Journal of Experimental Psychology:
Human Perception and Performance
2004, Vol. 30, No. 3, 519–537

0096-1523/04/$12.00 DOI: 10.1037/0096-1523.30.3.519

Constructing Visual Representations of Natural Scenes: The Roles of Short- and Long-Term Visual Memory

Andrew Hollingworth
University of Iowa

A "follow-the-dot" method was used to investigate the visual memory systems supporting accumulation of object information in natural scenes. Participants fixated a series of objects in each scene, following a dot cue from object to object. Memory for the visual form of a target object was then tested. Object memory was consistently superior for the two most recently fixated objects, a recency advantage indicating a visual short-term memory component to scene representation. In addition, objects examined earlier were remembered at rates well above chance, with no evidence of further forgetting when 10 objects intervened between target examination and test and only modest forgetting with 402 intervening objects. This robust prerecency performance indicates a visual long-term memory component to scene representation.

A fundamental question in cognitive science is how people represent the highly complex environments they typically inhabit. Consider an office scene. Depending on the tidiness of the inhabitant, an office likely contains at least 50 visible objects, often many more (over 200 in my office). Although the general identity of a scene can be obtained very quickly within a single eye fixation (Potter, 1976; Schyns & Oliva, 1994), acquisition of detailed visual information from local objects depends on the serial selection of objects by movements of the eyes (Hollingworth & Henderson, 2002; Nelson & Loftus, 1980). As a result, visual processing of scenes is typically a discrete, serial operation. The eyes are sequentially oriented to objects of interest (Henderson & Hollingworth, 1998), bringing each object onto the fovea, where acuity is highest (Riggs, 1965). During eye movements, however, visual perception is suppressed (Matin, 1974). Thus, eye movements divide scene perception into a series of discrete perceptual episodes, corresponding to fixations, punctuated by brief periods of blindness resulting from saccadic suppression. To construct a representation of a complex scene, visual memory is required to accumulate detailed information from attended and fixated objects as the eyes and attention are oriented from object to object within the scene (Hollingworth, 2003a; Hollingworth & Henderson, 2002).

The present study investigated the visual memory systems that contribute to the construction of scene representations. Current research suggests there are four different forms of visual memory (see Irwin, 1992b; Palmer, 1999, for reviews) and thus four potential contributors to the visual representation of complex scenes: visible persistence, informational persistence, visual short-term memory (VSTM),[1] and visual long-term memory (VLTM). Visible persistence and informational persistence constitute a precise, high-capacity, point-by-point, low-level sensory trace that decays very quickly and is susceptible to masking (Averbach & Coriell, 1961; Coltheart, 1980; Di Lollo, 1980; Irwin & Yeomans, 1986). Together, visible persistence and informational persistence are often termed *iconic memory* or *sensory persistence*. Visible persistence is a visible trace that decays within approximately 130 ms after stimulus onset (Di Lollo, 1980). Informational persistence is a nonvisible trace that persists for approximately 150 to 300 ms after stimulus offset (Irwin & Yeomans, 1986; Phillips, 1974). Although such sensory representations certainly support visual perception within a fixation, sensory persistence does not survive an eye movement and thus could not support the construction of a scene representation across shifts of the eyes and attention (Henderson & Hollingworth, 2003c; Irwin, 1991; Irwin, Yantis, & Jonides, 1983; Rayner & Pollatsek, 1983). Such accumulation is more likely supported by VSTM and VLTM.

VSTM maintains visual representations abstracted away from precise sensory information. It has a limited capacity of three to four objects (Luck & Vogel, 1997; Pashler, 1988) and less spatial precision than point-by-point sensory persistence (Irwin, 1991; Phillips, 1974). However, VSTM is considerably more robust than sensory persistence. It is not significantly disrupted by backward pattern masking and can be maintained for longer durations (on the order of seconds; Phillips, 1974) and across saccades (Irwin, 1992b). These characteristics make VSTM a plausible contributor to the construction of visual scene representations. VLTM maintains visual representations of similar format to those maintained in VSTM (see General Discussion, below) but with remarkably large capacity and robust storage. The capacity of VLTM is not exhausted by retention of the visual properties of hundreds of objects (Hollingworth, 2003b; see also Standing, Conezio, & Haber, 1970). I use the term *higher level visual representation* to

This research was supported by National Institute of Mental Health Grant R03 MH65456. Aspects of this research were presented at the Third Annual Meeting of the Vision Sciences Society, Sarasota, Florida, May 2003.

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

[1] Other authors prefer the term *visual working memory* (see, e.g., Luck & Vogel, 1997). The two terms refer to the same concept.

describe the type of abstracted visual information retained in VSTM and VLTM.

Current theories of scene perception differ greatly in their claims regarding the role of visual memory in scene representation. O'Regan (1992; O'Regan & Noë, 2001) has argued that there is no memory for visual information in natural scenes; the world itself acts as an "outside memory." In this view, there is no need to store visual information in memory because it can be acquired from the world when needed by a shift of attention and the eyes. Rensink (2000, 2002; Rensink, O'Regan, & Clark, 1997) has argued that visual memory is limited to the currently attended object in a scene. For an attended object, a coherent visual representation can be maintained across brief disruptions (such as a saccade, blink, or brief interstimulus interval [ISI]). However, when attention is withdrawn from an object, the visual object representation disintegrates into its elementary visual features, with no persisting memory (for similar claims, see Becker & Pashler, 2002; Scholl, 2000; Simons, 1996; Simons & Levin, 1997; Wheeler & Treisman, 2002; Wolfe, 1999).

Irwin (Irwin & Andrews, 1996; Irwin & Zelinsky, 2002) has proposed that visual memory plays a larger role in scene representation. In this view, higher level visual representations of previously attended objects accumulate in VSTM as the eyes and attention are oriented from object to object within a scene. However, this accumulation is limited to the capacity of VSTM: five to six objects at the very most (Irwin & Zelinsky, 2002). As new objects are attended and fixated and new object information is entered into VSTM, representations from objects attended earlier are replaced. The scene representation is therefore limited to objects that have been recently attended. This proposal is based on evidence that memory for the identity and position of letters in arrays does not appear to accumulate beyond VSTM capacity (Irwin & Andrews, 1996) and that memory for the positions of real-world objects, which generally improves as more objects are fixated, does not improve any further when more than six objects are fixated (Irwin & Zelinsky, 2002).

Finally, Hollingworth and Henderson (2002; Hollingworth, 2003a, 2003b; Hollingworth, Williams, & Henderson, 2001; see also Henderson & Hollingworth, 2003b) have proposed that both VSTM and VLTM are used to construct a robust visual scene representation that is capable of retaining information from many more than five to six objects. Under this *visual memory theory of scene representation*, visual memory plays a central role in the online representation of complex scenes. During a fixation, sensory representations are generated across the visual field. In addition, for the attended object, a higher level visual representation is generated, abstracted away from precise sensory properties. When the eyes move, sensory representations are lost, but higher level visual representations are retained in VSTM and in VLTM. Across multiple shifts of the eyes and attention to different objects in a scene, the content of VSTM reflects recently attended objects, with objects attended earlier retained in VLTM. Both forms of representation preserve enough detail to perform quite subtle visual judgments, such as detection of object rotation or token substitution (replacement of an object with another object from the same basic-level category) (Hollingworth, 2003a; Hollingworth & Henderson, 2002).

This proposal is consistent with Irwin's (Irwin & Andrews, 1996; Irwin & Zelinsky, 2002), except for the claim that VLTM plays a significant role in online scene representation. This difference has major consequences for the proposed content of scene representations. Because VLTM has very large capacity, visual memory theory holds that the online representation of a natural scene can contain a great deal of information from many individual objects. Irwin's proposal, on the other hand, holds that scene representations are visually sparse, with visual information retained from five to six objects at most, certainly a very small proportion of the information in a typical scene containing scores of discrete objects.

Support for the visual memory theory of scene representation comes from three sets of evidence. First, participants can successfully make subtle visual judgments about objects in scenes that have been, but are not currently, attended (Hollingworth, 2003a; Hollingworth & Henderson, 2002; Hollingworth et al., 2001). Theories claiming that visual memory is either absent (O'Regan, 1992) or limited to the currently attended object (see, e.g., Rensink, 2000) cannot account for such findings. Visual memory is clearly robust across shifts of attention.

Second, visual memory representations can be retained over relatively long periods of time during scene viewing, suggesting a possible VLTM component to online scene representation. Hollingworth and Henderson (2002) monitored eye movements as participants viewed 3-D rendered images of complex, natural scenes. The computer waited until the participant fixated a particular target object. After the eyes left the target, that object was masked during a saccadic eye movement to a different object in the scene, and memory for the visual form of the target was tested in a two-alternative forced-choice test. One alternative was the target, and the other alternative was either the target rotated 90° in depth (orientation discrimination) or another object from the same basic-level category (token discrimination). Performance on the forced-choice test was measured as a function of the number of fixations intervening between the last fixation on the target and the initiation of the test. Performance was quite accurate overall (above 80% correct) and remained accurate even when many fixations intervened between target fixation and test. The data were binned according to the number of intervening fixations. In the bin collecting trials with the largest number of intervening fixations, an average of 16.7 fixations intervened between target fixation and test for orientation discrimination and 15.3 for token discrimination. Yet, in each of these conditions, discrimination performance remained accurate (92.3% and 85.3% correct, respectively). Discrete objects in this study received approximately 1.8 fixations, on average, each time the eyes entered the object region. Thus, on average, more than eight objects were fixated between target and test in each condition. Given current estimates of three-to-four-object capacity in VSTM (Luck & Vogel, 1997; Pashler, 1988), it is unlikely that VSTM could have supported such performance, leading Hollingworth and Henderson to conclude that online scene representation is also supported by VLTM.

Third, memory for previously attended objects during scene viewing is of similar specificity to object memory over the long term. In a change detection paradigm, Hollingworth (2003b) presented scene stimuli for 20 s followed by a test scene. The test scene contained either the original target or a changed version of the target (either rotation or token substitution). To examine memory for objects during online viewing, the test scene was displayed 200 ms after offset of the initial scene. To examine memory under

conditions that unambiguously reflected VLTM, the test was delayed either one trial or until the end of the session, after all scene stimuli had been viewed. Change detection performance was generally quite accurate, and it did not decline from the test administered during online viewing to the test delayed one trial. There was a small reduction in change detection performance when the test was delayed to the end of the session, but only for rotation changes. Because visual memory representations during online viewing were no more specific than representations maintained one trial later (when performance must have been based on VLTM), these data suggest that the online representations themselves were also likely to have been retained in VLTM.

