encoding does not play a significant role in this paradigm, as a verbal working memory load and articulatory suppression did not significantly reduce change detection performance or alter the pattern of results between immediate and delayed tests (see also Hollingworth, in press). Moreover, successful change detection in the present task would not have been possible using conceptual representations of object identity, because rotations, in particular, do not alter the object’s identity. The present data therefore allow the conclusion that visual representations support both online memory for previously attended objects and object memory under conditions that unambiguously require LTM and that these visual representations are of similar specificity.

If LTM plays a significant role in online scene representation, what role does VSTM play? A good deal of indirect evidence and some direct evidence suggest that the higher-level visual representations maintained in VSTM and LTM have similar characteristics. VSTM across saccades shows sensitivity to object token (Henderson & Hollingworth, in press-a; Pollatsek, Rayner, & Collins, 1984), object orientation (Henderson & Hollingworth, 1999, in press-a; Henderson & Siefert, 1999, 2001), and structural relationships between object parts (Carlson-Radvansky, 1999; Carlson-Radvansky & Irwin, 1995), but shows insensitivity to the precise contours in an image (Henderson, 1997; Henderson & Hollingworth, in press-b) and object size (Pollatsek et al., 1984). These characteristics mirror those found for long-term object memory in the object recognition literature (e.g., Biederman & Cooper, 1991, 1992; Tarr et al., 1997). Most directly, Tarr et al. tested viewpoint sensitivity in novel object recognition in both a short-term memory task (sequential matching with 750 ms ISI) and a LTM task (speeded naming after an initial study session). They observed similar viewpoint-dependent effects across the two methods.

Despite this evidence of similar representational format in VSTM and VLTM, other evidence suggests that very recently attended objects are either more likely to be retained or are retained more veridically compared with objects attended earlier. Phillips and Christie (1977) presented a sequence of novel checkerboard objects and found a robust recency effect that was limited to the most recent object. Memory for earlier objects remained significantly above chance, however, with no evidence of further decline with more intervening stimuli (up to seven). Phillips and Christie interpreted these results as demonstrating a VSTM component (limited to one object) and a visual LTM component, with the former maintaining a more precise representation than the latter. Using more naturalistic stimuli, Irwin & Zelinsky (2002) found that memory for the positions of objects in a free-viewing paradigm exhibited a recency effect (recency here was the number of objects fixated between fixation on the target and test) that was limited to the two or three most recently fixated objects. There was no evidence of further decline with more intervening objects (up to five).

Hollingworth (in preparation) examined the roles of VSTM and VLTM in the online representation of objects in natural scenes. Participants followed an abruptly appearing dot-cue from object-to-object within a scene. During this sequence, a change detection test was introduced by masking an object and then revealing that object either changed (rotation or token change) or unchanged. The number of objects intervening between inspection of the target and the test was manipulated. There was a reliable recency effect, but this advantage was limited to the two or three most recently attended objects (the currently attended object and one object back). There was no further reduction in change detection performance with more intervening objects (up to 10), and this pre-recency performance was quite accurate ($A'$ greater than .80). Thus, the serial position effects of Hollingworth (in preparation) and Irwin and Zelinsky (2002) suggest a limited capacity VSTM component to scene representation, responsible for the recency advantage, and a higher capacity LTM component, responsible for the robust pre-recency performance.

In the present paradigm, the onset dot ensured that the target object was not the most recently attended object in the scene in the immediate test condition. The target may have been the second most recently attended object on some trials, but given the complexity of these scenes, such trials would have been rare. Performance in the immediate test condition was therefore dominated by memory for objects that were attended earlier than the two-object recency window. The serial position evidence from Hollingworth (in preparation) implicates LTM as
supporting the robust pre-recency memory for previously attended objects. The present data demonstrate that memory for these previously attended objects does indeed appear to be the same as memory for objects retained under conditions that unambiguously require LTM. Together, these data provide strong converging support for the claim that LTM plays a central role in the online visual representation of previously attended objects.

To summarize, the representation of objects in scenes appears to be composed of the following forms of visual representation. During a fixation, precise and complete sensory processing (to the limits of acuity) takes place across the visual field. However, such sensory (i.e., iconic) representations are fleeting (e.g., Sperling, 1960; Di Lollo, 1980), do not survive a saccadic eye movement (e.g., Irwin, 1991; Henderson & Hollingworth, in press-b), and thus do not accumulate as the eyes and attention are oriented from object-to-object within the scene. In addition to sensory representation during a fixation, for attended objects, higher-level visual representations (abstracted from precise sensory information) are retained briefly in VSTM and then more robustly in LTM. Both of these forms of visual memory survive shifts of attention and the eyes, and thus can accumulate information from previously attended and fixated regions. Object representations in VSTM are of similar format to object representations stored in LTM, although the VSTM representation appears to be slightly more accurate than the LTM representation, either because there is a higher probability of object retention in VSTM or because VSTM representations are of greater specificity. The contents of VSTM are replaced by the time attention has shifted to approximately two subsequent objects. Visual LTM, on the other hand, is remarkably robust, with no observable loss of information within a trial or across a one-trial delay, and only very moderate loss of information across numerous intervening scenes.

