VISUAL COGNITION, 2006, 14 (4/5/6/7/8), 781–807

Psychology Press
Taylor & Francis Group

Visual memory for natural scenes: Evidence from change detection and visual search

Andrew Hollingworth

University of Iowa, Iowa City, IA, USA

This paper reviews research examining the role of visual memory in scene perception and visual search. Recent theories in these literatures have held that coherent object representations in visual memory are fleeting, disintegrating upon the withdrawal of attention from an object. I discuss evidence demonstrating that, far from being transient, visual memory supports the accumulation of information from scores of individual objects in scenes, utilizing both visual short-term memory and visual long-term memory. In addition, I review evidence that memory for the spatial layout of a scene and memory for specific object positions can efficiently guide search within natural scenes.

In the past decade, the interaction between perception and memory has received a great deal of attention from cognitive scientists. Much of this interest has originated from increased understanding that perception is a dynamic, serial process, extended over space and time. In this paper, I will discuss two related lines of research in which the relationship between perception and memory has come to the fore: Scene perception and visual search. While viewing natural scenes, the eyes shift (via saccadic eye movements) approximately three times each second to bring different scene regions onto the fovea, where visual acuity is highest (see Henderson & Hollingworth, 1998, for a review). Across saccades, visual encoding is suppressed (Matin, 1974), dividing visual input into discrete episodes. Given the discrete, serial nature of scene perception, memory is required to retain and accumulate visual information from local objects as the eyes and attention are oriented from object-to-object within a scene. Similarly, visual search typically requires the serial allocation of attention to individual objects in the course of finding a target and rejecting distractors (Woodman & Luck, 2003). Roles for memory in search include keeping track of objects

---

Please address all correspondence to Andrew Hollingworth, Department of Psychology, University of Iowa, 11 Seashore Hall E, Iowa City, IA 52242-1407, USA.
E-mail: andrew-hollingworth@uiowa.edu

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

http://www.psypress.com/viscog DOI: 10.1080/13506280500193818

that have already been examined (e.g., Klein, 1988) and guiding attention to targets that appear in predictable locations (e.g., Chun & Jiang, 1998).

The following discussion of scene perception, search, and visual memory will be placed within the context of recent claims that visual memory is transient, playing little or no role in the representation of natural scenes or in search (Rensink, 2000; Wolfe, 1999). I will review evidence that, contrary to these claims, visual memory supports robust accumulation of visual information from scores of individual objects in scenes and that visual memory can efficiently guide search within real-world environments.

SCENE PERCEPTION, CHANGE BLINDNESS, AND VISUAL MEMORY

The phenomenon of change blindness has shaped recent thinking on the role of visual memory in scene perception. In change blindness studies, participants often fail to detect otherwise salient changes when detection depends on visual memory. Dependence on memory has been achieved either by introducing an interstimulus interval (ISI) between differing images (e.g., Rensink, O'Regan, & Clark, 1997), by introducing a change to an image during an eye movement (e.g., Grimes, 1996; Henderson & Hollingworth, 1999, 2003c), or by occluding a change with a physical object (e.g., Simons & Levin, 1998). In perhaps the most well-known change blindness paper, Rensink et al. (1997) presented photographs of real-world scenes and introduced a change to a portion of the image on each trial, such as the deletion of an airplane's engine in an airport scene. Each image in this *flicker paradigm* was presented for 240 ms, with an 80 ms neutral grey ISI. The change was repeated by alternating the two images until the participant detected the change (i.e., the airplane's engine would disappear, then reappear, then disappear, and so on). Rensink et al. found that for many changes, participants required extended viewing (often more than 30 s) before they detected the change. Researchers have concluded from such effects that very little visual information (at an extreme, no visual information, O'Regan, 1992; O'Regan & Noë, 2001) is retained from one view of a scene to the next, and therefore that visual representations of complex, natural scenes must be impoverished (Becker & Pashler, 2002; Irwin & Andrews, 1996; Rensink, 2000, 2002; Rensink et al., 1997; Simons, 1996; Simons & Levin, 1997; Wheeler & Treisman, 2002; Wolfe, 1999).

Rensink (2000, 2002) has provided the most theoretically elaborated account of this *visual transience* hypothesis. Rensink's view, which he terms *coherence theory*, provides a broad account of vision, attention, and memory. Coherence theory can be distilled into the following set of claims. First, low-level vision produces a description of the visual field in terms of *proto-objects*. Proto-objects, according to Rensink, are the earliest form of object representation in the visual system, in which local regions of the visual field are parsed, and sensory features corresponding to individual objects are loosely assembled. Proto-objects are computed in a bottom-up fashion, in parallel across the visual field, and independently of attention. They are also highly volatile: Proto-objects decay very quickly after visual stimulation is removed and are highly susceptible to interference from new sensory information (i.e., they are susceptible to backward masking, which Rensink terms *overwriting*). In the context of change detection, proto-objects are not by themselves sufficient to detect a change, because they will typically have decayed prior to the appearance of the changed image, or if they have not yet decayed, sensory processing of the new image will mask them (overwrite them), leaving the visual system with no informational basis upon which to detect the change.

