

The Relationship Between Online Visual Representation of a Scene and Long-Term Scene Memory

Andrew Hollingworth
The University of Iowa

In 3 experiments the author investigated the relationship between the online visual representation of natural scenes and long-term visual memory. In a change detection task, a target object either 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. Change detection during online scene viewing was compared with change detection after delay of 1 trial (Experiments 2A and 2B) until the end of the study session (Experiment 1) or 24 hr (Experiment 3). There was little or no decline in change detection performance from online viewing to a delay of 1 trial or delay until the end of the session, and change detection remained well above chance after 24 hr. These results demonstrate that long-term memory for visual detail in a scene is robust.

Human beings spend their lives within complex environments—offices, parks, living rooms—that typically contain scores of individual objects. All of the visual detail within a scene cannot be perceived in a single glance, as high acuity vision is limited to a relatively small, foveal region of the visual field (Riggs, 1965), so the eyes and attention are oriented serially to local scene regions to obtain high resolution, foveal information from individual objects (see Henderson & Hollingworth, 1998, for a review). Visual scene perception is therefore extended over time and space as the eyes and attention are oriented from object to object. To construct a visual representation of a scene as a whole, visual information from previously fixated and attended objects must be retained in memory and integrated with information from subsequently attended objects.

Once constructed during viewing, how robustly are visual representations of scenes retained in memory? Early evidence from the picture memory literature suggested that visual memory for scenes exhibits exceedingly large capacity and is highly robust (Nickerson, 1965, 1968; Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). Shepard (1967) presented participants with 612 photographs of varying subject matter for 6 s each. Immediately after the study session, mean performance on a two-alternative recognition test was essentially perfect (98% correct). Even more impressive capacity was observed by Standing (1973), who presented each participant with 10,000 different photographs for 5 s each. On a two-alternative recognition test, mean percentage correct was 83%, suggesting the retention of approximately 6,600 pictures. In addition to such demonstrations of capacity, robust retention of pictorial stimuli has been observed over long durations, with above chance picture recognition after a delay of almost 1 year (Nickerson, 1968).

These picture memory studies show that some form of representation is retained robustly from many individual pictures after a single trial of study, and this information is sufficient to discriminate studied pictures from pictures that were not studied. The distractors used in these experiments, however, were typically chosen to be highly different from studied images, making it difficult to identify the type of information supporting such feats of memory. Participants might have remembered studied pictures by maintaining visual representations (coding visual properties such as shape, color, orientation, texture, and so on), by maintaining conceptual representations of picture identity (i.e., scene gist, such as *bedroom*), or by maintaining verbal descriptions of picture content. It is possible that these latter types of nonvisual representation, and not visual memory, accounted for observations of high-capacity, robust picture memory (Chun, 2003; Simons, 1996; Simons & Levin, 1997).

Long-term memory (LTM) for the specific visual properties of objects in scenes was observed in a set of experiments examining the influence of scene contextual structure on object memory (Friedman, 1979; Mandler & Johnson, 1976; Mandler & Parker, 1976; Mandler & Ritchey, 1977; Parker, 1978). Participants reliably remembered token-specific details of individual objects within line drawings of natural scenes. However, the scenes were very simple, typically consisting of a handful of discrete objects; scene items were repeated many times; and verbal encoding was not controlled. Thus, memory for visual object detail could easily have been based on verbal descriptions generated for a small number of objects during the many repetitions of each scene.

Subsequent work has questioned whether LTM reliably retains the visual details of complex stimuli. In a frequently cited study, Nickerson and Adams (1979) tested memory for the visual form of a common object, the penny. Participants were shown a line drawing of the “head” side of a correct penny among 14 alternatives that differed from the correct penny in the surface details. Forty-two percent of participants chose the correct alternative, leading Nickerson and Adams to conclude that memory for pennies is largely schematic, retaining little visual detail. This conclusion must be qualified in two respects, however. First, memory

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

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

encoding was not controlled; the authors assumed that pennies were familiar from everyday experience. Although pennies might be handled quite often, most people rarely attend to the surface details of a penny. Second, although fewer than half the participants (42%) chose the correct penny on the recognition test, the correct alternative was chosen at a rate well above chance of 6.7%, and participants were more than twice as likely to choose the correct penny than to choose any particular distractor. Further, the distractors were highly similar to the correct alternative, often differing in only a single surface feature, such as the position of a text element. Thus, one could just as validly interpret the Nicker-son and Adams results as demonstrating relatively accurate memory for the perceptual details of a common object.

More recently, the phenomenon of change blindness has been used to argue that scene representations in memory are primarily schematic and gist-based, retaining little, if any, specific visual information (Becker & Pashler, 2002; Irwin & Andrews, 1996; O'Regan, 1992; O'Regan & Nöe, 2001; O'Regan, Rensink, & Clark, 1999; Rensink, 2000; Rensink, O'Regan, & Clark, 1997; Simons, 1996; Simons & Levin, 1997; Wolfe, 1999). In change blindness studies, 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, 1999, 2003b) or brief interstimulus interval (ISI; e.g., Rensink et al., 1997). Poor change detection performance in these paradigms has led researchers to propose that a memory representation of the visual detail in a scene is either nonexistent (O'Regan, 1992; O'Regan & Nöe, 2001), limited to the currently attended object (Rensink, 2000, 2002), or limited to the small number of objects (three or four) that can be maintained in visual short-term memory (VSTM) (Becker & Pashler, 2002; Irwin & Andrews, 1996; Irwin & Zelinsky, 2002). With the exception of the currently attended object (or very recently attended objects), scene memory is proposed to be schematic, retaining the gist of the scene and the spatial layout of the scene (i.e., the spatial configuration of surfaces and object positions), but not retaining episodic visual details from local objects. Because these views hold that no robust representation of the visual details of a scene is constructed during online viewing, researchers have necessarily concluded that high capacity, robust LTM in the picture memory literature (e.g., Nickerson, 1968; Standing, 1973) must have been based on the retention of abstract scene gist or layout information (Chun, 2003; Simons, 1996; Simons & Levin, 1997).

Hollingworth and Henderson (2002) sought to resolve the apparent discrepancy between claims of visual transience in the change blindness literature and evidence of high capacity, robust memory for pictures. Participants viewed 3D-rendered images of real-world scenes, and eye movements were monitored. After a target object in the scene had been fixated, and during a subsequent saccadic eye movement to a different, nontarget object in the scene, the target was either rotated 90° in depth, replaced by another object from the same basic-level category (token change), or remained the same. These changes altered the visual details of the target object without changing the gist of the scene, the spatial layout of the scene, or the basic-level identity of the target. And, because attention precedes the eyes to the goal of an eye movement (e.g., Hoffman & Subramaniam, 1995), the target object was no longer attended when the change occurred; attention was allocated to the nontarget object that was the goal of the eye move-

ment. Theories claiming that memory for visual detail is either nonexistent (O'Regan, 1992; O'Regan & Nöe, 2001) or limited to the currently attended object (Rensink, 2000, 2002) predict that such changes should be undetectable. 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 upon the withdrawal of attention (see also Hollingworth, Williams, & Henderson, 2001). Subsequent work has demonstrated that verbal encoding could not account for accurate memory for the visual details of previously attended objects, as change detection performance remained accurate in the presence of a verbal working memory load and articulatory suppression (Hollingworth, 2003, 2004).

In addition to testing object memory during online scene viewing, Hollingworth and Henderson (2002) examined longer-term retention of visual object representations. 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, introducing a delay of approximately 15–20 min, on average, between scene viewing and test. Participants were surprisingly accurate on this delayed test, performing above 80% correct both when discriminating targets from rotated distractors and when discriminating targets from different token distractors. These data provide preliminary evidence that memory for the visual details of individual objects in scenes is robust not only during the online viewing of a scene but also over time scales directly implicating LTM, as tested in the picture memory literature.