The results of Hollingworth and Henderson (2002) and Hollingworth (2003b) provide evidence of a VLTM component to online scene representation. They do not provide direct evidence of a VSTM component, however; the results could be accounted for by a VLTM-only model. The goal of the present study was to examine whether and to what extent VSTM contributes to online scene representation and, in addition, to confirm the role of VLTM.

A reliable marker of a short-term/working memory (STM) contribution to a serial memory task, such as extended scene viewing, is an advantage in the recall or recognition of recently examined items, a *recency effect* (Glanzer, 1972; Glanzer & Cunitz, 1966; Murdock, 1962). In the visual memory literature, recency effects have been consistently observed for the immediate recognition of sequentially presented visual stimuli, ranging from novel abstract patterns (Broadbent & Broadbent, 1981; Neath, 1993; Phillips, 1983; Phillips & Christie, 1977; Wright, Santiago, Sands, Kendrick, & Cook, 1985) to pictures of common objects and scenes (Korsnes, 1995; Potter & Levy, 1969).[2] Phillips and Christie (1977) presented a series of between five and eight randomly configured checkerboard objects at fixation. Memory was probed by a change detection test, in which a test pattern was displayed that was either the same as a presented pattern or the same except for the position of a single filled square. Phillips and Christie observed a recency advantage that was limited to the last pattern viewed.[3] In addition, performance at earlier serial positions remained above chance, with no further decline in performance for earlier serial positions. Phillips and Christie interpreted this result as indicating the contribution of two visual memory systems: a VSTM component, responsible for the one-item recency advantage, and a VLTM component, responsible for stable prerecency performance. If such a data pattern were observed for visual object memory during scene viewing, it would provide evidence of both VSTM and VLTM components to online scene representation.

Before proceeding to examine serial position effects for object memory in scenes, it is important to note that the association between recency effects and STM has not gone unchallenged. The strongest evidence that recency effects reflect STM comes from the fact that the recency and prerecency portions of serial position curves are influenced differently by different variables, such as presentation rate (Glanzer & Cunitz, 1966) and list length (Murdock, 1962), both of which influence prerecency performance without altering performance for recent items. In contrast, the introduction of a brief interfering activity after list presentation, which should displace information from STM, typically eliminates the recency advantage, while leaving prerecency portions of the serial position curve unaltered (Baddeley & Hitch, 1977; Glanzer & Cunitz, 1966). Phillips and Christie (1977) replicated most of these findings in the domain of visual memory, and in particular, they found that a brief period of mental arithmetic or visual pattern matching after stimulus presentation eliminated their one-item recency effect without influencing the stable prerecency performance. Additional evidence connecting recency effects to STM comes from the neuropsychological literature, in which patients with anterograde amnesia exhibited impaired prerecency performance with normal recency performance (Baddeley & Warrington, 1970), whereas patients with STM deficits exhibited normal prerecency performance and impaired recency performance (Shallice & Warrington, 1970). Such behavioral and neuropsychological dissociations strongly suggest the contribution of two memory systems to serial tasks, with the recency advantage attributable to STM.

The strongest challenges to the view that recency advantages are attributable to STM have come on two fronts (see Baddeley, 1986; Pashler & Carrier, 1996, for reviews). First, recency effects can be observed in tasks that clearly tap into long-term memory (LTM), such as recall of U.S. Presidents, a *long-term recency effect* (Baddeley & Hitch, 1977; Bjork & Whitten, 1974; Crowder, 1993). However, the finding that recency effects can be observed in LTM does not demonstrate that recency effects in immediate recall and recognition also arise from LTM; the two effects could be generated from different sources. Indeed, this appears to be the case. Long-term and immediate recency effects are doubly dissociable in patients with LTM deficits, who have shown normal immediate recency effects but impaired long-term recency effects (Carlesimo, Marfia, Loasses, & Caltagirone, 1996), and in patients with STM deficits, who have shown normal long-term recency effects and impaired immediate recency effects (Vallar, Papagno, & Baddeley, 1991). A second challenge has come from Baddeley and Hitch (1977), who found that the recency effect for an auditorily presented list of words was not eliminated by the addition of a digit span task (using visually presented digits) during list presentation. Assuming that the digit span fully occupied STM, then STM could not be responsible for the recency effect. However, as argued by Pashler and Carrier (1996), if one accepts that there are separate STM systems for visual and auditory material (Baddeley, 1986), then the digits in the span task may have been stored visually (see Pashler, 1988, for evidence that alphanumeric stimuli are efficiently maintained in VSTM), explaining the lack of interference with short-term auditory retention. Thus, on balance, present evidence strongly favors the position that recency effects in immediate recall and recognition are a reliable marker of STM.

Three studies have examined serial position effects during the sequential examination of objects in complex scenes. As reviewed above, Hollingworth and Henderson (2002) examined forced-choice discrimination performance as a function of the number of

[2] In contrast, primacy effects are very rare, likely because visual stimuli are difficult to rehearse (Shaffer & Shiffrin, 1972).

[3] Recently, Potter, Staub, Rado, and O'Connor (2002) failed to find a recency advantage for sequences of rapidly presented photographs. However, they never tested memory for the very last picture in the sequence. Given the results of Phillips and Christie (1977), it is likely that VSTM for complex images is limited to the last item viewed, explaining the absence of a recency effect in the Potter et al. study, in that the last item was never tested.

fixations intervening between target fixation and test. There was no evidence of a recency effect in these data, but the paradigm was not an ideal one for observing such an effect. First, the number of intervening objects between target fixation and test could be estimated only indirectly. Second, the analysis was post hoc; serial position was not experimentally manipulated. Third, the data were quite noisy and were likely insufficient to observe such an effect, if one were present.

Irwin and Zelinsky (2002) and Zelinsky and Loschky (1998) examined serial position effects in memory for the location of objects in object arrays (displayed against a photograph of a real-world background). In Irwin and Zelinsky, a set of seven baby-related objects was displayed against a crib background. The same set of seven objects appeared on each of the 147 trials; only the spatial positions of the objects varied. Eye movements were monitored, and a predetermined number of fixations were allowed on each trial. After the final fixation, the scene was removed, and a particular location was cued. Participants then chose which of the seven objects appeared in the cued location. Irwin and Zelinsky found a recency effect: Position memory was reliably higher for the three most recently fixated objects compared with objects fixated earlier. In a similar paradigm, Zelinsky and Loschky presented arrays of nine objects (three different sets, with each set repeated on 126 trials). On each trial, the computer waited until a prespecified target object had been fixated and then counted the number of objects fixated subsequently. After a manipulated number of subsequent objects (between one and seven), the target position was masked, and participants were shown four of the nine objects, indicating which of the four had appeared at the masked location. Zelinsky and Loschky observed a serial position pattern very similar to that of Phillips and Christie (1977). A recency effect was observed: Position memory was reliably higher when only one or two objects intervened between target fixation and test. In addition, prerecency performance was above chance and did not decline further with more intervening objects.

The data from Irwin and Zelinsky (2002) and Zelinsky and Loschky (1998) demonstrate that memory for the spatial position of objects in arrays is supported by an STM component. The stable prerecency data from Zelinsky and Loschky suggest an LTM component as well. However, these studies cannot provide strong evidence regarding memory for the visual form of objects in scenes (i.e., information such as shape, color, orientation, texture, and so on). The task did not require memory for the visual form of array objects; only position memory was tested. Previous studies of VSTM have typically manipulated the visual form of objects (Irwin, 1991; Luck & Vogel, 1997; Phillips, 1974), so it is not clear whether a position memory paradigm requires VSTM, especially given evidence that STM for visual form is not significantly disrupted by changes in spatial position (Irwin, 1991; Phillips, 1974) and given evidence of potentially separate working memory systems for visual and spatial information (see, e.g., Logie, 1995). In addition, both in Irwin and Zelinsky and in Zelinsky and Loschky, the individual objects must have become highly familiar over the course of more than 100 array repetitions, each object was easily encodable at a conceptual level (such as a basic-level identity code), and each object was easily discriminable by a simple verbal label (such as *bottle* or *doll*). Participants could have performed the task by binding a visual representation of each object to a particular spatial position, but they also could have performed the task by associating identity codes or verbal codes with particular positions. Thus, although the Irwin and Zelinsky and the Zelinsky and Loschky studies demonstrate recency effects in memory for what objects were located where in a scene (the binding of identity and position), they do not provide strong evidence of a specifically visual STM component to scene representation.

Present Study and General Method

The present study sought to test whether VSTM and VLTM contribute to the online representation of complex, natural scenes, as claimed by the visual memory theory of scene representation (Hollingworth & Henderson, 2002). A serial examination paradigm was developed in which the sequence of objects examined in a complex scene could be controlled and memory for the visual form of objects tested. In this follow-the-dot paradigm, participants viewed a 3-D-rendered image of a real-world scene on each trial. To control which objects were fixated and when they were fixated, a neon-green dot was displayed on a series of objects in the scene. Participants followed the dot cue from object to object, shifting gaze to fixate the object most recently visited by the dot. A single target object in each scene was chosen, and the serial position of the dot on the target was manipulated. At the end of the sequence, the target object was masked, and memory for the visual form of that object was tested. Serial position was operationalized as the number of objects intervening between the target dot and the test. For example, in a 4-back condition, the dot visited four intervening objects between target dot and test. In a 0-back condition, the currently fixated object was tested.

The sequence of events in a trial of Experiment 1 is displayed in Figure 1. Sample stimuli are displayed in Figure 2. On each trial, participants first pressed a pacing button to initiate the trial. Then, a white fixation cross on a gray field was displayed for 1,000 ms, followed by the initial scene for 1,000 ms (see Figure 2A). The dot sequence began at this point. A neon-green dot appeared on an object in the scene and remained visible for 300 ms (see Figure 2B). The dot was then removed (i.e., the initial scene was displayed) for 800 ms. The cycle of 300-ms dot cue and 800-ms initial scene was repeated as the dot visited different objects within the scene. At a predetermined point in the dot sequence, the dot visited the target object. After the final 800-ms presentation of the initial scene, the target object was obscured by a salient mask for 1,500 ms (see Figure 2C). The target mask served to prevent further target encoding and to specify the object that was to be tested.

In Experiment 1, a sequential forced-choice test immediately followed the 1,500-ms target mask. Two versions of the scene were displayed in sequence. One alternative was the initial scene. The other alternative was identical to the initial scene except for the target object. In the latter case, the target object distractor was either a different object from the same basic-level category (token substitution; see Figure 2D) or the original target object rotated 90° in depth (see Figure 2E). After the 1,500-ms target mask, the first alternative was presented for 4 s, followed by the target mask again for 1,000 ms, followed by the second alternative for 4 s, followed by a screen instructing participants to indicate, by a button press, whether the first or second alternative was the same as the original target object.