In addition to comparing online and LTM scene representation, the present study makes two additional contributions. First, it confirms the feats of memory found in the earlier literature on picture memory (Nickerson, 1965; Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). Experiments 3 and 4 exhibited remarkably little loss of visual information from individual objects in scenes across numerous intervening scenes and scores of intervening objects (see also Hollingworth & Henderson, 2002). Whereas many of the early picture memory studies used recognition tasks that could not isolate visual memory, the present results were obtained under conditions that required retention of specifically visual attributes of objects. In addition, whereas many of the early picture memory studies used non-studied pictures as distractors, the present study used old items (same condition) and new items (changed conditions) that were identical except for a single object in the scene. Thus, the present study provided a much more stringent test targeted to visual memory, yet it replicated the earlier findings of large capacity and very gradual forgetting.

These results contrast with results from a recent study by Potter, Staub, Rado, and O’Connor (2002). Potter et al. presented a series of between 5 and 20 unrelated photographs in rapid serial visual presentation (RSVP) mode, with each picture displayed at fixation for 173 ms and 0 ISI. The RSVP stream was followed by a sequential recognition memory test including the viewed photographs and an equal number of distractor photographs. Potter et al. found that with the exception of the first item in the RSVP stream (which was remembered better than pictures at other serial positions), serial position did not matter, even with a series of 20 photographs. That is, the 2nd item in the RSVP series was remembered as accurately as the 20th. This result is consistent with the present finding of very little interference in memory from numerous subsequently viewed scenes. However, Potter et al. found a significant effect of serial position on the recognition memory test, with quite accurate recognition if the photograph was the first one tested (over 80% correct) but rapidly decreasing performance for items tested later. Potter et al. argued that memory for the visual details of pictures presented in RSVP is fleeting. In addition, they concluded that because the RSVP method introduces the same sort of discrete perceptual events as produced by eye movements during normal scene perception, these data suggest that very little visual information is retained from one fixation on a scene to subsequent fixations.

The Potter et al. conclusion is based on the assumption that an RSVP stream of different photographs provides a close analog to the perceptual events generated by eye movements during scene viewing. But during normal viewing of a scene, each eye movement does not result in the appearance of an entirely new scene. The median saccade length during scene viewing is approximately 2° (Henderson &
Hollingworth, 1998), so there is considerable overlap in the information available in consecutive fixations. Thus, the perceptual and memorial demands of processing an entirely new image every 173 ms may yield rapid forgetting that is not typical of normal scene viewing, which does not introduce equivalent demands. The present data demonstrate that these RSVP effects do not generalize to visual accumulation across eye movements during free-viewing of scenes; memory performance was essentially unreduced by multiple intervening scenes and multiple intervening tests, despite the fact that participants made many eye movements during viewing. The difference between the Potter et al. (2002) results and those of the present study are likely caused by differences in the time available for consolidation (Potter, 1976). When participants have sufficient opportunity to consolidate visual information into memory during free viewing, visual memory is robust.

The second additional contribution of the present study was to demonstrate that visual memory is sensitive to task demands and viewer strategies. Experiment 4 examined the possibility that attending to an object yields a memory representation of maximal specificity, with all available visual information encoded regardless of task or strategy. This hypothesis could be generated from evidence that the encoding of additional object features may not consume additional memory capacity (Luck & Vogel, 1997). In Experiment 4, however, delayed test performance was significantly worse when preceded by a different type of change detection task in the initial session than when preceded by the same type of change detection task in the initial session. Participants initial encoding was clearly selective. Such selective processing may be due to the complexity of the real-world scenes and objects used in the present study. Hollingworth and Henderson (2002) found that change detection performance was highest when the target had been fixated at study for 1500 ms or longer. This encoding duration for a single object is an order of magnitude longer than the entire presentation duration producing maximum levels of performance for multiple objects in studies using simple stimuli (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). The extended encoding period required by complex, natural objects provides opportunity to selectively attend to relevant object attributes. Within 1500 ms, multiple individual fixations are directed to different object regions. Thus, just as selection over an entire picture is guided by task demands and viewer goals (Yarbus, 1967), the eyes and attention may be oriented to regions within an object that are particularly relevant to the task at hand.