To account for these and concomitant results, Hollingworth and Henderson (2002) proposed a descriptive model of scene perception and memory, a visual memory theory of scene representation. In this view, during a fixation, visual sensory representations are initially generated across the visual field. If the scene is removed or perceptual processing otherwise interrupted (e.g., across an eye movement), precise sensory persistence (iconic memory) decays very quickly (Averbach & Coriell, 1961; Di Lollo, 1980; Irwin & Yeomans, 1986; Sperling, 1960) and is not integrated from one view of the scene to the next (Bridgeman & Mayer, 1983; Henderson, 1997; Irwin, 1991; Irwin, Yantis, & Jonides, 1983; McConkie & Zola, 1979; O'Regan & Lévy-Schoen, 1983; Rayner & Pollatsek, 1983). However, directing attention to an object allows the formation of a higher level visual object representation and its consolidation into more stable VSTM (Averbach & Coriell, 1961; Irwin, 1992a; Schmidt, Vogel, Woodman, & Luck, 2002; Sperling, 1960). Higher level visual representations in VSTM are abstracted away from the precise sensory activation characteristic of early, iconic forms of visual memory (Henderson, 1997; Henderson & Hollingworth, 2003b; Irwin, 1991; Phillips, 1974; Pollatsek, Rayner, & Collins, 1984) but retain visual information sufficient to specify object token (Henderson & Hollingworth, 1999, 2003a; Henderson & Siefert, 2001; Pollatsek et al., 1984) and orientation (Henderson & Hollingworth, 1999, 2003a; Henderson & Siefert, 1999, 2001).

Further, the higher level visual representation of an object, activated in VSTM, is associated with a position within a spatial representation of the scene, forming an object file (Henderson, 1994; Hollingworth, in press-a; Kahneman, Treisman, & Gibbs, 1992; Irwin, 1992a). This representation is then consolidated into LTM. As subsequent objects are attended and fixated, higher-level

visual representations of those objects are activated in VSTM. Because VSTM has a limited capacity of, at most, 3 or 4 objects (Luck & Vogel, 1997; Pashler, 1988) representations of objects attended earlier are replaced by new object representations in VSTM. However, the LTM representation is retained robustly and accumulates with visual representations from other previously attended objects. Thus, over the course of scene viewing, VSTM and visual long-term memory (VLTM) support the construction of a relatively detailed representation of the scene, with higher level visual representations of individual objects episodically linked to the larger scene representation through binding of object representations to scene locations, a process likely supported by medial temporal lobe brain regions (Epstein & Kanwisher, 1998; McNaughton et al., 1996; Rolls, 1999).

One of the principal claims of this view is that both VSTM and VLTM contribute to the online representation of a scene. To test this claim, Hollingworth (2004) manipulated serial position during object examination within scenes. On each trial of this follow-the-dot paradigm, participants followed a green dot as it appeared on a series of objects in a scene, shifting gaze to fixate the object most recently visited by the dot. At the end of the dot sequence, a single object was tested in a token-change detection test. Object memory was consistently superior for the two objects fixated most recently before the test. This recency advantage, characteristic of short-term memory retention (Glanzer & Cunitz, 1966; Murdock, 1962; Phillips & Christie, 1977), indicates a VSTM component to online scene representation, apparently limited to two objects, an estimate consistent with independent evidence of VSTM capacity for complex objects (Alvarez & Cavanagh, 2004). Irwin and Zelinsky (2002) observed a similar recency effect for object position memory.

Objects examined earlier than 2 objects before the test were nevertheless remembered at rates well above chance ($A' = \sim .80$),¹ and there was no evidence of further forgetting: An object fixated 3 objects before the test was remembered no more accurately than an object fixated 10 objects before the test. Memory capacity at 10 objects back easily exceeded 3–4 object estimates of VSTM capacity, eliminating the possibility that the online visual representation of a scene is limited to VSTM (Becker & Pashler, 2002; Irwin & Andrews, 1996; Irwin & Zelinsky, 2002). Robust memory for objects fixated and attended early in viewing therefore indicates a significant VLTM component to online scene representation. In summary, VSTM appears to support memory for the last two objects fixated and attended in a scene, with memory for objects attended earlier supported by VLTM.

To examine the capacity of VLTM for visual object representations in scenes, Hollingworth (2004) used the follow-the-dot method but delayed the change detection test until after all scenes had been viewed initially. The mean delay between target examination and test was 12 min. In addition, 31 scenes and 400 objects, on average, intervened between target examination and test. Despite these memory demands, participants performed the token-change detection task at a rate well above chance ($A' = .75$), only moderately lower than when object memory was tested during online viewing. As with the delayed discrimination test in Hollingworth and Henderson (2002), these results suggest high-capacity, robust retention of the visual details of objects in scenes, consistent with the literature on LTM for pictures.

The Present Study

The present study sought to investigate systematically the retention and forgetting of visual detail in scenes. Memory for previously attended objects during the online viewing of a scene was compared with object memory after variable delays, from less than 1 min to 24 hr. In each experiment, participants viewed a series of 3D-rendered scenes for 20 s each. For half of the scenes, a change detection test probing memory for the visual form of a single object in the scene (token or rotation change) was administered 200 ms after the 20-s scene presentation (*immediate test* condition). For the other half (*delayed test* condition), the test was delayed either one trial (Experiments 2A and 2B), until the end of the initial session (Experiment 1), or 24 hr (Experiment 3).

The most basic contribution of this work is to provide foundational evidence regarding the long-term retention of visual scene representations in memory. As reviewed above, the picture memory literature did not isolate visual memory, and thus evidence of high-capacity, robust picture memory cannot be taken as strong evidence for high-capacity, robust visual memory for scenes. The present study isolated visual memory by testing object token and orientation, information more specific than scene gist, scene layout, or even the basic-level identities of individual objects. Initial evidence for robust retention of object detail in scenes comes from the delayed object tests in Hollingworth and Henderson (2002) and Hollingworth (2004), but in each of these studies, the test was delayed only 10–20 min after study. The present work extends retention to as long as 24 hr. If preserved memory for the visual form of objects in scenes were observed after a 24-hr delay, this would provide strong evidence that scene representations in LTM can be precise and accurate, and are not limited to broad, schematic information, such as scene gist or spatial layout. No study to date has examined longer-term scene memory in a paradigm that isolates visual representation.

The second major goal of this study was to examine the relationship between the scene representation maintained during online viewing and scene representations maintained after participants have subsequently viewed other scenes. With the exception of Hollingworth (2004)—and that study used artificial encoding conditions to manipulate serial position—no study to date has directly compared online scene representation with longer-term scene memory. Armchair reflection suggests that people maintain a richer representation of a visual environment when they are currently examining it versus later, when they are no longer present in the environment. If visual memory is consistent with this impression, change detection performance should be higher when the test is administered during online scene viewing (*immediate test* condition) versus when it is administered after the scene has been removed for a significant duration and participants have subsequently viewed other scenes (*delayed test* conditions).

In contrast, the visual memory theory of scene representation (Hollingworth & Henderson, 2002) makes the counterintuitive prediction that change detection performance will not show a significant decline from online test to delayed test, based on the following reasoning. First, visual memory theory holds that with

¹ A' is a signal detection measure of sensitivity with a functional range of .5, chance, to 1.0, perfect sensitivity.

the exception of the two most recently attended objects in a scene, object memory during scene viewing is supported by VLTM (Hollingworth, 2004). The present experiments ensured that in the immediate test condition, the target object was not focally attended when the change occurred, so online change detection should have been almost entirely dependent on VLTM. Second, the VLTM component to online scene representation shows no evidence of decay during viewing, at least over the range of delays tested so far (Hollingworth, 2004). If VLTM is primarily responsible for the online representation of previously attended objects, and if VLTM is highly resistant to decay and interference, then the representation of a scene maintained after other scenes have been subsequently viewed (when scene memory must depend entirely on VLTM) should be essentially equivalent to the representation of previously attended objects maintained during online viewing, also dependent on VLTM. This leads to the prediction of equivalent change detection performance in the immediate and delayed test conditions. Of course, it is likely that there would be some decay of VLTM information if a very long delay were introduced between scene viewing and test (such as 24 hr, Standing et al., 1970), so the critical experiments testing this prediction compared online scene memory with memory after a delay of one trial (Experiments 2A and 2B). The delay of one trial ensured that the scene representation must have been entirely dependent on VLTM but was likely short enough to avoid significant forgetting in VLTM itself (Shepard, 1967).

In summary, if the online representation of previously attended objects in a scene is supported primarily by robust VLTM, as held by visual memory theory, then online scene representations might be no more specific than scene memory maintained after a scene has been removed for a significant duration, when scene representation must be entirely dependent on VLTM.

Experiment 1

Experiment 1 examined the retention and forgetting of visual information in natural scenes, comparing change detection performance during online viewing with change detection on a delayed test administered after all scenes had been viewed initially. Participants viewed 42 different scene stimuli in an initial session. For half of these, a change detection test was presented immediately after viewing the scene (immediate test condition). For the other half, the test was delayed until after all 42 stimuli had been viewed in the initial session (delayed test condition). In the latter condition, successful change detection depended on retaining object information over multiple minutes of delay and numerous intervening scenes.