Further, coherence theory holds that visual attention is critical to the consolidation of proto-objects into a coherent, robust representation that does not immediately decay and is not susceptible to backward masking. When attention selects a set of proto-objects that correspond to an object in the visual field, the object's features, which were only loosely assembled prior to the allocation of attention, are bound into a coherent object representation (which Rensink terms a *nexus*). Activation from the nexus feeds back to the proto-objects, and this recurrent flow of information to the proto-object level allows proto-objects to be maintained robustly across delays and to resist masking from subsequent sensory input. Rensink terms the nexus plus stabilized proto-objects a *coherence field*. Under coherence theory, it is the continued maintenance of these proto-objects in the coherence field that allows one to perceive a change to an attended object in a scene across disruptions such as a brief ISI or eye movement.

Finally, coherence theory holds that once attention is removed from an object, the coherent object representation (the nexus) comes unbound, the recurrent activation from the nexus to the proto-objects is lost, and the proto-objects return to their original state as fleeting and susceptible to masking. Thus, perceiving a change to a previously attended object is equivalent to perceiving a change to an object that has never been attended. In neither case is there a coherent representation to support change detection: "After focused attention is released, the object loses its coherence and dissolves back into its constituent proto-objects. There is little or no 'after effect' of having been attended" (Rensink, 2000, p. 20). Although Rensink claims that coherent visual representation is limited to the currently attended object, he does allow that other forms of representation may be retained robustly from a complex scene. In particular, the gist (or basic identity of a scene, such as "kitchen" or "airport") is remembered robustly, as well as the abstract spatial organization of the scene, or layout. But neither of these representations preserves information about the visual details of individual objects in the scene.

In summary, coherence theory claims that a coherent visual representation sufficient to support change detection is available only for the currently attended object. When attention is withdrawn from an object, feature binding comes undone, and any coherent object representation dissolves back into its constituent features. As a result, visual representations do not accumulate as attention is oriented from object-to-object within a scene, and visual scene representations are therefore impoverished, leading to change blindness. These claims are not unique to coherence theory. Many other researchers have made similar proposals regarding the transience of visual memory. Some prominent examples include the following.

- "Binding information [in memory] can be lost when new visual objects are presented and attention is withdrawn, causing bound objects to fall apart. Bound visual objects may survive in memory across distraction only when they are recoded into a nonvisual form, such as a verbal label." (Wheeler & Treisman, 2002, p. 62)
- "... when a scene is viewed, observers create two representations. One contains the gist, or meaning, of a scene, and the other represents the visual details of a small portion of the scene. The gist representation is thought to be relatively stable, and more conceptual than visual in nature (Wolfe, 1998). By contrast, the representation of visual details is thought to be volatile and fleeting. At any instant, observers represent only the small portion of the visual observer's attention shifts, so do the contents of VSTM [visual short-term memory], leaving no memory or representation of the previously attended visual stimuli ...." (Becker & Pashler, 2002, p. 744)
- "During any fixation, we have a rich visual experience. From that visual experience, we abstract the meaning or gist of a scene. During the next visual fixation, we again have a rich visual experience, and if the gist is the same, our perceptual system assumes the details are the same." (Simons & Levin, 1997, p. 267)
- "When attention is deployed elsewhere, the visual representation of an object appears to revert to its preattentive state." (Wolfe, 1999, p. 78)

As is evident from these statements, the concept of visual transience and its consequences (such as visually impoverished representations of scenes) have been highly influential in recent thinking on vision and memory. Thus, the visual transience view deserves close scrutiny. My discussion will focus on Rensink's coherence theory, because it is the most prominent and most clearly specified visual transience theory, but coherence theory stands for a larger class of theory that has been widely accepted within the vision community.