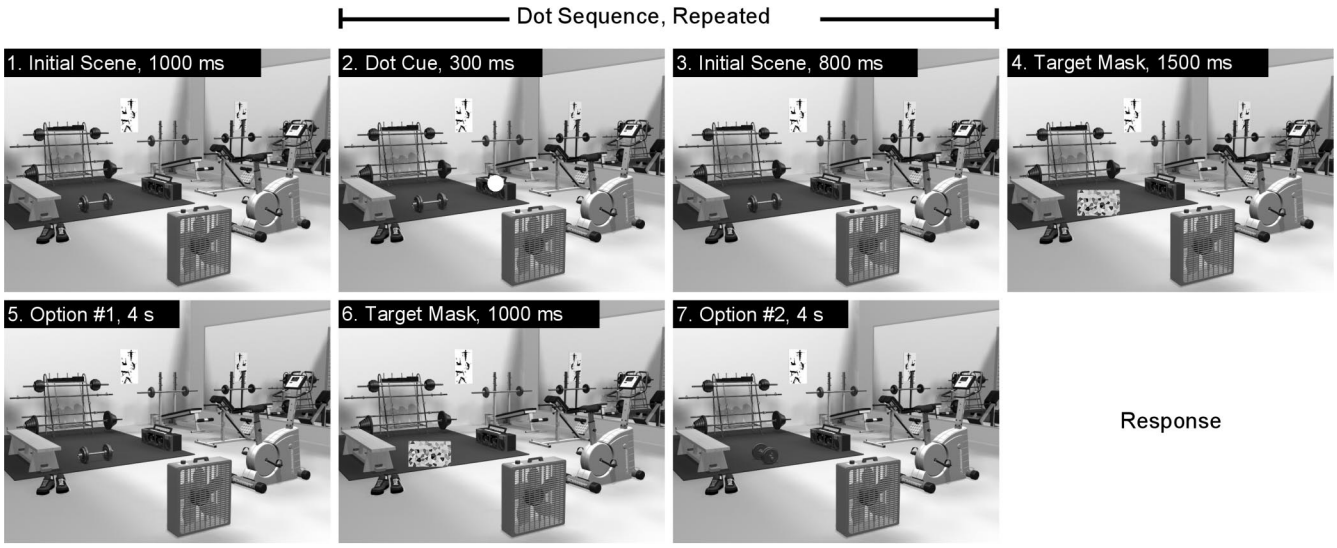

Figure 1. Sequence of events in a trial of Experiment 1. Each trial began with a 1,000-ms fixation cross (not shown). After fixation, the initial scene was displayed for 1,000 ms, followed by the dot sequence, which was repeated as the dot visited different objects in the scene. The dot sequence was followed by a target object mask and presentation of the two test options. The trial ended with a response screen. Participants responded to indicate whether the first or the second option was the same as the original target object. The sample illustrates an orientation discrimination trial with the target appearing as Option 1 and the rotated distractor as Option 2. In the experiments, stimuli were presented in full color.

In Experiments 2–6, a change detection test followed the 1,500-ms target mask. A test scene was displayed until response. In the same condition, the test scene was the initial scene. In the changed condition, the test scene was the token substitution scene (see Figure 2D). Participants responded to indicate whether the target object had or had not changed from the version displayed initially.

In all experiments, participants were instructed to shift their gaze to the dot when it appeared and to look directly at the object the dot had appeared on until the next dot appeared. Participants did not have any difficulty complying with this instruction.[4] Because attention and the eyes are reflexively oriented to abruptly appearing objects (Theeuwes, Kramer, Hahn, & Irwin, 1998; Yantis & Jonides, 1984), following the dot required little effort. In addition, the 300-ms dot duration was long enough that the dot was typically still visible when the participant came to fixate the cued object, providing confirmation that the correct object had been fixated. The 800-ms duration after dot offset was chosen to approximate typical gaze duration on an object during free viewing (Hollingworth & Henderson, 2002). Finally, the dot sequence was designed to mimic a natural sequence of object selection during free viewing, based on previous experience with individual eye movement scan patterns on these and on similar scenes (Hollingworth & Henderson, 2002).

The position of the target object in the sequence was manipulated in a manner that introduced the smallest possible disparity between the dot sequences in different serial position conditions. Table 1 illustrates the dot sequence for a hypothetical scene item in each of three serial position conditions: 1-back, 4-back, and 10-back, as used in Experiments 1 and 2. The dot sequence was identical across serial position conditions, except for the position of the target object in the sequence. The total number of dots was varied from scene item to scene item, from a minimum of 14 total dots to a maximum of 19 total dots, depending on the number of discrete objects in the scene. With fewer total dots, the absolute position of the target appeared earlier in the sequence, with more total dots, later, ensuring that participants could not predict the ordinal position of the target dot.

Experiments 1 and 2 tested serial positions 1-back, 4-back, and 10-back. Experiments 3 and 4 provided targeted tests of recent serial positions, between 0-back and 4-back. Experiment 5 examined memory for earlier positions and included a condition in which the test was delayed until after all scenes had been viewed (an average delay of 402 objects). In Experiments 1–5, each of the 48 scene items appeared once; there was no scene repetition. Experiment 6 examined 10 serial positions (0-back through 9-back) within participants by lifting the constraint on scene repetition. To preview the results, robust recency effects were observed throughout the study, and this memory advantage was limited to the two most recently fixated objects. Prerecency performance was quite accurate, however, and robust: There was no evidence of further forgetting with more intervening objects (up to 10-back) and only modest forgetting when the test was delayed until after all scenes had been viewed (402 intervening objects). Consistent with the visual memory theory of scene representation, these data suggest a VSTM

[4] The present method could not eliminate the possibility that participants covertly shifted attention to other objects while maintaining fixation on the cued object. However, considering that participants were receiving high-resolution, foveal information from the currently fixated object, there would seem to have been little incentive to attend elsewhere.

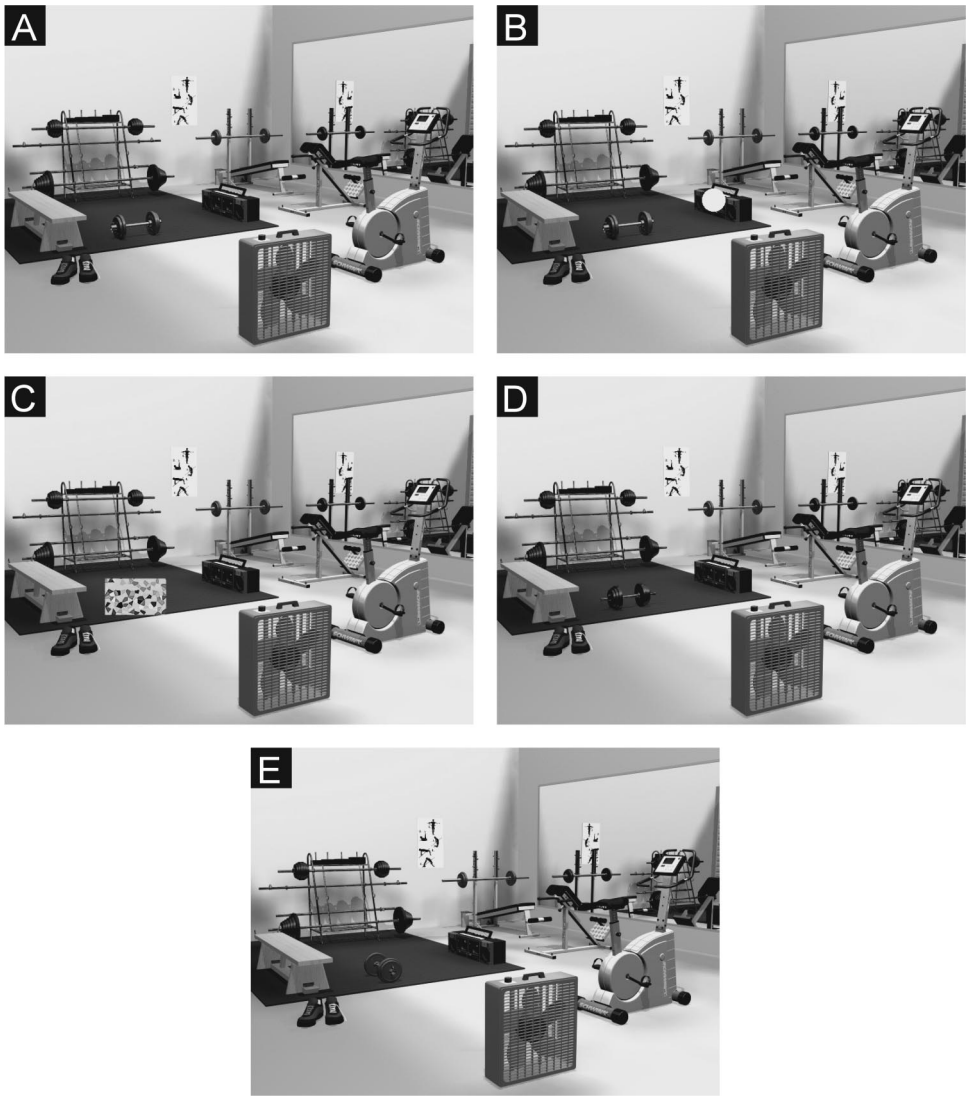

Figure 2. Stimulus manipulations used in Experiments 1–5 for a sample scene item. A: The initial scene (the barbell was the target object). B: The onset dot appearing on an object in the scene. C: The target object mask. D and E: The two altered versions of the scene, token substitution and target rotation, respectively.

Table 1
Sequence of Dots on a Hypothetical Scene Item for Serial Position Conditions 1-Back, 4-Back, and 10-Back

Condition	Objects visited by dot, in order
1-back	A, B, C, D, E, F, G, H, I, J, K, L, M, N, Target, O, (target mask)
4-back	A, B, C, D, E, F, G, H, I, J, K, Target, L, M, N, O, (target mask)
10-back	A, B, C, D, E, Target, F, G, H, I, J, K, L, M, N, O, (target mask)

Note. Letters represent individual objects in the scene.

component to scene representation, responsible for the recency advantage, and a VLTM component, responsible for robust prerecency performance.