The paradigm was based on that of Hollingworth (2003). The sequence of events in a trial of the immediate test condition 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 object in the scene would be fixated and attended before the change on the vast majority of trials (in Hollingworth & Henderson, 2002, with 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 in a region of the scene not occupied by the target object, followed by the initial scene again for 200 ms, a pattern mask for 200 ms, and a test scene. The dot onset was

included to ensure that the target object was not focally attended when the change was introduced, on the assumption that an abruptly appearing object (the dot) would capture attention (Hollingworth, 2003; Jonides, 1981; Yantis & Jonides, 1984). In the test scene, a target object was either the same as it had been in the initial scene or changed in one of the following 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 (e.g., in the scene displayed in Figure 1, the plastic watering can, blue in the experimental stimulus, was replaced by a metallic watering can). The target object was specified in the test scene by an arrow postcue to ensure that participants could limit retrieval and comparison processes to the target (Hollingworth, 2003).

To examine online memory for previously attended objects, the immediate test condition presented the test scene immediately after the pattern mask, as illustrated in Figure 1. Although a short delay is introduced in this condition (the 200-ms mask), the phenomenology was of a single presentation of the scene, interrupted briefly by the mask. Online viewing in the real world is itself discontinuous, interrupted every 300 ms or so by a saccadic eye movement and interrupted for ~300 ms every 3 or 4 s by a blink. The masked interval was necessary to ensure that change detection performance could not be based on low-level iconic persistence, but it was significantly shorter than ISIs used to examine VSTM (e.g., Hollingworth, 2004; Vogel, Woodman, & Luck, 2001). The delayed test condition was identical to the immediate test condition, except that the test scene was not displayed in the initial session (the trial ended with the mask). After completion of the initial session, participants completed a delayed test session. The test scenes for the items not tested initially were displayed, and participants reported whether the cued target object had changed. Note that each scene item (and thus each target object) was viewed once in the initial session and tested only once, either immediately after viewing or in the delayed test session. In addition, when viewing each scene initially, participants did not know whether the test of that scene would be immediate or delayed.

The delay until the end of the session is similar to delayed memory tests in Hollingworth and Henderson (2002) and Hollingworth (2004). However, neither of these studies was ideal for examining retention and forgetting within visual scene memory. In Hollingworth and Henderson (2002), the online test (change detection) and delayed test (forced-choice discrimination) were different, precluding the examination of forgetting from online viewing to delayed test. In addition, the target object was postcued in the delayed test but not in the online test, which almost certainly led to an underestimation of memory performance in the online test (see Hollingworth, 2003). In the Hollingworth (2004) follow-the-dot experiment, the online and delayed tests were equivalent (change detection), and there was a significant decline for token-change detection from the test administered online ($A' = .83$) to the test delayed until the end of the session ($A' = .75$). However, the follow-the-dot method did not allow participants to control memory encoding, as they were required to follow the dot from object to object, and thus might have limited their ability to maximize memory performance. To examine forgetting of visual scene information, Experiment 1 used equivalent change detection tests for the immediate and delayed test and allowed participants free viewing during the initial scene presentation.

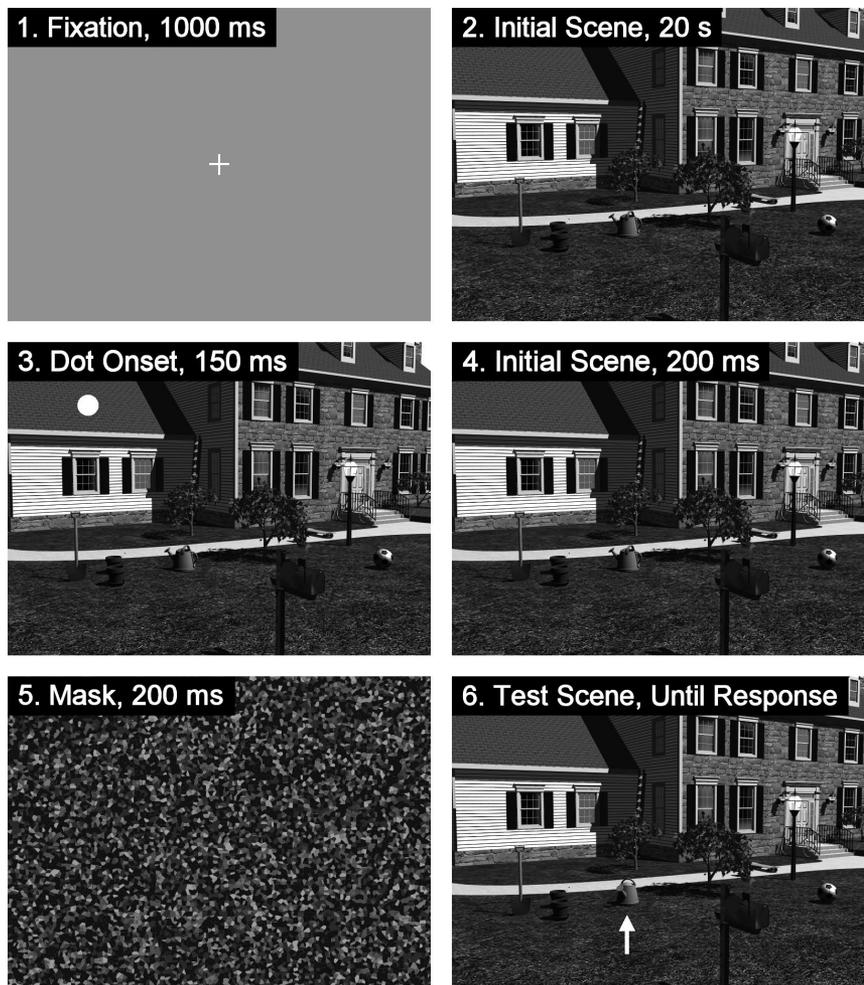


Figure 1. Sequence of events in a trial in the immediate test condition. In this example, the watering can is the target object, and the change is 90° in-depth rotation. In the experiments, stimuli were presented in color.

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. In addition, all object changes were constructed so that the changed object did not obscure or reveal other discreet objects in the scene. Scene images subtended 16.9° (height) \times 22.8° (width) of 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 800×600 pixels by 24-bit color on a 17-in. video monitor at a refresh rate of 100 Hz. The initiation of image presentation was synchronized to the monitor's vertical refresh. Responses were collected by using a serial button box. The presentation of stimuli and collection of responses were 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. Participants were further instructed that for half of the scenes, 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 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.

Participants pressed a pacing button to initiate each trial in the initial session. Then, a white fixation cross on a gray field was displayed for 1,000 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 had either changed or was the same. For the delayed test items in the initial session, the trial ended after the mask (i.e., the test scene was not displayed). There was a 1,500 ms delay between trials.

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 the same-changed response to each. 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 in the delayed test session was determined randomly, with the restriction 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.

Each participant viewed seven scene items in each of the six conditions created by the 2 (test delay: immediate, delayed) \times 3 (change type: same, token change, rotation) factorial design. Item-condition assignments were counterbalanced across participants by Latin square so that each scene item appeared equally often in each condition. A group of 6 participants produced a completely counterbalanced design. The experiment lasted approximately 45 min.

Results

Percentage correct data were used to calculate the sensitivity measure A' . For each participant in each change condition, A' was calculated by 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 by using the participant's hit rate in the immediate token-change condition and the false-alarm rate in the immediate same condition.² All figures show the full functional range of A' , from .50 (chance) to 1.00 (perfect sensitivity). 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 produced the same pattern of results and are reported in the appendix. Given the use of a limited set of scenes and objects, analyses treating scene item as a random effect were conducted (but are not reported individually). Item analyses were conducted over the raw percentage correct data, as A' requires estimating participant decision criteria. Across all the experiments, the reliable A' effects in analyses treating participant as a random effect were also reliable for percentage correct analyses treating scene item as a random effect, and vice versa.

Mean A' in each of the test delay and change conditions is displayed in Figure 2. The test delay manipulation did not produce a reliable main effect, $F(1, 23) = 1.40, p = .25$. 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 a reliable advantage for the immediate test in the rotation condition, $F(1, 23) = 4.43, p < .05$, but there was no effect of test delay in the token-change condition ($F < 1$).

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. Because 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 3. Neither analysis produced a reliable correlation [rotation: $r_{pb} = -.04, t(311) = -.78, p = .43$; token: $r_{pb} = -.08, t(479) = -1.40, p = .16$], although the trends indicated a very slight negative relationship.