**Conclusion**

The present study compared the specificity of visual object representations maintained during online viewing of a natural scene with object representation maintained in LTM. The data clearly demonstrate that there is no loss of information from object representations maintained while one is actively viewing a scene to object representations maintained after subsequent inspection of an entirely different scene. In addition, there is very little loss of information when memory is tested after minutes of delay and numerous intervening scenes. These results demonstrate that online representations of previously attended objects are of similar specificity to objects representation maintained under conditions that unambiguously require LTM, supporting the proposal that LTM plays a central role in the online representation of previously attended objects (Hollingworth & Henderson, 2002). In contrast to claims that visual memory is transient (Irwin & Andrews, 1996; Rensink, 2000), the present study demonstrates that visual memory is instead quite remarkably robust.

**References**


### Author Notes
This research was supported by National Institute of Mental Health Grant R03 MH65456. Correspondence concerning this article should be addressed to Andrew Hollingworth, The University of Iowa, Department of Psychology, 11 Seashore Hall E, Iowa City, IA, 52242-1407. Email should be sent to andrew-hollingworth@uiowa.edu.

### Appendix

Mean Percentages Correct Data for Experiments 1-4

<table>
<thead>
<tr>
<th>Test Condition</th>
<th>Same</th>
<th>Rotation</th>
<th>Token Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>86.9</td>
<td>73.8</td>
<td>76.2</td>
</tr>
<tr>
<td>Delayed (one trial)</td>
<td>87.5</td>
<td>80.3</td>
<td>80.9</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>82.1</td>
<td>70.2</td>
<td>71.4</td>
</tr>
<tr>
<td>Delayed (one trial)</td>
<td>81.5</td>
<td>75.0</td>
<td>78.0</td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>82.1</td>
<td>76.2</td>
<td>70.2</td>
</tr>
<tr>
<td>Delayed (end of session)</td>
<td>73.2</td>
<td>71.4</td>
<td>78.6</td>
</tr>
<tr>
<td>Experiment 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same test (rotation-rotation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>80.0</td>
<td>90.0</td>
<td>—</td>
</tr>
<tr>
<td>Delayed (end of session)</td>
<td>88.8</td>
<td>77.5</td>
<td>—</td>
</tr>
<tr>
<td>Same test (token-token)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>86.3</td>
<td>—</td>
<td>73.8</td>
</tr>
<tr>
<td>Delayed (end of session)</td>
<td>88.8</td>
<td>—</td>
<td>67.5</td>
</tr>
<tr>
<td>Different test (token-rotation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>92.5</td>
<td>—</td>
<td>82.5</td>
</tr>
<tr>
<td>Delayed (end of session)</td>
<td>68.8</td>
<td>68.8</td>
<td>—</td>
</tr>
<tr>
<td>Different test (rotation-token)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>90.0</td>
<td>81.3</td>
<td>—</td>
</tr>
<tr>
<td>Delayed (end of session)</td>
<td>78.8</td>
<td>—</td>
<td>72.5</td>
</tr>
</tbody>
</table>
Figure 1. Sequence of events in a trial in the immediate test condition. In the delayed test conditions, the test scene was delayed either one trial (Experiments 1 and 2) or until after all scene had been viewed (Experiments 3 and 4). In this example, the alarm clock is the target object, and the change is token substitution.
Figure 2. Experiment 1: Mean $A'$ as a function of test delay and change type. Error bars represent 95% confidence intervals based on the interaction error term.

Figure 3. Experiment 2: Mean $A'$ as a function of test delay and change type. Error bars represent 95% confidence intervals based on the interaction error term.
Figure 4. Experiment 3: Mean $A'$ as a function of test delay and change type. Error bars represent 95% confidence intervals based on the interaction error term.

Figure 5. Experiment 3: Mean percentage correct change detection in the delayed test condition as a function of the number of scenes intervening between study and test. In each change condition, the mean of each intervening scene quartile is plotted against mean percentage correct detection in that quartile.
Figure 6. Experiment 4: Mean $A'$ as a function of same/different test, test delay, and change type.
Figure 7. Experiment 4: Mean percentage correct change detection in the delayed test condition as a function of the number of scenes intervening between study and test. For each participant group, the mean of each intervening scene quartile is plotted against mean percentage correct detection in that quartile.

Figure 8. Omnibus regression analysis, combining delayed test data from Experiments 3 and 4 (same test only): Mean percentage correct change detection in the delayed test condition as a function of the number of scenes intervening between study and test. For each participant group, the mean of each intervening scene sextile is plotted against mean percentage correct detection in that sextile.
Figure 9. Delayed test data from token change conditions in Experiments 3 and 4 (same test only): Mean percentage correct change detection in the delayed test condition as a function of the number of scenes intervening between study and test. The mean of each intervening scene sextile is plotted against mean percentage correct detection in that sextile. Mean percentage correct for token change detection in the immediate test conditions of Experiments 3 and 4 (same test only) is plotted for comparison.