Experiment 1

Experiment 1 examined three serial positions of theoretical interest: 1-back, 4-back, and 10-back. The 1-back condition was chosen because object memory in this condition should fall squarely within typical three-to-four-object estimates of VSTM capacity (Luck & Vogel, 1997; Pashler, 1988). The 4-back condition was chosen as pushing the limits of VSTM capacity. The 10-back condition was chosen as well beyond the capacity of VSTM. Evidence of a recency effect—higher performance in the 1-back condition compared with the 4-back and/or 10-back conditions—would provide evidence of a VSTM component to online

scene representation. Evidence of robust prerecency performance—relatively accurate performance in the 10-back condition—would provide evidence of a VLTM component. Performance in the 10-back condition was compared with the prediction of a VSTM-only model of scene representation derived from Irwin and Zelinsky (2002).

Method

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

Stimuli. Forty-eight scene items were created from 3-D models of real-world environments, and 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 (token substitution). The objects for token substitution were chosen to be approximately the same size as the initial target object. Scene images subtended a 16.9° × 22.8° visual angle at a viewing distance of 80 cm. Target objects subtended 3.3° on average along the longest dimension in the picture plane. The object mask was made up of a patchwork of small colored shapes and was large enough to occlude not only the target object but also the two potential distractors and the shadows cast by each of these objects. Thus, the mask provided no information useful to performance of the task except to specify the relevant object (see Hollingworth, 2003a). The dot cue was a neon-green disc (red, green, blue: 0, 255, 0), with a diameter of 1.15°.

Apparatus. The stimuli were displayed at a resolution of 800 × 600 pixels in 24-bit color on a 17-in. video monitor with a refresh rate of 100 Hz. The initiation of image presentation was synchronized to the monitor's vertical refresh. Responses were collected using a serial button box. The presentation of stimuli and collection of responses 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. For each, they should follow the dot, fixating the object most recently visited by the dot, until a single object was obscured by the target mask. Participants were instructed to fixate the mask, view the two object alternatives, and respond to indicate whether the first or second alternative was the same as the original object at that position. The possible distractor objects (rotation or token substitution) were described. Participants pressed a pacing button to initiate each trial. This was followed by the dot sequence, target mask, and forced-choice alternatives, as described in the General Method, above.

Participants first completed a practice session. The first 2 practice trials simply introduced participants to the follow-the-dot procedure, without an object test. These were followed by 4 standard practice trials with a variety of target serial positions (1-back, 4-back, 6-back, and 9-back). Two of the practice trials were token discrimination, and 2 were orientation discrimination. The practice scenes were not used in the experimental session. The practice trials were followed by 48 experimental trials, 4 in each of the 12 conditions created by the 3 (1-back, 4-back, 10-back) × 2 (token discrimination, orientation discrimination) × 2 (target first alternative, second alternative) factorial design. The final condition was for counterbalancing purposes and was collapsed in the analyses that follow. Trial order was determined randomly for each participant. Across participants, each of the 48 experimental items appeared in each condition an equal number of times. The entire experiment lasted approximately 45 min.

Results and Discussion

Mean percentage correct performance in each of the serial position and discrimination conditions is displayed in Figure 3.

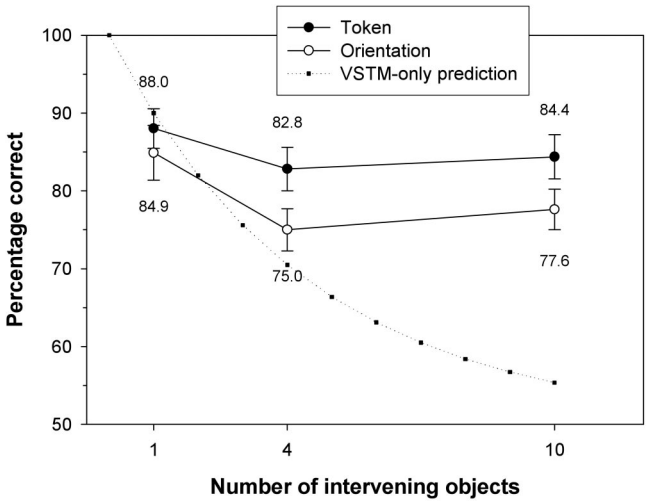

Figure 3. Experiment 1: Mean percentage correct as a function of serial position (number of objects intervening between target dot and test) and discrimination type (token and orientation). Error bars represent standard errors of the means. The dotted line is the prediction of a visual-short-term-memory-only (VSTM-only) model of scene representation.

There was a reliable main effect of discrimination type, with higher performance in the token discrimination condition (85.1%) than in the orientation discrimination condition (79.2%), $F(1, 23) = 12.91$, $p < .005$. There was also a reliable main effect of serial position, $F(2, 23) = 4.96$, $p < .05$. Serial position and discrimination type did not interact, $F < 1$. Planned comparisons of the serial position effect revealed that 1-back performance was reliably higher than 4-back performance, $F(1, 23) = 10.61$, $p < .005$, and that 4-back performance and 10-back performance were not reliably different, $F < 1$. In addition, there was a strong trend toward higher performance in the 1-back condition compared with the 10-back condition, $F(1, 23) = 3.82$, $p = .06$.

Figure 3 also displays the prediction of a VSTM-only model of scene representation (Irwin & Andrews, 1996; Irwin & Zelinsky, 2002). The prediction was based on the following assumptions, drawn primarily from Irwin and Zelinsky (2002). A generous (and thus conservative, for present purposes) VSTM capacity of five objects was assumed. In addition, it was assumed that the currently attended target object (0-back) is reliably maintained in VSTM, yielding correct performance. Furthermore, as attention shifts to other objects, replacement in VSTM is stochastic (Irwin & Zelinsky, 2002), with a .2 (i.e., $1/k$, where k is VSTM capacity) probability that an object in VSTM will be purged from VSTM when a new object is attended and entered into VSTM. The probability that the target object is retained in VSTM (p) after n subsequently attended objects would be

$$p = \left(1 - \frac{1}{k}\right)^n.$$

Correcting for guessing in the two-alternative forced-choice paradigm on the assumption that participants will respond correctly on trials when the target is retained in VSTM and will get 50% correct on the remaining trials by guessing, percentage correct performance under the VSTM-only model (P_{vstm}) can be expressed as

$$P_{vstm} = 100[p + .5(1 - p)].$$

As is evident from Figure 3, this prediction is not supported by the Experiment 1 data.[5] In particular, the VSTM-only model predicted much lower discrimination performance in the 10-back condition than was observed. The present data therefore suggest that the online visual representation of scenes is supported by more than just VSTM. The logical conclusion is that relatively high levels of performance in the 4-back and 10-back conditions were supported by VLTM.

In summary, the Experiment 1 results demonstrate a recency effect in memory for the visual form of objects, suggesting a VSTM component to the online representation of natural scenes. This finding complements the recency advantage observed by Irwin and Zelinsky (2002) and Zelinsky and Loschky (1998) for object position memory. However, the present results do not support the Irwin and Zelinsky claim that scene representation is limited to the capacity of VSTM. Performance was no worse when 10 objects intervened between target dot and test compared with when 4 objects intervened between target and test. This robust prerecency performance suggests a significant VLTM contribution to the online representation of scenes, as held by the visual memory theory of scene representation (Hollingworth, 2003a; Hollingworth & Henderson, 2002).

Experiments 2–5

Experiments 2–5 tested additional serial positions of theoretical interest. In addition, the paradigm from Experiment 1 was improved with the following modifications. First, the two-alternative method used in Experiment 1 may have introduced memory demands at test (associated with processing two sequential alternatives) that could have interfered with target object memory. Therefore, Experiments 2–5 used a change detection test, in which a single test scene was displayed after the target mask. Because token and orientation discrimination produced similar patterns of performance in Experiment 1, the change detection task in Experiments 2–5 was limited to token change detection: The target object in the test scene either was the same as the object presented initially (same condition) or was replaced by a different object token (token change condition).[6] Finally, a four-digit verbal working memory load and articulatory suppression were added to the paradigm to minimize the possibility that verbal encoding was supporting object memory (see Hollingworth, 2003a; Vogel, Woodman, & Luck, 2001, for similar methods).

Experiment 2

Experiment 2 replicated the serial position conditions from Experiment 1 (1-back, 4-back, and 10-back) to determine whether the modified method would produce the same pattern of results as in Experiment 1.

Method

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

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

Procedure. The procedure was identical to Experiment 1, with the following exceptions. In this experiment, the initial screen instructing participants to press a button to start the next trial also contained four randomly chosen digits. Participants began repeating the four digits aloud before initiating the trial and continued to repeat the digits throughout the trial. Participants were instructed to repeat the digits without interruption or pause, at a rate of at least two digits per second. The experimenter monitored digit repetition to ensure that participants complied.

The trial sequence ended with a test scene, displayed immediately after the target mask. In the same condition, the test scene was identical to the initial scene. In the token change condition, the test scene was identical except for the target object, which was replaced by another token. Participants pressed one button to indicate that the test object was the same as the object displayed originally at that position or a different button to indicate that it had changed. This response was unspeeded; participants were instructed only to respond as accurately as possible.

The practice session consisted of the 2 trials of follow-the-dot practice followed by 4 standard trials. Two of these were in the same condition, and 2 were in the token change condition. The practice trials were followed by 48 experimental trials, 8 in each of the six conditions created by the 3 (1-back, 4-back, 10-back) $\times$ 2 (same, token change) factorial design. Trial order was determined randomly for each participant. Across participants, each of the 48 experimental items appeared in each condition an equal number of times. The entire experiment lasted approximately 45 min.

Results and Discussion

Percentage correct data were used to calculate the signal detection measure A' (Grier, 1971). A' has a functional range of .5 (chance) to 1.0 (perfect sensitivity). A' models performance in a two-alternative forced-choice task, so A' in Experiment 2 should produce similar levels of performance as proportion correct in Experiment 1. For each participant in each serial position condition, A' was calculated using the mean hit rate in the token change condition and the mean false alarm rate in the same condition.[7] Because A' corrects for potential differences in response bias in the percentage correct data, it forms the primary data for interpreting

[5] The VSTM-only prediction is based on stochastic replacement in VSTM. Another plausible model of replacement in VSTM is first-in-first-out (Irwin & Andrews, 1996). A VSTM-only model with the assumption of first-in-first-out replacement and five-object capacity would predict ceiling levels of performance for serial positions 0-back through 4-back and chance performance at earlier positions. Clearly, this alternative VSTM-only model is also inconsistent with the performance observed in the 10-back condition.

[6] This change detection task is equivalent to an old/new recognition memory task in which new trials present a different token distractor.

[7] 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 where a participant performed below chance in a particular condition, A' was calculated using the below-chance equation developed by Aaronson and Watts (1987):

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

these experiments. Raw percentage correct data for Experiments 2–6 are reported in the Appendix.