Discussion

In Experiment 1, the delayed test condition introduced, on average, 9 min of delay and 31 intervening scenes between scene viewing and test. Despite this delay and considerable amount of potentially interfering scene information, there was only modest forgetting of orientation-specific information from scene representations maintained during online viewing to the scene representations available during the delayed test. More strikingly, there was no evidence of any forgetting at all of token-specific information, as token-change detection performance was actually numerically higher in the delayed test condition than in the immediate test condition. These data provide strong evidence that visual object representations are retained robustly over periods of time that clearly implicate VLTM, consistent with the picture memory literature. Scene representations in VLTM are not limited to gist or abstract layout information.

These data do not replicate the Hollingworth (2004) finding of a reliable drop in token-change detection from immediate test to the test delayed until the end of the session. However, that earlier study artificially controlled participants' eye movement scan path

² For above-chance performance, A' 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): $A' = \frac{1}{2} - \frac{(x-y)(1+x-y)}{4x(1-y)}$

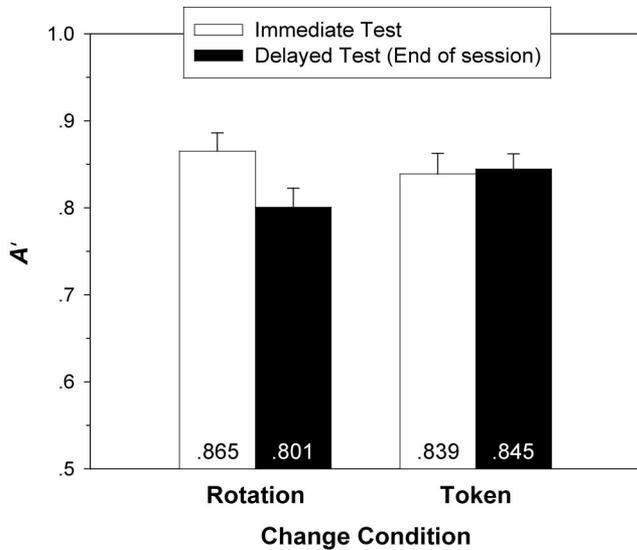


Figure 2. Experiment 1: Mean A' as a function of test delay and change type. Error bars are standard errors of the means.

on the scene to manipulate the serial position of the target object. In addition, fixation time on the target during study in Hollingworth (2004) was approximately 800 ms, whereas target fixation times during 20-s free viewing average approximately 1,500 ms (Hollingworth & Henderson, 2002). Thus, the Hollingworth (2004) data likely underestimated participants' ability to form

robust visual object representations. The reliable drop in orientation change detection performance from immediate to delayed test in Experiment 1 suggests that information specific to orientation might be 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.

These data are generally consistent with the visual memory theory of scene representation (Hollingworth & Henderson, 2002). In that view, online scene representations are primarily supported by VLTM, and VLTM representations of objects show little evidence of decay or susceptibility to interference (Hollingworth, 2004). Thus, the representation of previously attended objects maintained online, when a participant is actively viewing a scene, should be of similar specificity to object representation maintained after the scene has been removed and subsequent scenes viewed, when the representation of a scene must be entirely dependent on VLTM. The token-change detection condition supported this claim, whereas there was evidence of significant forgetting from online to delayed test in the rotation condition. But Experiment 1 provided a very strong test of the equivalence claim, as it required VLTM retention across many intervening scenes and hundreds of individual objects. Experiments 2A and 2B used a delay of one trial, which is sufficient to ensure that scene representation must be entirely dependent on VLTM, but is likely short enough to avoid significant forgetting in VLTM itself, providing an ideal test of the visual memory theory equivalence prediction.

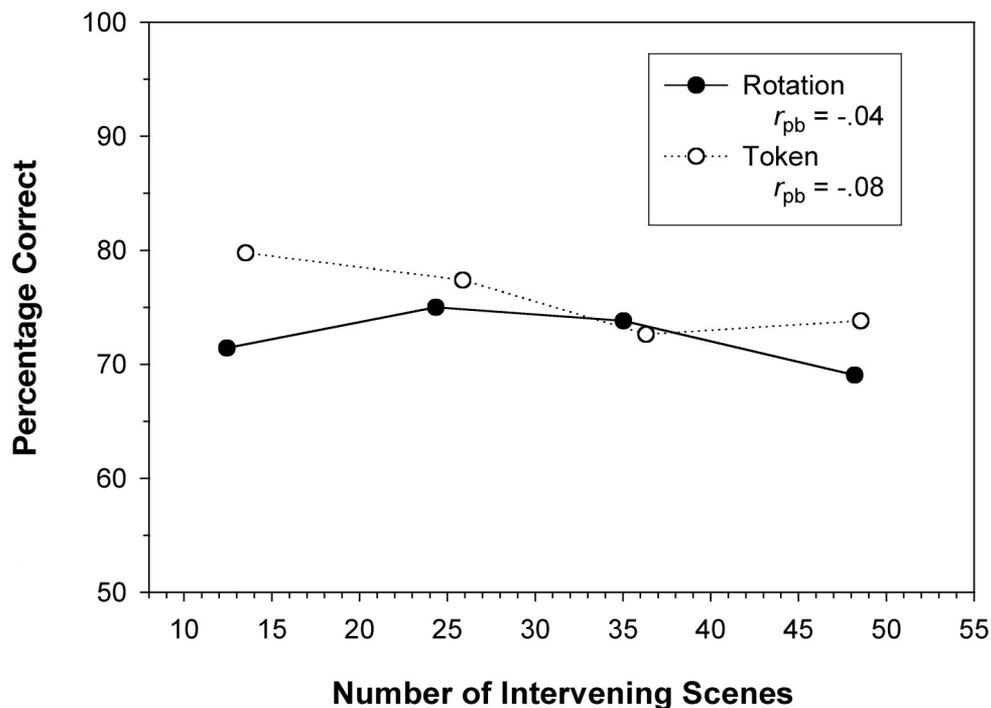


Figure 3. Experiment 1: Mean percentage correct change detection in the delayed test condition as a function of the number of scenes intervening between study and test. In each change condition, the mean of each intervening scene quartile is plotted against mean percentage correct detection in that quartile.

Experiment 2A

In Experiment 2A, the immediate test condition was the same as in Experiment 1. In this condition, the onset dot ensured that the target object was not focally attended when the change occurred. In the delayed test condition, the change detection test was delayed for one trial. A delay of one trial introduced the following events between scene study and test of that scene: (a) change detection test for the previous scene item and (b) viewing of the subsequent scene item 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, at most, three or four objects and the fact that each scene stimulus contained many more than four individual objects. Thus, this experiment compared the specificity of online visual representations of previously attended objects (in the immediate test condition) with the specificity of visual object representations when retention must depend on VLTM. If, as held by visual memory theory, the online visual representation of previously attended objects is itself supported primarily by VLTM, and VLTM representations are resistant to decay and interference, then change detection performance should be no lower in the one-trial delay condition than in the immediate test condition.

The immediate and one-trial delay conditions were conducted in two separate blocks. In the immediate test block, participants viewed half of the scene items, and scene viewing was followed immediately by the test scene for that item (as in the immediate test condition of Experiment 1). In the one-trial delay block, participants viewed the second half of the scene items, but the viewing of each scene was followed by the test scene for the scene item presented one trial earlier. Block order was counterbalanced between participant groups. Experiment 2A implemented this method. Experiment 2B replicated Experiment 2A with the addition of a verbal working memory load and articulatory suppression to minimize the possible influence of verbal encoding.

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 similar to Experiment 1. 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 completed two blocks of trials, an immediate test block and a one-trial delay block. In the immediate test block, each of 21 scenes was presented for 20 s, followed by the dot onset (150 ms), offset (200 ms), mask (200 ms), and test scene, as in Experiment 1 and as illustrated in Figure 1. In the one-trial delay block, the other 21 scenes were likewise each presented for 20 s, followed by the dot onset (150 ms), offset (200 ms), and mask (200 ms). However, the mask was followed by a gray screen with the message *prepare for test of previous scene* displayed for 1,500 ms. This was followed by the test image for the scene item displayed one trial earlier. Thus, on a particular trial, participants viewed scene n , followed by the test scene for scene $n - 1$. On the next trial, participants viewed scene $n + 1$, followed by the test scene for scene n , and so on. Participants responded in the same manner as in the immediate test condition. An extra

scene viewing (dummy item) was added to the end of the one-trial delay block so that the test of the final experimental item could be delayed one trial. The dummy item was not tested. In the one-trial delay block, the mean delay between the end of scene viewing and the test scene for that item was 30.0 s ($SD = 3.3$ s).