Mean A' performance in each of the serial position conditions is displayed in Figure 4. The pattern of data was very similar to that in Experiment 1. There was a reliable effect of serial position, $F(2, 23) = 5.13$, $p < .01$. Planned comparisons of the serial position effect revealed that 1-back performance was reliably higher than 4-back performance, $F(1, 23) = 11.35$, $p < .005$; that 1-back performance was reliably higher than 10-back performance, $F(1, 23) = 7.16$, $p < .05$; and that 4-back and 10-back performance were not reliably different, $F < 1$. These data replicate the recency advantage found in Experiment 1, suggesting a VSTM component to scene representation, and they also replicate the robust prerecency memory, suggesting a VLTM component to scene representation.

Experiment 3

Experiments 1 and 2 demonstrated a reliable recency effect for object memory in scenes. That advantage did not extend to the 4-back condition, suggesting that only objects retained earlier than four objects back were maintained in VSTM. However, these data did not provide fine-grained evidence regarding the number of objects contributing to the recency effect. To provide such evidence, Experiment 3 focused on serial positions within the typical three-to-four-object estimate of VSTM capacity: 0-back, 2-back, and 4-back. In the 0-back condition, the last dot in the sequence appeared on the target object, so this condition tested memory for the currently fixated object. The 0-back and 2-back conditions were included to bracket the 1-back advantage found in Experiments 1 and 2. The 4-back condition was included for comparison because it clearly had no advantage over the 10-back condition in Experiments 1 and 2 and thus could serve as a baseline measure of prerecency performance. If the recency effect includes the currently fixated object, performance in the 0-back condition should be higher than that in the 4-back condition. If the recency effect extends to three objects (the currently fixated object plus two objects back), then performance in the 2-back condition should be higher than that in the 4-back condition.

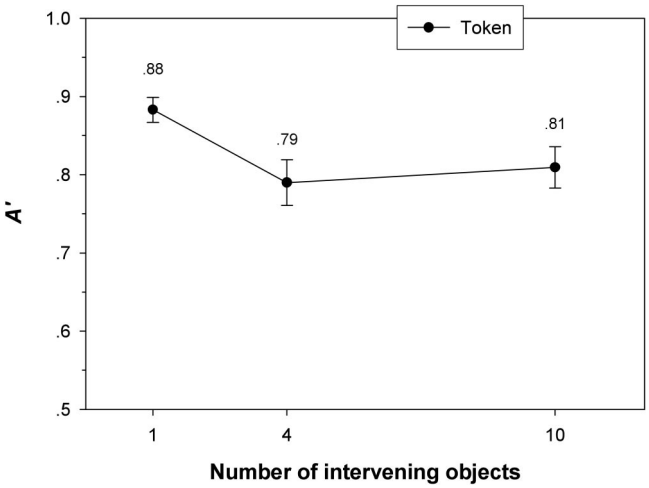

Figure 4. Experiment 2: Mean A' for token change as a function of serial position (number of objects intervening between target dot and test). Error bars represent standard errors of the means.

Method

Participants. Twenty-four new participants from the University of Iowa community completed the experiment. They either received course credit or were paid. All participants reported normal or corrected-to-normal vision. One participant did not perform above chance and was replaced.

Stimuli and apparatus. The stimuli and apparatus were the same as in Experiments 1 and 2.

Procedure. The procedure was identical to Experiment 2, with the following exception. Because only relatively recent objects were ever tested in Experiment 3, the total number of objects in the dot sequence was reduced in each scene by six. Otherwise, participants could have learned that objects visited by the dot early in the sequence were never tested, and they might have ignored them as a result. The objects visited by the dot in each scene and the sequence of dot onsets were modified to ensure a natural transition from object to object. The target objects, however, were the same as in Experiments 1 and 2. As is evident from the Experiment 3 results, these differences had little effect on the absolute levels of change detection performance.

Results and Discussion

Mean A' performance in each of the serial position conditions is displayed in Figure 5. There was a reliable effect of serial position, $F(2, 23) = 8.10$, $p < .005$. Planned comparisons of the serial position effect revealed that 0-back performance was reliably higher than 2-back performance, $F(1, 23) = 8.71$, $p < .01$; that 0-back performance was reliably higher than 4-back performance, $F(1, 23) = 14.89$, $p < .005$; and that 2-back and 4-back performance were not reliably different, $F < 1$. The recency advantage clearly held for the currently fixated object (0-back condition), but there was no statistical evidence of a recency advantage for two objects back. Taken together, the results of Experiments 1–3 suggest that the VSTM component of online scene representation may be limited to the two most recently fixated objects (the currently fixated object and one object back). The issue of the number of objects contributing to the recency advantage will be examined again in Experiment 6.

Experiment 4

So far, the recency advantage has been found at positions 0-back and 1-back, but in different experiments. Experiment 4 sought to compare 0-back and 1-back conditions directly. Rensink (2000) has argued that visual memory is limited to the currently attended object. Clearly, the accurate memory performance for objects visited 1-back and earlier (i.e., previously attended objects) in Experiments 1–3 is not consistent with this proposal (see also Hollingworth, 2003a; Hollingworth & Henderson, 2002; Hollingworth et al., 2001). Visual memory representations do not necessarily disintegrate after the withdrawal of attention. Experiment 4 examined whether there is any memory advantage at all for the currently fixated object (0-back) over a very recently attended object (1-back). In addition to 0-back and 1-back conditions, the 4-back condition was again included for comparison.

Method

Participants. Twenty-four new participants from the University of Iowa community completed the experiment. They either received course

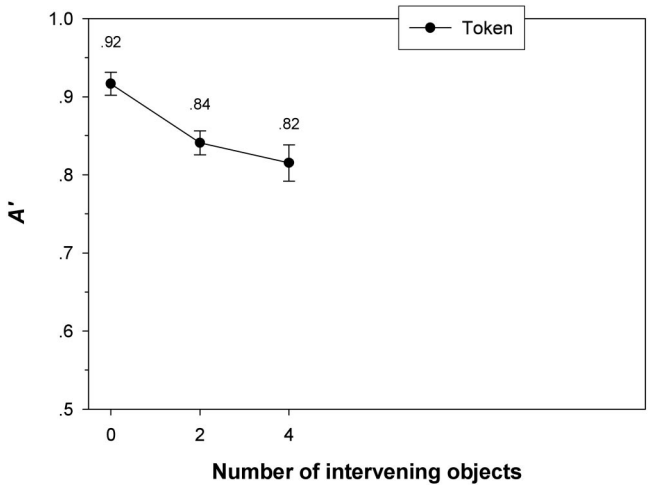

Figure 5. Experiment 3: Mean A' for token change as a function of serial position (number of objects intervening between target dot and test). Error bars represent standard errors of the means.

credit or were paid. All participants reported normal or corrected-to-normal vision. One participant did not perform above chance and was replaced.

Stimuli, apparatus, and procedure. The stimuli and apparatus were the same as in Experiments 1–3. The procedure was the same as in Experiment 3.

Results and Discussion

Mean A' performance in each of the serial position conditions is displayed in Figure 6. There was a reliable effect of serial position, $F(2, 23) = 9.92$, $p < .001$. Planned comparisons of the serial position effect revealed that 0-back performance was reliably higher than 1-back performance, $F(1, 23) = 9.31$, $p < .01$; that 0-back performance was reliably higher than 4-back performance, $F(1, 23) = 17.19$, $p < .001$; and that 1-back performance was also reliably higher than 4-back performance, $F(1, 23) = 4.19$, $p < .05$. The advantage for the 0-back over the 1-back condition demonstrates that the withdrawal of attention is accompanied by the loss of at least some visual information, but performance was still quite high after the withdrawal of attention, consistent with prior reports of robust visual memory (Hollingworth, 2003a; Hollingworth & Henderson, 2002; Hollingworth et al., 2001). In addition, the reliable advantages for 0-back and 1-back over 4-back replicate the finding that the recency effect includes the currently fixated object and one object earlier.

Experiment 5

Experiment 5 examined portions of the serial sequence relatively early in scene viewing. Experiments 1 and 2 demonstrated a trend toward higher performance at 10-back compared with 4-back. These conditions were compared in Experiment 5 with a larger group of participants to provide more power to detect a difference, if a difference exists. In addition, as a very strong test of the robustness of prerecency memory, a new condition was included in which the change detection test was delayed until the end of the session. If performance at serial positions 4-back and 10-back does indeed reflect LTM retention, then one might expect to find evidence of similar object memory over even longer retention intervals. Such memory has already been demonstrated in a free viewing paradigm (Hollingworth, 2003b), in which change detection performance was unreduced or only moderately reduced when the test was delayed until the end of the session compared with when it was administered during online viewing. Experiment 5 provided an opportunity to observe such an effect using the present dot method. In addition, the dot method provides a means to estimate the number of objects intervening between study and test for the test delayed until the end of the session. In this condition, the mean number of objects intervening between target dot and test was 402.

Method

Participants. Thirty-six new participants from the University of Iowa community completed the experiment. They either received course credit or were paid. All participants reported normal or corrected-to-normal vision.

Stimuli and apparatus. The stimuli and apparatus were the same as in Experiments 1–4.

Procedure. Because Experiment 5 tested earlier serial positions, the dot sequence from Experiments 1 and 2 was used. The procedure was identical to that in Experiment 2, except for the condition in which the test was delayed until the end of the session. In the initial session, one third of the trials were 4-back, the second third were 10-back, and the final third were not tested. For this final set of trials (delayed test condition), the dot sequence was identical to that in the 10-back condition. However, the trial simply ended after the final 800-ms view of the scene, without presentation of the target mask or the test scene.

After all 48 stimuli had been viewed in the initial session, participants completed a delayed test session in which each of the 16 scenes not tested initially was tested. For the delayed test session, each trial started with the 1,500-ms target mask image, followed by the test scene.

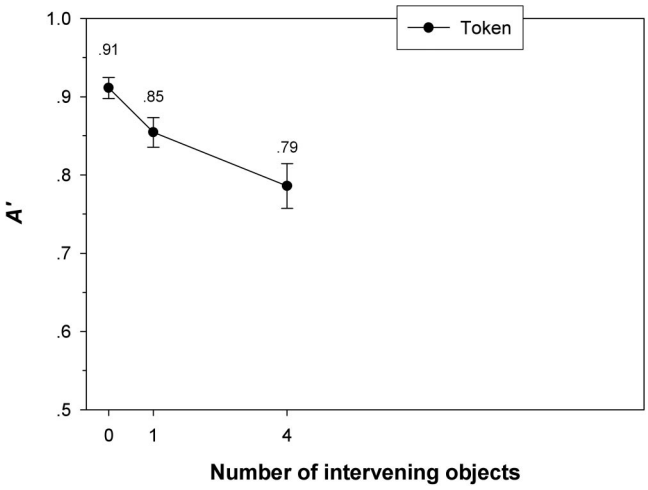

Figure 6. Experiment 4: Mean A' for token change as a function of serial position (number of objects intervening between target dot and test). Error bars represent standard errors of the means.

Participants responded to indicate whether the target had changed or had not changed from the version viewed initially. Thus, participants saw the same set of stimuli in the 10-back and delayed test conditions, except that in the latter condition, the target mask and test scene were delayed until after all scene stimuli had been viewed initially. As in previous experiments, the order of trials in the initial session was determined randomly. The order of trials in the delayed test session was yoked to that in the initial session. In the delayed test condition, the mean number of objects intervening between target dot and test was 402. The mean temporal delay was 12.1 min.

Results and Discussion

Mean A' performance in each of the serial position conditions is displayed in Figure 7. There was a reliable effect of serial position, $F(2, 23) = 4.18$, $p < .05$. Mean A' in the 4-back and 10-back conditions was identical. However, both 4-back performance and 10-back performance were reliably higher than that in the delayed test condition, $F(1, 23) = 5.29$, $p < .05$, and $F(1, 23) = 5.88$, $p < .05$, respectively.

Experiment 5 found no evidence of a difference in change detection performance between the 4-back and 10-back conditions, suggesting that there is little or no difference in the token-specific information available for an object fixated 4 objects ago versus 10 objects ago. In addition, Experiment 5 found that memory for token-specific information is quite remarkably robust. Although change detection performance was reliably worse when the test was delayed until the end of the session, it was nonetheless well above chance, despite the fact that 402 objects intervened, on average, between target viewing and test. These data complement evidence from Hollingworth (2003b; see also Hollingworth & Henderson, 2002) demonstrating that memory for previously attended objects in natural scenes is of similar specificity to memory under conditions that unambiguously require LTM, such as delay until the end of the session. Such findings provide converging evidence that the robust prerecency memory during scene viewing is indeed supported by VLTM.

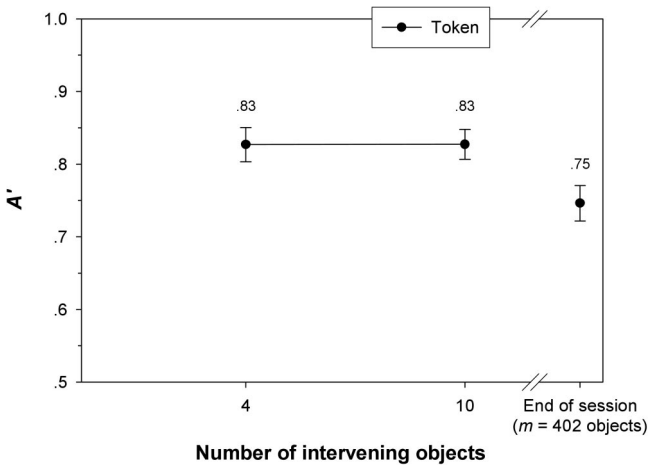

Figure 7. Experiment 5: Mean A' for token change as a function of serial position (number of objects intervening between target dot and test). Error bars represent standard errors of the means.

In addition, the results of Experiment 5 address the issue of whether LTM for scenes retains specific visual information. Most studies of scene and picture memory have used tests that were unable to isolate visual representations. The most common method has been old/new recognition of whole scenes, with different, unstudied pictures as distractors (Nickerson, 1965; Potter, 1976; Potter & Levy, 1969; Shepard, 1967; Standing, 1973; Standing et al., 1970). Participants later recognized thousands of pictures. The distractor pictures used in these experiments, however, were typically chosen to be maximally different from studied images, making it difficult to identify the type of information supporting recognition. Participants may have remembered studied pictures by maintaining visual representations (coding visual properties such as shape, color, orientation, and so on), by maintaining conceptual representations of picture identity, or by maintaining verbal descriptions of picture content. This ambiguity has made it difficult to determine whether long-term picture memory maintains specific visual information or, instead, depends primarily on conceptual representations of scene gist (as claimed by Potter and colleagues; Potter, 1976; Potter, Staub, & O'Connor, 2004). A similar problem is found in a recent study by Melcher (2001), who examined memory for objects in scenes using a verbal free recall test. Participants viewed an image of a scene and then verbally reported the identities of the objects present. Again, such a test cannot distinguish between visual, conceptual, and verbal representation.[8]

The present method, however, isolates visual memory. Distractors (i.e., changed scenes) were identical to studied scenes except for the properties of a single object. The token manipulation preserved basic-level conceptual identity, making it unlikely that a representation of object identity would be sufficient to detect the difference between studied targets and distractors. Similar memory performance has been observed for object rotation (Experiment 1, above; Hollingworth, 2003b; Hollingworth & Henderson, 2002), which does not change the identity of the target object at all. Furthermore, verbal encoding was minimized by a verbal working memory load and articulatory suppression. Thus, the present method provided a particularly stringent test of visual memory. Despite the difficulty of the task, participants remembered token-specific details of target objects across 402 intervening objects, 32 intervening scenes, and 24 intervening change detection tests, on average. Clearly, long-term scene memory is not limited to conceptual representations of scene gist. Visual memory for the details of individual objects in scenes can be highly robust.

Experiment 6

Experiments 1–5 tested serial positions of particular theoretical interest. Only a small number of serial positions could be tested in each experiment because of the limited set of scenes (48) and the

[8] Melcher (2001) did include an experiment to control for verbal encoding. In this experiment, the objects in the scene were replaced by printed words. This manipulation changed the task so drastically—instead of viewing objects in scenes, participants read words in scenes—that its value as a control is unclear.

requirement that scenes not be repeated. The combined data from the different serial positions tested in Experiments 2–5 are plotted in Figure 8. Experiment 6 sought to replicate the principal results of Experiments 1–5 with a within-participants manipulation of 10 serial positions (0-back through 9-back). A single scene item was displayed on each of 100 trials, 10 in each of the 10 serial position conditions. The scene image is displayed in Figure 9. Two token versions of 10 different objects were created. On each trial, one of the 10 objects was tested at one of the 10 possible serial positions. The token version of the 9 other objects was chosen randomly. The dot sequence was limited to this set of 10 objects, with one dot onset on each object. With the exception of the serial position of the dot on the object to be tested, the sequence of dots was generated randomly on each trial.

This method is similar to aspects of the Irwin and Zelinsky (2002) position memory paradigm, in which the same crib background and seven objects were presented on each of 147 trials. In Irwin and Zelinsky, the same object stimuli were presented on every trial; only the spatial position of each object varied. In Experiment 6, the same object types were presented in the same spatial positions on each trial; only the token version of each object varied. One issue when stimuli are repeated is the possibility of proactive interference from earlier trials. Irwin and Zelinsky found no such interference in their study; position memory performance did not decline as participants completed more trials over similar stimuli. Experiment 6 provided an opportunity to examine possible proactive interference when memory for the visual properties of objects was required.

Method

Participants. Twenty-four new participants from the University of Iowa community completed the experiment. They either received course credit or were paid. All participants reported normal or corrected-to-normal vision. One participant did not perform above chance and was replaced.

Stimuli. The workshop scene from Experiments 1–5 was modified for this experiment. Ten objects were selected (bucket, watering can, wrench, lantern, scissors, hammer, aerosol can, electric drill, screwdriver, and fire extinguisher), and two tokens were created for each. The objects and the two token versions are displayed in Figure 9. On each trial, all 10 objects were presented in the spatial positions displayed in Figure 9. Only the token version of each object varied from trial to trial. The token version of the to-be-tested object was presented according to the condition assignments described in the *Procedure* section, below. The token versions of the other 9 objects were chosen randomly on each trial.

Apparatus. The apparatus was the same as in Experiments 1–5.

Procedure. Participants were instructed in the same way as in Experiments 2–5, except they were told that the same scene image would be presented on each trial; only the object versions would vary.

There were a total of 40 conditions in the experiment: 10 (serial positions) × 2 (same, token change) × 2 (target token version initially displayed). Each participant completed 100 trials in the experimental session, 10 in each serial position condition. Half of these were same trials, and half were token change trials. The target token version condition, an arbitrary designation, was counterbalanced across participant groups. The analyses that follow collapsed this factor. A group of four participants created a completely counterbalanced design. Each of the 10 objects appeared in each condition an equal number of times.

On each trial, the sequence of events was the same as in Experiments 2–5, including the four-digit verbal working memory load. However, in Experiment 6, there was a total of 10 dot onsets on every trial, one on each of the 10 possibly changing objects. With the exception of the position of the target dot in the sequence, the sequence of dots was determined randomly. Participants first completed a practice session of 6 trials, randomly selected from the complete design. They then completed the experimental session of 100 trials. The entire experiment lasted approximately 50 min.

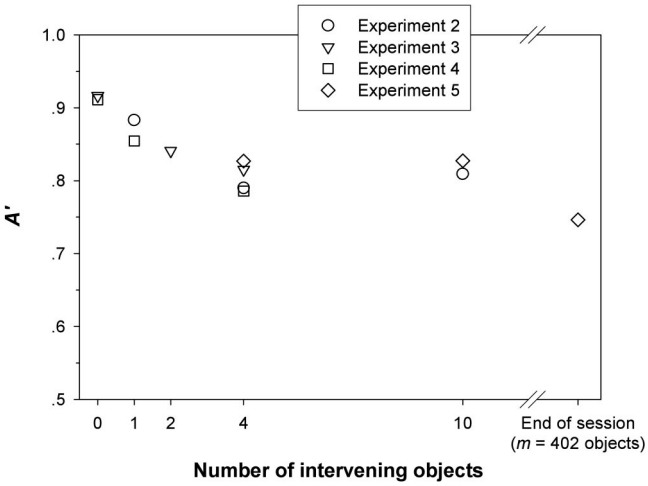

Figure 8. Compilation of results from Experiments 2–5.

Results and Discussion

Mean A' performance in each of the serial position conditions is displayed in Figure 10. There was a reliable effect of serial position, $F(9, 207) = 2.65$, $p < .01$. Planned contrasts were conducted for each pair of consecutive serial positions. A' in the 0-back condition was reliably higher than that in the 1-back condition, $F(1, 23) = 5.12$, $p < .05$, and there was a trend toward higher A' in the 1-back condition compared with the 2-back condition, $F(1, 23) = 2.41$, $p = .13$. No other contrasts approached significance: All Fs < 1, except 7-back versus 8-back, $F(1, 23) = 1.22$, $p = .28$.