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 in each block, seven in each of the three change conditions (same, rotation, token change). Thus, participants viewed a total of 55 different scenes, which consisted of 12 practice items, 42 experimental items, and one dummy item at the end of the one-trial delay block. Item-condition assignments were counterbalanced across participants so that each of the 42 experimental items appeared in each condition an equal number of times. Block order was counterbalanced between participant groups. The entire experiment lasted approximately 50 min.

Results

Because block order neither produced a reliable main effect nor interacted with the change type or test delay variables, the following analysis collapsed block order. Mean A' in each of the test delay and change conditions is displayed in Figure 4. 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 ($F < 1$).

It is important, given the null effect of test delay, to demonstrate that the experiment had sufficient power to detect an effect of test delay, if such an effect had been present. The 95% within-subject confidence interval (CI) was calculated on the basis of the error term of the test delay factor (Loftus & Masson, 1994). This CI was $A' = \pm .041$, and thus the experiment had sufficient power to

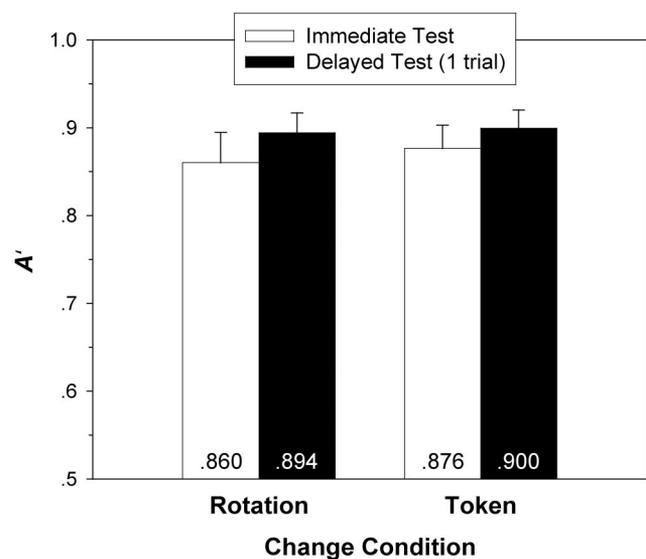


Figure 4. Experiment 2A: Mean A' as a function of test delay and change type. Error bars are standard errors of the means.

detect a difference between test delay means of .058 or larger ($CI * \sqrt{2}$). In Experiment 1, the reliable effect of test delay for rotation-change detection was .064. If the effect of test delay observed in Experiment 1 had been present in Experiment 2A, Experiment 2A had sufficient power to detect it.

However, this assessment of power does not take into account the fact that in Experiment 2A, mean detection performance was actually numerically higher in the one-trial delay condition, which should increase confidence that the absence of a significant drop in performance with test delay did not arise from low power. To provide a quantitative measure reflecting this circumstance, difference scores were computed by subtracting each participant's A' mean in the one-trial delay condition from his or her mean in the immediate test condition. A positive difference score indicated a drop in performance from immediate test to delayed test. Across participants, the mean difference score ($-.029$) was negative because mean performance was numerically higher after one-trial delay. The unidirectional (i.e., one-tailed, as the hypothesis in question is that performance was lower with a one-trial delay) 95% CI was computed for the set of difference scores. This CI was $A' = .048$. Therefore, if a drop in performance from immediate to delayed tests exists, we have 95% confidence that that difference is no greater than .019 ($-.029 + .048$). Such a drop would be quite small, especially in comparison with the .064 effect found in the rotation condition of Experiment 1.

Discussion

The principal issue in this experiment was the relationship between the online representation of a previously attended object in a scene and the representation of that same object when the scene had been removed and object memory must have depended on VLTm. Change detection performance was no worse when the test was delayed one trial compared with when the test was administered during online scene viewing. In fact, the numerical trend was in the direction of higher change detection performance on the delayed test. Thus, Experiment 2A indicates that little or no token- or orientation-specific object information is lost from the representation of a scene maintained while one is actively viewing a scene to the representation maintained after the participant has (a) completed a test for the previous scene item and (b) inspected another, different scene for 20 s. Consistent with the visual memory theory claim that online scene representation of previously attended objects is supported primarily by VLTm, object representations maintained after a one-trial delay appear to be just as 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 or basic-level verbal codes could not support accurate change detection in this paradigm. To examine the potential contribution of verbal description to change detection performance, Hollingworth (2003) added a four-digit verbal working memory load and articulatory suppression to the change detection task used in the immediate test

conditions of Experiments 1 and 2A. Performance was not reduced compared with 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 might play a larger role in scene memory across longer delays, such as those introduced by the delayed test condition in Experiment 2A. Thus, Experiment 2B replicated Experiment 2A but with the addition of a four-digit verbal working memory load and articulatory suppression.

Experiment 2B

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 2A, except for the addition of a four-digit working memory load. On each trial, the initial screen instructing participants to *press a button to begin 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. In the one-trial delay block, the mean delay between the end of scene viewing and the test scene for that item was 30.6 s ($SD = 3.0$ s).

Results and Discussion

A' analysis. Mean A' in each of the test delay and change conditions is displayed in Figure 5. Block order neither produced a reliable main effect nor interacted 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

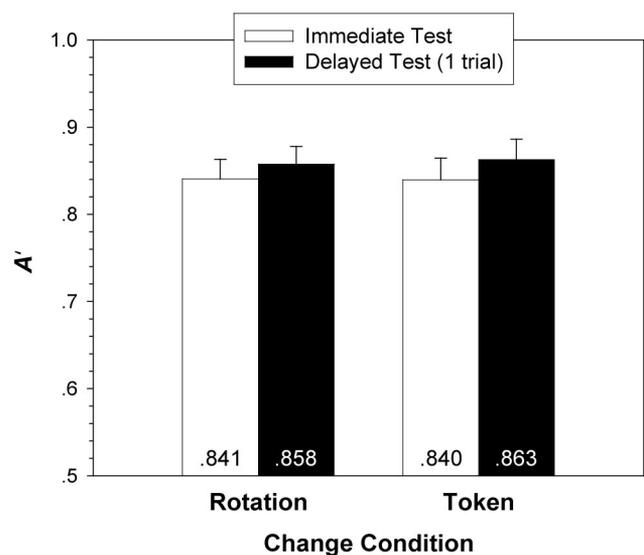


Figure 5. Experiment 2B: Mean A' as a function of test delay and change type. Error bars are standard errors of the means.

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 2A and 2B were combined in an omnibus analysis, with experiment treated as a between-participants factor. The effect of verbal load was not reliable, $F(1, 46) = 1.60$, $p = .21$, nor was the effect of test delay, $F(1, 46) = 1.97$, $p = .17$. Change detection performance was not reliably reduced with the addition of a verbal working memory load, and Experiment 2B replicated the pattern of results found in Experiment 2A, with no decline in change detection sensitivity with a test delay of one trial. Thus, the Experiment 2B data demonstrate that the robust retention effect is supported by visual, not verbal, memory.

As in Experiment 2A, interpreting the null effect of test delay requires evidence that Experiment 2B had sufficient power to detect an effect of test delay, if such an effect had been present. The 95% within-subject CI based on the error term of the test delay factor was $A' = +/- .030$, and thus the experiment had sufficient power to detect a difference between test delay means of .042 or larger. If the reliable .064 effect of test delay in the Experiment 1 rotation condition had been present in Experiment 2B, Experiment 2b had sufficient power to detect it.

To account for the fact that mean accuracy was actually numerically higher in the one-trial delay condition, difference scores were computed by subtracting each participant's A' mean in the one-trial delay condition from his or her mean in the immediate test condition. The mean difference score was $-.020$. The unidirectional 95% CI for the set of difference scores was .035. Thus, if a drop in performance from immediate to delayed tests exists, we have 95% confidence that that difference is no greater than .015 ($-.020 + .035$). Again, such an effect would be quite small, much smaller than the .064 effect found in the rotation condition of Experiment 1.

Finally, to further assess power, the test delay difference scores from Experiments 2A and 2B were pooled. The mean difference score was $A' = -.024$. The unidirectional 95% CI for the pooled difference scores was .029. If there exists a drop in performance from immediate to delayed tests, we have 95% confidence that that difference is no greater than .005 ($-.024 + .029$), a negligible drop unlikely to be psychologically relevant if present. Thus, we can safely infer that there is no meaningful drop in performance with one-trial delay, consistent with the visual memory theory prediction.