The pattern of performance was very similar to that in Experiments 1–5. A reliable recency effect was observed, and this was limited, at most, to the two most recently fixated objects. In addition, prerecency performance was quite stable, with no evidence of further forgetting from serial position 2-back to 9-back. These data confirm a VSTM contribution to scene representation, responsible for the recency advantage, and a VLTM contribution, responsible for prerecency stability. Experiment 6 repeated the same scene stimulus and objects for 100 trials, yet performance was not significantly impaired relative to earlier experiments, in which scene stimuli were unique on each trial. Consistent with Irwin and Zelinsky (2002), this suggests very little proactive interference in visual memory.

The Experiment 6 results also argue against the possibility that the results of Experiments 1–5 were influenced by strategic factors based on the particular serial positions tested in each of those experiments or the particular objects chosen as targets. In Exper-

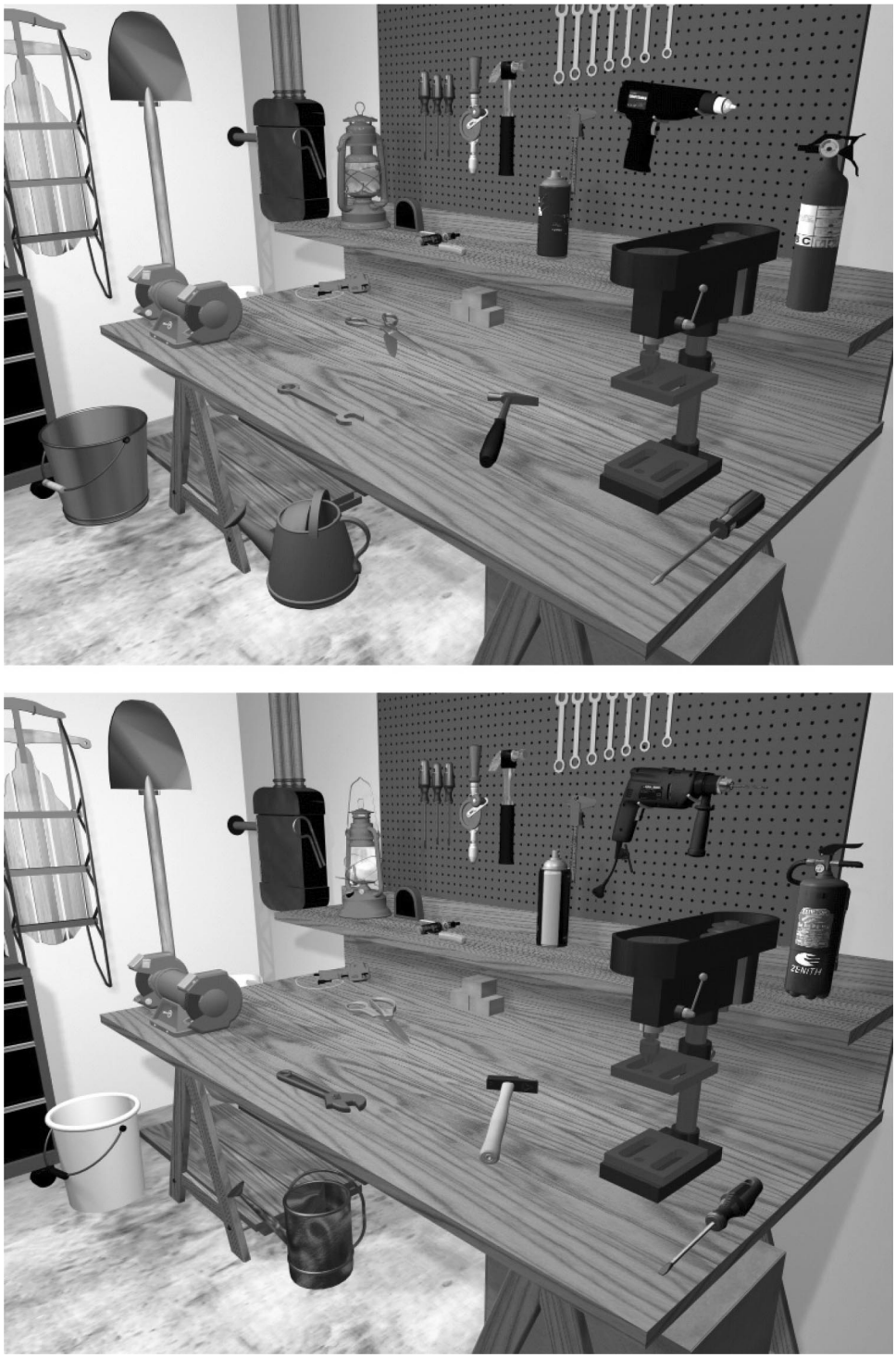

Figure 9. Scene stimulus used in Experiment 6. The two panels show the two token versions of the 10 potentially changing objects (bucket, watering can, wrench, lantern, scissors, hammer, aerosol can, electric drill, screwdriver, and fire extinguisher).

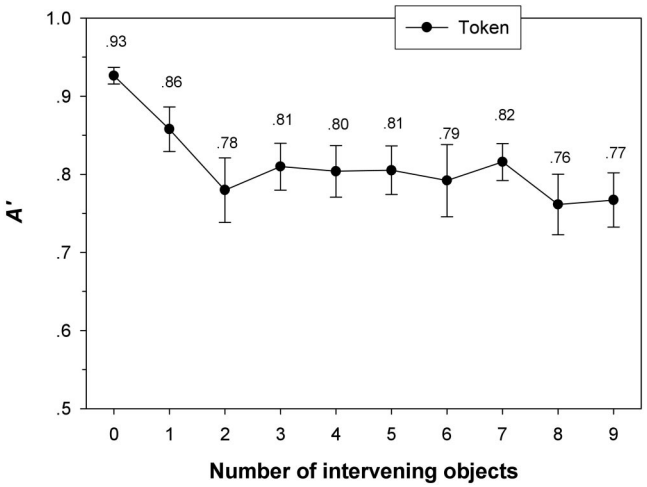

Figure 10. Experiment 6: Mean *A′* for token change as a function of serial position (number of objects intervening between target dot and test). Error bars represent standard errors of the means.

iment 6, each of the 10 objects visited by the dot was equally likely to be tested, and each of the 10 serial positions was also equally likely to be tested. Therefore, there was no incentive to preferentially attend to any particular object or to bias attention to any particular serial position or set of positions. Yet, Experiment 6 replicated all of the principal results of Experiments 1–5.

General Discussion

Experiments 1–6 demonstrate that the accumulation of visual information from natural scenes is supported by VSTM and VLTM. The basic paradigm tested memory for the visual properties of objects during scene viewing, controlling the sequence of objects attended and fixated within each scene. On each trial of this follow-the-dot paradigm, participants followed a neon-green dot as it visited a series of objects in a scene, shifting gaze to fixate the object most recently visited by the dot. At the end of the sequence, a single target object was masked in the scene, followed by a forced-choice discrimination or change detection test. The serial position of the target object in the sequence was manipulated. Object memory was consistently superior for the two most recently fixated objects, the currently fixated object and one object earlier. This recency advantage indicates a VSTM component to online scene representation. In addition, objects examined earlier than the two-object recency window were nonetheless remembered at rates well above chance, and there was no evidence of further forgetting with more intervening objects. This robust prerecency performance indicates a VLTM component to online scene representation.

Theories claiming that visual memory makes no contribution to scene representation (O'Regan, 1992) or that visual object representations disintegrate on the withdrawal of attention (Rensink, 2000) cannot account for the present data because accurate memory performance was observed for objects that had been, but were no longer, attended when the test was initiated (see also Hollingworth, 2003a, 2003b; Hollingworth & Henderson, 2002). Experiment 4 did find evidence that the currently fixated object was remembered more accurately than the object fixated one object earlier, so the withdrawal of attention from an object is at least accompanied by the loss of some visual information.

In addition, the present results demonstrate that online visual scene representations retain visual information that exceeds the capacity of VSTM. In particular, performance in the early serial positions—such as 10-back in Experiments 1, 2, and 5—exceeded maximum predicted performance based on the hypothesis that visual scene representation is limited to VSTM (Irwin & Andrews, 1996; Irwin & Zelinsky, 2002). The logical conclusion is that this extra memory capacity for the visual form of objects reflects the contribution of VLTM. Furthermore, the VLTM component exhibits exceedingly large capacity and very gradual forgetting, as memory performance remained well above chance when the test was delayed until the end of the experimental session, a condition in which an average of 402 objects intervened between target examination and test.

Together, these data support the claim that both VSTM and VLTM are used to construct scene representations with the capability to preserve visual information from large numbers of individual objects (Hollingworth, 2003a, 2003b; Hollingworth & Henderson, 2002). Under this visual memory theory of scene representation, during a fixation on a particular object, complete and precise sensory representations are produced across the visual field. In addition, a higher level visual representation, abstracted away from precise sensory information, is constructed for the attended object. When the eyes are shifted, the sensory information is lost (Henderson & Hollingworth, 2003c; Irwin, 1991). However, higher level visual representations survive shifts of attention and the eyes and can therefore support the accumulation of visual information within the scene. Higher level visual representations are maintained briefly in VSTM. Because of capacity limitations, only the two most recently attended objects occupy VSTM. Higher level visual representations are also maintained in VLTM, and VLTM has exceedingly large capacity, supporting the accumulation of information from many individual objects as the eyes and attention are oriented from object to object within a scene.

The present finding of a VSTM component to online scene representation, preserving information about the visual form of individual objects, complements evidence of an STM component to online memory for the spatial position of objects in scenes (Irwin & Zelinsky, 2002; Zelinsky & Loschky, 1998). Taken together, these results are consistent with the possibility that objects are maintained in VSTM, and perhaps also in VLTM (Hollingworth & Henderson, 2002), as episodic representations binding visual information to spatial position, that is, as object files (Hollingworth & Henderson, 2002; Irwin, 1992a; Kahneman, Treisman, & Gibbs, 1992; Wheeler & Treisman, 2002[9]). However, further experimental work manipulating visual object properties,

[9] Note that Wheeler and Treisman (2002) stressed the fragility of visual–spatial binding in VSTM and its susceptibility to interference from other perceptual events requiring attention. This emphasis is a little puzzling considering that memory for binding in their study was generally very good, with memory for the binding of visual and spatial information equivalent to or only slightly less accurate than memory for either visual or spatial information alone.

spatial position, and the binding of the two is needed to provide direct evidence that object representations in scenes bind visual information to spatial positions.