Experiment 3

As a very strong test of the robustness of visual memory for objects in scenes, Experiment 3 compared immediate change detection performance with change detection performance after a delay of 24 hr. This length of delay makes a direct connection to the literature on LTM for pictures, but the current method isolates visual memory, which the picture memory literature did not do.

The method was the same as in Experiment 1, except that the delayed test session was administered 24 hr after the initial session. In addition, a verbal working memory load and articulatory suppression were used to minimize verbal encoding.

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. As in Experiment 1, participants completed an initial session in which they viewed all 42 scene items. For half of these, the test scene was displayed immediately after scene viewing. The remaining scenes were tested in a delayed test session administered 24 hr after the initial session. As in Experiment 2B, a verbal working memory load and articulatory suppression minimized verbal encoding. Except for the length of the delay between initial and delayed test sessions and the verbal WM load, Experiment 3 was identical to Experiment 1.

Results and Discussion

Mean A' in each of the test delay and change conditions is displayed in Figure 6. The test delay manipulation produced a reliable main effect, $F(1, 23) = 36.6$, $p < .001$. Mean A' was .892 in the immediate test condition and .696 in the delayed test condition. The main effect of change type was not reliable ($F < 1$), but there was a marginal interaction between test delay and change type, $F(1, 23) = 3.9$, $p = .10$. The effect of test delay was reliable both for the token-change condition, $F(1, 23) = 27.4$, $p < .001$, and for the rotation condition, $F(1, 23) = 23.7$, $p < .001$.

The 24-hr delay significantly impaired change detection performance relative to the test administered during online viewing. This effect was observed both for token-change detection and rotation-change detection. The marginal interaction between delay and

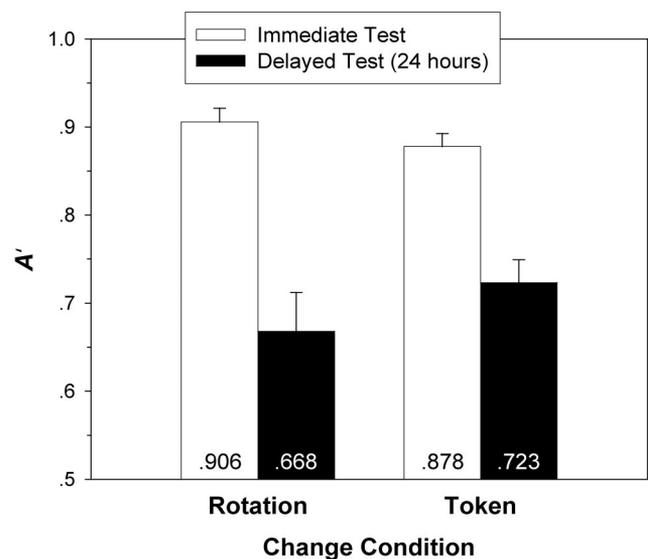


Figure 6. Experiment 3: Mean A' as a function of test delay and change type. Error bars are standard errors of the means.

change type suggests that orientation information was potentially more susceptible to forgetting than token information, consistent with the results of Experiment 1. Despite reliably poorer performance after a delay of 24 hr, change detection performance was still well above chance after 24 hr, both for token-change detection, $t(23) = 8.57, p < .001$, and for rotation-change detection, $t(23) = 3.81, p < .001$. The results of Experiment 3 demonstrate robust retention of visual detail in natural scenes. Although there was significant forgetting of orientation- and token-specific information from online test to the test delayed 24 hr, change detection performance after 24 hr delay was well above chance.³

General Discussion

The present study asked a fundamental question in visual cognition: How robustly are the visual details of a complex scene retained in memory? Three experiments investigated the retention and forgetting of visual detail in scenes, comparing the online representation of previously attended objects in a scene with object memory after delays that implicated LTM. 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 was either presented immediately after scene viewing or delayed for one trial (Experiment 2A and 2B), until after all stimuli had been viewed initially (Experiment 1), or 24 hr (Experiment 3). The delay of one trial required retention of object information across (a) the test for the previous scene item and (b) the 20 s viewing of the subsequent scene item. Change detection performance was no lower when the test was delayed one trial versus when it was presented during the online viewing of the scene, suggesting no loss of orientation- or token-specific information from the scene representation maintained while viewing a scene to the representation available one trial later. The delay until the end of the session in Experiment 1 required object retention across 9 min and 31 intervening scenes, on average. Despite these considerable memory demands, orientation change detection was only modestly reduced compared with the immediate test. Strikingly, token-change detection performance was unreduced by delay until the end of the session. Finally, when the change detection test was delayed 24 hr in Experiment 3, significant forgetting of visual detail was observed, but change detection performance remained well above chance, for both token- and orientation-change detection.

Together, these results demonstrate that memory for the visual details of a scene is robust. One of the principal goals of this study was to examine visual scene memory over delays similar to those used in the early picture memory literature (Nickerson, 1965, 1968; Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). In those studies, participants exhibited high-capacity, robust memory for thousands of individual photographs. However, the method of testing memory—typically forced-choice recognition of a studied photograph paired with an unstudied photograph—did not isolate visual memory. Participants might have discriminated studied scenes from unstudied scenes by remembering broad categorical information (the gist of the scene), verbal descriptions of picture content, or abstract spatial representations of scene layout. Indeed, recent reviews of the picture memory literature have concluded that high-capacity, robust pic-

ture memory was unlikely to have been supported by memory for visual detail but instead by memory for scene gist (Chun, 2003) or spatial layout (Simons, 1996; Simons & Levin, 1997). In addition, evidence of change blindness has led many researchers to propose that memory for visual detail in a scene is inherently transient, decaying soon after the withdrawal of attention from an object (Becker & Pashler, 2002; Irwin & Andrews, 1996; O'Regan, 1992; O'Regan & Nöe, 2001; Rensink, 2000; Rensink, O'Regan, & Clark, 1997; Simons, 1996; Simons & Levin, 1997; Wolfe, 1999).

The present study isolated visual memory. Studied scenes (same condition) and new scenes (changed conditions) were identical except for the visual properties of a single object. Thus, performance in the present study depended on memory for scene information far more specific than scene gist and abstract spatial layout. In addition, a verbal working memory load and articulatory suppression minimized possible verbal encoding. The critical data bearing on the issue of robust visual memory came from Experiment 3, which compared change detection immediately after viewing to change detection after a delay of 24 hr. Both orientation and token-change detection remained well above chance after a 24 hr delay. To put this in concrete terms, after having viewed more than 40 different scenes containing hundreds of individual objects, and after a 24-hr delay, participants could still detect, for example, that the watering can in the front yard scene had changed orientation (see Figure 1). This specificity in long-term scene memory stands in contrast with the view that visual memory is transient and that memory representations of scenes are highly schematic. On the basis of change blindness evidence, Simons and Levin (1997) speculated that only the gist of a scene is retained across a saccade from one eye fixation on a scene to the next, a delay of only 20–60

³ In Experiments 1–3, the scene viewed for the initial 20 s always contained the same version of the target object, which either remained the same or was replaced by the different token or rotated target. Although the initial version of the target object was chosen randomly in Experiments 1–3, it is possible that version-specific effects led to an overestimation of memory performance. To eliminate possible target version effects, a control experiment was conducted in which target object version was counterbalanced, with each possible version of the target object equally likely to appear as initial target. As in Experiment 3, memory for the visual form of the target was tested either during online scene viewing or after a delay of 24 hr. Given the increased number of conditions this entailed, two groups of participants were run, with 8 participants completing only rotation change detection and 8 participants completing only token change detection. In addition, the set of experimental scenes was increased to 48. Each participant completed six trials in each of the eight conditions created by the 2 (immediate test, 24-hr delayed test) \times 2 (target same, changed) \times 2 (initial target version) factorial design. In all other respects, the control experiment was identical to Experiment 3.

For rotation change detection, mean A' was .924 in the immediate test condition and .743 in the 24-hr delay condition, which were reliably different, $F(1, 7) = 37.5, p < .001$. Rotation detection performance after a 24-hr delay was reliably above chance, however, $t(7) = 7.32, p < .001$. For token change detection, mean A' was .835 in the immediate test condition and .701 in the 24-hr delay condition, which were reliably different, $F(1, 7) = 8.29, p < .05$. Again, performance after a 24-hr delay was reliably above chance, $t(7) = 4.10, p < .005$. The control data replicated the Experiment 3 results, with a reliable drop in performance with 24-hr delay but with performance after a 24-hr delay significantly above chance.

ms. Instead, participants are capable of retaining the visual details of scores of individual objects in scenes across delays of at least 24 hr.

Given the ability to accumulate and retain visual object representations over significant periods of time in the present study, why have many change blindness studies observed such poor change detection performance? There appear to be five principal causes of change blindness (see also Hollingworth, *in press*; Simons, 2000), none of which was strongly operational in the present experiments. First, because the encoding of scene information into stable VSTM and VLTM is dependent on the allocation of attention (Averbach & Coriell, 1961; Irwin, 1992a; Sperling, 1960; Schmidt et al., 2002) and of the eyes (Hollingworth, Schrock, & Henderson, 2001; Nelson & Loftus, 1980), participants are unlikely to detect a change if the relevant region has not been fixated and attended before to the change. When Hollingworth and Henderson (2002) introduced object changes before fixation of the object, change detection performance was no higher than the false-alarm rate. In change blindness paradigms that use repeated change (e.g., Rensink et al., 1997), a local change may go undetected for a significant duration simply because the changing object has not yet been fixated and attended. In the present study, however, the 20-s scene duration ensured that participants had the opportunity to fixate and attend the target object prior to the change, thus minimizing limitations on encoding.

Second, change blindness occurs if detecting the change requires information more specific than can be represented in the abstract visual format of VSTM and VLTM. As reviewed above, researchers have long known that visual sensory memory is fleeting and does not survive disruptions such as eye movements. In almost all change blindness paradigms, change detection therefore depends on visual representations abstracted away from precise sensory information, maintained in VSTM and VLTM. If change detection requires a representation of sensory-level precision, then the change will be missed (Henderson, 1997; Henderson & Hollingworth, 2003b). In the present study, object changes were designed to be detectable on the basis of abstracted visual representations maintained in VSTM and VLTM.

Third, change blindness can be caused by failures of retention. Early theories proposed that change blindness derived primarily from retention failure (Rensink, 2000; Rensink et al., 1997). In this view, a visual object representation of the currently attended object, maintained in VSTM, disintegrates upon the withdrawal of attention, and thus visual memory is incapable of accumulating information from local objects in a scene. Although retention failure does play a role in change blindness, as there is significant forgetting of information from VSTM to VLTM (Hollingworth, 2004), it is clearly not the case that visual object representations disintegrate upon the withdrawal of attention.

Fourth, even if a visual representation of sufficient fidelity has been encoded and retained across a disruption, change blindness occurs if that representation is not retrieved and compared with perceptual information after the change. To examine the role of retrieval and comparison failure in change blindness, Hollingworth (2003) used a change detection task similar to the immediate test condition of the present study. In the test scene, the target object was either postcued by an arrow or not postcued. The latter is characteristic of change blindness studies. When the postcue limited decision processes to a single object, minimizing retrieval and

comparison demands, change detection performance was significantly improved relative to the no postcue condition. Retrieval and comparison failure also play a significant role in change blindness under real world, incidental conditions (Simons & Levin, 1998). Simons, Chabris, Schnur, and Levin (2002) covertly removed a basketball as a participant engaged in conversation with an experimenter dressed for the gym. Participants often failed to report the fact that the basketball had been removed, but when given a direct retrieval cue mentioning the basketball, participants could often report distinctive perceptual features of the basketball. A visual memory representation was retained, but that representation was not retrieved and compared with current perceptual information to support change detection during the interaction. To minimize retrieval and comparison failure, the present study used a target postcue as was done in Hollingworth (2003).

Finally, change blindness occurs because changes detected by the visual system are not always sufficient to support explicit report of change. Many studies have found evidence of sensitivity to change on indirect measures, such as fixation duration, in the absence of explicit change report (Angelone, Levin, & Simons, 2003; Fernandez-Duque & Thornton, 2000, *in press*; Hayhoe, Bensinger, & Ballard, 1998; Henderson & Hollingworth, 2003a; Hollingworth & Henderson, 2004; Hollingworth, Williams, & Henderson, 2001; Ryan, Althoff, Whitlow, & Cohen, 2000; Williams & Simons, 2000). Although explicit change detection in the present study also may have underestimated visual memory, the postcue should have facilitated exhaustive assessment of stored information relevant to detecting the change. In summary, when other causes of change blindness—such as limitations on encoding, retrieval, and comparison—were minimized in the present study, retention failure was not a major limiting factor. Object representations supporting change detection were retained robustly.

The second major goal of this study was to test the claim that the online representation of previously attended objects in a scene is primarily supported by VLTM (Hollingworth, 2004; Hollingworth & Henderson, 2002). Under the visual memory theory of scene representation (Hollingworth & Henderson, 2002), as the eyes and attention are oriented from object to object within a scene, higher level visual representations of attended objects are activated, maintained briefly in VSTM, and consolidated into VLTM. The VSTM representation is soon replaced as attention moves on to other objects. However, higher level visual representations of previously attended objects accumulate in VLTM, forming, over time, a robust and relatively detailed representation of the scene. VSTM supports memory for approximately the last two objects fixated and attended in a scene (the currently attended object and one object back), with VLTM supporting memory for all objects fixated and attended earlier in viewing (Hollingworth, 2004). If memory for previously attended objects during online viewing is primarily supported by VLTM, and if VLTM representations are highly resistant to decay and interference (Hollingworth, 2004), then online memory for previously attended objects should be of similar specificity to object memory under conditions that unambiguously require VLTM, such as a delay of one trial in Experiments 2A and 2B.

Consistent with the prediction of visual memory theory, change detection performance after a delay of one trial was unreduced from the test administered during online viewing. In fact, the trend

was in the direction of higher change detection performance on the delayed test. These data, in conjunction with evidence from serial position manipulations (Hollingworth, 2004) provide strong support for the claim that VLTM plays a central role in the online visual representation of previously attended objects (Hollingworth & Henderson, 2002). Memory for previously attended objects during online viewing is highly robust and resistant to decay, exceeds VSTM capacity, and, as demonstrated in the present study, is of similar specificity to object memory under conditions that unambiguously require VLTM. The clear implication is that the online representation of previously attended objects is itself supported primarily by VLTM.

With the present data, it is now possible to generate a fairly complete account of memory for the visual properties of natural scenes. Scene representation is characterized by three main stages of visual retention over three time scales, as illustrated in Figure 7 (see also Irwin, 1992b). First, precise and complete (i.e., iconic) sensory-level representations are generated during an eye fixation on a scene. Precise sensory information is fleeting, however, persisting for less than 300 ms after a stimulus event (Averbach & Coriell, 1961; Di Lollo, 1980; Irwin & Yoemans, 1986; Sperling, 1960). In addition, sensory information is not retained or integrated across visual disruptions such as eye movements (Bridgeman & Mayer, 1983; Henderson, 1997; Irwin, 1991; Irwin, Yantis, & Jonides, 1983; McConkie & Zola, 1979; O'Regan & Lévy-Schoen, 1983; Rayner & Pollatsek, 1983). To examine possible sensory integration across eye movements on scenes, Henderson and Hollingworth (2003b) monitored eye movements as participants viewed images of scenes, with a set of vertical gray bars occluding half the image (i.e., as if one were looking at the scene from behind a picket fence). During eye movements, the bars were

shifted so that all previously occluded portions of the scene became visible and all previously visible portions occluded. This change drastically altered the low-level sensory information in the scene but left higher level visual information, such as the general shape of objects and the spatial relationships between objects, essentially intact. Participants were almost entirely insensitive to these changes, demonstrating that sensory information is not retained and integrated from one fixation on a scene to the next. Thus, the sensory contribution to scene representation is limited to each individual fixation on a scene, which lasts approximately 300 ms, on average, during scene viewing (Henderson & Hollingworth, 1998).

Second, VSTM supports more robust retention of object information in scenes. VSTM maintains object-based visual representations (Irwin & Andrews, 1996; Luck & Vogel, 1997) abstracted away from precise sensory information (Henderson, 1997; Irwin, 1991; Phillips, 1974), and VSTM has a limited capacity of approximately two complex, natural objects (Alvarez & Cavanagh, 2004; Hollingworth, 2004). Thus, there is quite drastic loss of visual information from sensory persistence (complete and precise visual information to the limits of acuity) to VSTM (abstracted visual representation of approximately two objects). Unlike transient sensory persistence, VSTM representations can be maintained robustly over the course of multiple seconds (Vogel, Woodman, & Luck, 2001; Phillips, 1974). In addition, VSTM representations are maintained across eye movements (Irwin, 1992a), supporting the active integration of visual information from one fixation to the next (Henderson, 1994, 1997; Henderson & Siefert, 2001; Irwin, 1991, 1992a; Pollatsek, Rayner, & Collins, 1984).