Recency effects provide evidence of a VSTM component to scene representation, but exactly how are VSTM representations to be distinguished from VLTM representations? There is a very clear dissociation between VSTM and sensory persistence (iconic memory) in terms of format and content (abstracted vs. sensory–pictorial), capacity (limited vs. large capacity), and time course (relatively robust vs. fleeting). The distinction between VSTM and VLTM, however, is not quite as clear cut. The format of visual representations retained over the short and long terms appears to be quite similar. Visual representations stored over the short term (e.g., across a brief ISI or saccadic eye movement) are sensitive to object token (Henderson & Hollingworth, 2003a; Henderson & Siefert, 2001; Pollatsek, Rayner, & Collins, 1984), orientation (Henderson & Hollingworth, 1999, 2003a; Henderson & Siefert, 1999, 2001; Tarr, Bülthoff, Zabinski, & Blanz, 1997), and the structural relationship between object parts (Carlson-Radvansky, 1999; Carlson-Radvansky & Irwin, 1995) but are relatively insensitive to absolute size (Pollatsek et al., 1984) and precise object contours (Henderson, 1997; Henderson & Hollingworth, 2003c). Similarly, visual representations retained over the long term (e.g., in studies of object recognition) show sensitivity to object token (Biederman & Cooper, 1991), orientation (Tarr, 1995; Tarr et al., 1997), and the structural relationship between object parts (Palmer, 1977) but are relatively insensitive to absolute size (Biederman & Cooper, 1992) and precise object contours (Biederman & Cooper, 1991). Short-term visual memory and long-term visual memory are clearly distinguishable in terms of capacity, however. Whereas VSTM has a limited capacity of three to four objects at maximum (Luck & Vogel, 1997; Pashler, 1988), VLTM has exceedingly large capacity, such that token change detection performance in the present study was still well above chance after 402 intervening objects, on average. Finally, VLTM representations can be retained over very long periods of time. In the picture memory literature, picture recognition remains above chance after weeks of delay (Rock & Engelstein, 1959; Shepard, 1967). Thus, there are also clear differences in the time course of retention in VSTM and VLTM.

Each of these three issues—format, capacity, and time course—deserves further consideration. If the format and content of VSTM and VLTM representations are similar, what then accounts for the recency advantage itself? The present paradigm was not designed to directly compare the representational format of VSTM and VLTM. One clear possibility, however, is that although VSTM and VLTM maintain representations of similar format, VSTM representations are more precise than VLTM representations. Support for this possibility comes from experiments examining VSTM as a function of retention interval (Irwin, 1991; Phillips, 1974). Such studies have consistently observed that memory performance declines with longer retention intervals, suggesting loss of information from VSTM during the first few seconds of retention, even without interference from subsequent stimuli (see also Vandenbeld & Rensink, 2003). Similar loss of relatively precise information in VSTM may explain the present recency advantage and the rapid decline to prerecency levels of performance.

The similar representational format in VSTM and VLTM also prompts consideration of the degree of independence between visual memory systems. Again, any discussion of such an issue must be speculative at present, given the paucity of evidence on the subject. However, the similar representational format does raise the possibility that VSTM may constitute just the currently active portion of VLTM, as proposed by some general theories of working memory (Lovett, Reder, & Lebiere, 1999; O'Reilly, Braver, & Cohen, 1999). However, can VSTM be just the activated contents of VLTM? It is unlikely that VSTM is entirely reducible to the activation of preexisting representations in VLTM when one considers that entirely novel objects can be maintained in VSTM (see, e.g., Phillips, 1974; Tarr et al., 1997). VSTM may represent novel objects by supporting novel conjunctions of visual feature codes. As an example, object shape may be represented as a set of 3-D or 2-D components (Biederman, 1987; Riesenhuber & Poggio, 1999). The shape primitives would clearly be VLTM representations, but they can be bound in VSTM in novel ways to produce representations of stimuli with no preexisting VLTM representation. Once constructed in VSTM, the new object representation may then be stored in VLTM. This view is consistent with theories stressing the active and constructive nature of working memory systems (Baddeley & Logie, 1999; Cowan, 1999).

The present study found that performance attributable to VLTM was observed at fairly recent serial positions. For the 2-back condition, in which performance was no higher than at earlier serial positions, the delay between target fixation and test was only 3,700 ms. If 2-back performance is indeed supported by VLTM, this would suggest that VLTM representations set up very quickly indeed. A retention interval of 3.7 s is significantly shorter than some retention intervals in studies seeking to examine VSTM (Irwin, 1991; Phillips, 1974; Vogel et al., 2001). In addition, the present data do not preclude the possibility that VLTM representations are established even earlier than two objects back. So, although VLTM clearly dissociates from VSTM when considering very long-term retention (over the course of days or weeks), the distinction is much less clear when considering retention over the course of a few seconds.

Previous studies examining VSTM have not typically considered the potential contribution of LTM to task performance or even the distinction between VSTM and VLTM (see Phillips & Christie, 1977, for a prominent exception). VSTM was originally defined as a separate memory system not with respect to LTM but rather with respect to sensory persistence, or iconic memory (see, e.g., Phillips, 1974). Subsequent studies examining VSTM have used retention intervals, typically on the order of 1,000 ms (Jiang, Olson, & Chun, 2000; Luck & Vogel, 1997; Olson & Jiang, 2002; Vogel et al., 2001; Wheeler & Treisman, 2002; Xu, 2002a, 2002b), that exceed the duration of sensory persistence but fit within intuitive notions of what constitutes the short term. Given the present evidence that VLTM representations are established very quickly, it is a real possibility that performance in studies seeking to examine VSTM have reflected both VSTM and VLTM retention, overestimating the

capacity of VSTM. As in Phillips and Christie (1977), the present serial examination paradigm provided a means to isolate the VSTM contribution to object memory. The recency advantage was limited to the two most recently fixated objects in the present study and to the very last object in Phillips and Christie, suggesting that the true capacity of VSTM may be smaller than three to four objects. However, any direct comparison between capacity estimates based on simple stimuli (see, e.g., Vogel et al., 2001) and complex objects, as in the present study, must be treated with caution, especially given evidence that more complex, multipart objects are not retained as efficiently as simple, single-part objects (Xu, 2002b).

Conclusion

The accumulation of visual information during scene viewing is supported by two visual memory systems: VSTM and VLTM. The VSTM component appears to be limited to the two most recently fixated objects. The VLTM component exhibits exceedingly large capacity and gradual forgetting. Together, VSTM and VLTM support the construction of scene representations capable of maintaining visual information from large numbers of individual objects.

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.

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

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and performance VI* (pp. 646–667). Hillsdale, NJ: Erlbaum.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge University Press.

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 176–189.

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

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

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

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

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes. *Cognitive Psychology, 6,* 173–189.

Broadbent, D. E., & Broadbent, M. H. P. (1981). Recency effects in visual memory. *Quarterly Journal of Experimental Psychology, 33A,* 1–15.

Carlesimo, G. A., Marfia, G. A., Loasses, A., & Caltagirone, C. (1996). Recency effect in anterograde amnesia: Evidence for distinct memory stores underlying enhanced retrieval of terminal items in immediate and delayed recall paradigms. *Neuropsychologia, 34,* 177–184.

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

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

Coltheart, M. (1980). The persistences of vision. *Philosophical Transactions of the Royal Society of London, Series B, 290,* 269–294.

Cowan, N. (1999). An embedded process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York: Cambridge University Press.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition, 21,* 142–145.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109,* 75–97.

Glanzer, M. (1972). Storage mechanisms in recall. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 129–193). New York: Academic Press.

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.

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

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

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

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

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

Hollingworth, A. (2003b). *The relationship between online visual representation of a scene and long-term scene memory*. Manuscript submitted for publication.

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., 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.

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

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

Korsnes, M. S. (1995). Retention intervals and serial list memory. *Perceptual and Motor Skills, 80,* 723–731.

Logie, R. H. (1995). *Visuo-spatial working memory.* Hove, England: Erlbaum.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). New York: Cambridge University Press.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

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

Melcher, D. (2001, July 26). Persistence of visual memory for scenes. *Nature, 412,* 401.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition, 21,* 689–698.

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.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong object" hypothesis. *Perception & Psychophysics, 64,* 1055–1067.

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'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). New York: Cambridge University Press.

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

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

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

Pashler, H., & Carrier, M. (1996). Structures, processes, and the "flow of information." In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 3–29). San Diego, CA: Academic Press.

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

Phillips, W. A. (1983). Short-term visual memory. *Philosophical Transactions of the Royal Society of London, Series B, 302,* 295–309.

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.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81,* 10–15.

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 478–489.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1163–1175.

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.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Riggs, L. A. (1965). Visual acuity. In C. H. Graham (Ed.), *Vision and visual perception* (pp. 321–349). New York: Wiley.

Rock, I., & Engelstein, P. (1959). A study of memory for visual form. *American Journal of Psychology, 72,* 221–229.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition, 7,* 377–396.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200.

Shaffer, W. O., & Shiffrin, R. M. (1972). Rehearsal and storage of visual information. *Journal of Experimental Psychology, 92,* 292–296.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22,* 261–273.

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

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 2500 visual stimuli. *Psychonomic Science, 19,* 73–74.

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

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

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9,* 379–385.

Vallar, G., Papagno, C., & Baddeley, A. D. (1991). Long-term recency effects and phonological short-term memory: A neuropsychological case study. *Cortex, 27,* 323–326.

Vandenbeld, L., & Rensink, R. A. (2003, May). *The decay characteristics of size, color and shape information in visual short-term memory.* Poster presented at the Third Annual Meeting of the Vision Sciences Society, Sarasota, FL.

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.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64.

Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985, July 19). Memory processing of serial lists by pigeons, monkeys, and people. *Science, 229,* 287–289.

Xu, Y. (2002a). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics, 64,* 1260–1280.

Xu, Y. (2002b). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 458–468.

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.

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

Appendix

Mean Percentage Correct Data for Experiments 2–6

Serial position condition	Same	Token change
Experiment 2		
1-back	74.0	87.0
4-back	70.8	73.5
10-back	73.0	74.5
Experiment 3		
0-back	77.1	93.8
2-back	69.8	81.3
4-back	70.3	77.1
Experiment 4		
0-back	75.0	93.2
1-back	63.5	88.5
4-back	63.0	79.7
Experiment 5		
4-back	70.5	80.6
10-back	73.6	77.4
End of session	70.5	64.9
Experiment 6		
0-back	82.8	89.2
1-back	79.6	77.8
2-back	75.9	72.4
3-back	72.9	71.3
4-back	68.1	74.4
5-back	73.2	74.2
6-back	66.6	74.8
7-back	71.5	73.5
8-back	68.0	70.0
9-back	75.3	66.0

Received June 24, 2003
Revision received November 17, 2003
Accepted November 17, 2003 ■