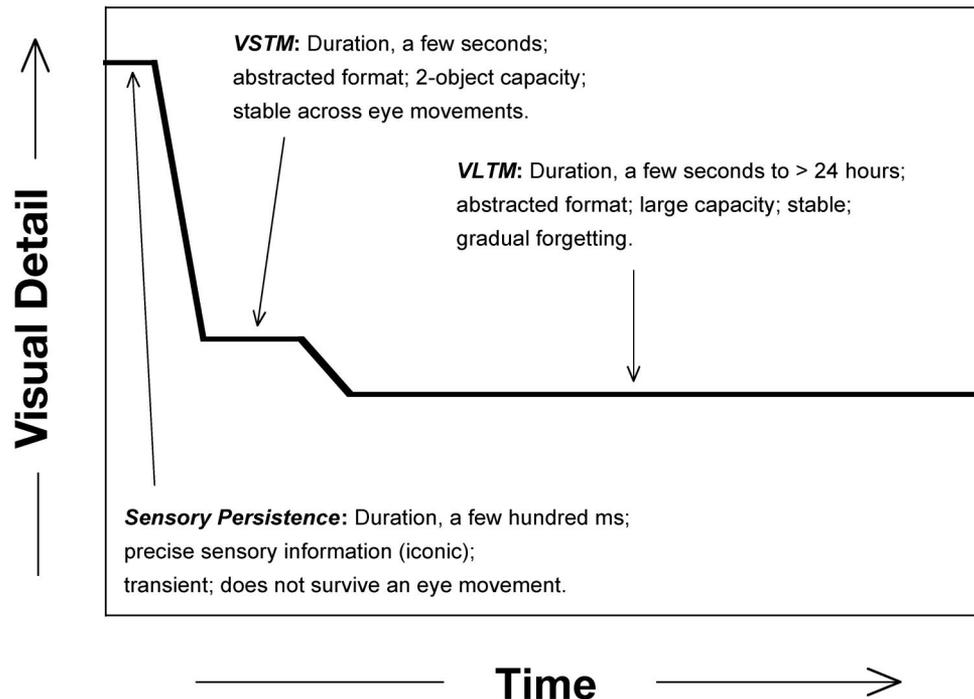


Figure 7. Three stages of visual memory for natural scenes. Ms = milliseconds.

Third, VLTM supports high-capacity, highly robust memory for visual detail within a scene. VLTM maintains object-based representations of a similar format to those maintained in VSTM (see Hollingworth, 2003). Although there is significant loss of precision from VSTM to VLTM (Hollingworth, 2004; Phillips & Christie, 1977), VLTM is still capable of maintaining fairly subtle visual properties, such as object token and orientation. VLTM representations are established relatively quickly during the online viewing of a scene. In Hollingworth (2004), object memory attributable to VLTM was observed for objects fixated and attended three objects before the test and at a temporal delay of only 3,700 ms. The present study demonstrates that VLTM representations are relatively stable, not only during online viewing, but also over extended delays of at least 24 hr. In summary, VLTM supports the representation of object visual detail in a scene, from delays of a few seconds to more than 24 hr, during online viewing and after a scene have been removed, exhibiting large capacity and gradual forgetting.

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.
- 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.
- Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception*, *32*, 947–962.
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal*, *40*, 309–328.
- Becker, M. W., & Pashler, H. (2002). Volatile visual representations: Failing to detect changes in recently processed information. *Psychonomic Bulletin & Review*, *9*, 744–750.
- Bridgeman, B., & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society*, *21*, 285–286.
- Chun, M. M. (2003). Scene perception and memory. *Psychology of Learning and Motivation*, *42*, 79–108.
- Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*, 75–97.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition*, *7*, 324–344.
- Fernandez-Duque, D., & Thornton, I. M. (in press). Explicit mechanisms do not account for implicit localization and identification of change: *Journal of Experimental Psychology: Human Perception and Performance*.
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316–355.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*, 351–360.
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception: Vancouver studies in cognitive science* (Vol. 5, pp. 89–110). Oxford, England: Oxford University Press.
- Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*, 125–137.
- 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. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–283). Oxford, England: Elsevier.
- Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*, 438–443.
- Henderson, J. M., & Hollingworth, A. (2003a). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics*, *65*, 58–71.
- Henderson, J. M., & Hollingworth, A. (2003b). Global transsaccadic change blindness during scene perception. *Psychological Science*, *14*, 493–497.
- Henderson, J. M., & Siefert, A. B. (1999). The influence of enantiomorphic transformation on transsaccadic object integration. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 243–255.
- Henderson, J. M., & Siefert, A. B. (2001). Types and tokens in transsaccadic object identification: Effects of spatial position and left-right orientation. *Psychonomic Bulletin & Review*, *8*, 753–760.
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*, 787–795.
- 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. (in press a). Memory for object position in natural scenes. *Visual Cognition: Special Issue on Scene Perception*.
- Hollingworth, A. (in press b). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition: Special Issue on Visual Search and Attention*.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 113–136.
- Hollingworth, A., & Henderson, J. M. (2004). Sustained change blindness to incremental scene rotation: A dissociation between explicit change detection and visual memory. *Perception & Psychophysics*, *66*, 800–807.
- Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition*, *29*, 296–304.
- Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8*, 761–768.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*, 420–456.
- Irwin, D. E. (1992a). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 307–317.
- Irwin, D. E. (1992b). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146–165). New York: Springer-Verlag.
- 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., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics*, *34*, 35–46.
- 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 perception: Memory for things observed. *Perception & Psychophysics*, *64*, 882–895.
- Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, *1*, 476–490.
- Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
- Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 529–540.
- Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 38–48.
- Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 386–396.
- McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception & Psychophysics*, *25*, 221–224.
- 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.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488.
- Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 391–399.
- Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, *19*, 155–160.
- Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science*, *11*, 58.
- Nickerson, R. S., & Adams, J. J. (1979). Long-term memory for a common object. *Cognitive Psychology*, *11*, 287–307.
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461–488.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–1011.
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999, March 4). Change blindness as a result of "mudsplashes." *Nature*, *398*, 34.
- O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, *23*, 765–768.
- Parker, R. E. (1978). Picture processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 284–293.
- 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.
- Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, *29*, 117–133.
- Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, *113*, 426–442.
- Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception & Psychophysics*, *34*, 39–48.
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277.
- 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.
- Riggs, L. A. (1965). Visual acuity. In C. H. Graham (Ed.), *Vision and Visual Perception* (pp. 321–349). New York: Wiley.
- Rolls, E. T. (1999). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, *9*, 467–480.
- Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science*, *8*, 368–373.
- 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.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.
- Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, *7*, 301–305.
- Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7*, 1–16.
- Simons, D. J., Chabris, C. F., Schnur, T. T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness & Cognition*, *11*, 78–97.
- 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. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2,500 visual stimuli. *Psychonomic Science*, *19*, 73–74.
- Vogel, E. K., Woodman, G. E., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 92–114.
- Williams, P., & Simons, D. J. (2000). Detecting changes in novel 3D objects: Effects of change magnitude, spatiotemporal continuity, and stimulus familiarity. *Visual Cognition*, *7*, 297–322.
- Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.
- Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 601–621.

Appendix

Mean Percentage Correct Data for Experiments 1–3

Test condition	Same	Rotation	Token change
Experiment 1			
Immediate	82.1	76.2	70.2
Delayed (end of session)	73.2	71.4	78.6
Experiment 2A			
Immediate	86.9	73.8	76.2
Delayed (one trial)	87.5	80.3	80.9
Experiment 2B			
Immediate	82.1	70.2	71.4
Delayed (one trial)	81.5	75.0	78.0
Experiment 3			
Immediate	92.3	75.6	66.1
Delayed (24 h)	65.5	58.9	64.9
Experiment 3: control			
Group 1			
Immediate	90.6	80.2	
Delayed (24 h)	72.9	60.4	
Group 2			
Immediate	87.5		61.5
Delayed (24 h)	72.9		53.1

Received July 26, 2004
Revision received October 18, 2004
Accepted October 20, 2004 ■