Visual memory systems

Before discussing whether visual memory is indeed transient, it is necessary to briefly review current knowledge regarding the properties of visual memory. Visual memory appears to be composed of four different memory stores: Visible persistence, informational persistence, visual short-term memory (VSTM), and visual long-term memory (VLTM) (see Irwin, 1992b, for an excellent review). Visible and informational persistence are often grouped together as *iconic memory* or, preferably, *sensory persistence* (Coltheart, 1980). Visible persistence and informational persistence preserve a precise, high-capacity, sensory trace that is generated across the visual field but is highly volatile. Visible persistence, as the name suggests, is phenomenologically visible (that is, one sees a stimulus as visibly persisting after it has been removed). The duration of visible persistence is extraordinarily brief, decaying within approximately 80–100 ms after the *onset* of a stimulus (Di Lollo, 1980). Informational persistence is a nonvisible sensory trace that persists for approximately 150–300 ms after stimulus offset (Irwin & Yeomans, 1986). Both visible persistence and informational persistence are highly susceptible to interference from new sensory processing; they are susceptible to backward masking. VSTM maintains visual representations abstracted away from precise sensory information. It has a limited capacity of 3–4 objects (Irwin, 1992a; Luck & Vogel, 1997; Pashler, 1988) and less spatial precision than point-by-point sensory persistence (Irwin, 1991; Phillips, 1974). However, VSTM is not significantly disrupted by backward masking (Pashler, 1988; Phillips, 1974) and can be maintained over durations on the order of seconds (Phillips, 1974) and across saccades (Irwin, 1992a). VLTM appears to maintain visual representations similar to those maintained in VSTM (see Hollingworth, 2004) but with the capability to accumulate visual information from scores of individual objects (Hollingworth, 2004, 2005b). Note that of the four visual memory stores, only visible persistence directly supports visual phenomenology. Other forms of visual memory certainly maintain visual information, but they do not directly support visual experience. With the exception of extraordinarily brief visible persistence, visual memory is nonvisible.

Aligning coherence theory with the visual memory literature, it is clear that proto-objects in coherence theory map onto sensory persistence (i.e., visible persistence and/or informational persistence). Both proto-objects and sensory persistence are low-level visual representations generated in parallel across the visual field independently of attention, both are highly volatile, and both are susceptible to backward masking. Further, the coherence field (the nexus plus the stabilized proto-objects) directly maps onto VSTM (Rensink, 2000).[1] Thus, coherence theory can be rephrased in the following manner using terminology from the visual memory literature. When looking at a visual image, low-level sensory representations are generated across the visual field. When the image is removed, low-level sensory (iconic) persistence is fleeting and is highly susceptible to masking. However, the visual system can consolidate a small number of visual objects into a more stable store, VSTM, which can then be maintained in the service of explicit report or comparison in a change detection task.

From this perspective, visual transience theories are perfectly consistent with existing visual memory research. Early research on sensory memory (Averbach & Coriell, 1961; Sperling, 1960) found that for briefly presented arrays of letters, low-level sensory representations were generated across the visual field, but after stimulus removal, these low-level representations were fleeting and highly susceptible to masking. Approximately three or four letters, however, could be attended and consolidated into a more stable memory store (which we now term VSTM) that could support letter report at longer delays. The transience of sensory representations has been a background assumption in the visual memory literature for the last 40 years. Thus, the novelty of visual transience theories lies not in the claim that sensory representations are fleeting, but rather in the claim that after an object is attended and consolidated into VSTM, that object representation comes unbound when attention is withdrawn, leaving no trace of the coherent object representation that had been previously formed. Under visual transience theories, it is this absence of visual accumulation that explains poor detection in change blindness studies. One must be attending the object that changes in order to detect the change, because a coherent object representation is maintained only for the currently attended object.

From this review of visual memory systems, it is clear that the only plausible candidates for the accumulation of visual information during scene viewing are VSTM and VLTM; visible and informational persistence decay too quickly and are highly susceptible to masking. The critical question,

---

[1] The coherence theory view of VSTM differs from standard models in that coherence theory claims low-level sensory representations (proto-objects) constitute a component of VSTM; attention allows the continued maintenance of proto-objects for the attended object in VSTM (Rensink, 2000). However, a great deal of evidence demonstrates that even for attended objects, VSTM is abstracted away from precise sensory information (Henderson, 1997; Henderson & Hollingworth, 2003b; Irwin, 1991; Phillips, 1974). For example, Phillips (1974) presented single, checkerboard objects at fixation in a change detection task. The checkerboard object was clearly attended, since it was the only stimulus on display. Yet, Phillips found that high-capacity sensory persistence was fleeting, and that VSTM maintained representations abstracted away from sensory persistence, even for an attended object.

then, is whether VSTM and VLTM are indeed used to accumulate visual information from individual objects as the eye and attention are oriented from object-to-object within a scene, or whether visual object representations come unbound upon the withdrawal of attention, with little or no accumulation of visual information.

Robust visual memory for objects in natural scenes

My colleagues and I have conducted a series of studies to answer this question (see Henderson & Hollingworth, 2003b, for an earlier review). Hollingworth and Henderson (2002; Hollingworth, Williams, & Henderson, 2001b) examined the basic issue of whether coherent visual object representations can be maintained after the withdrawal of attention. In these experiments, eye movements were monitored while participants viewed computer-generated depictions of real-world scenes. Figure 1 shows a sample scene. The computer waited until the participant had directly fixated a target object in the scene (to ensure it had been attended). Subsequently, the target object was changed during a saccade to a different (nontarget) object in the scene. Because visual attention is automatically and exclusively allocated to the goal of a saccade prior to the initiation of that eye movement (e.g., Hoffman & Subramaniam, 1995), the target object was no longer attended when the change occurred; attention had shifted to the nontarget object that was the goal of the saccade. The target object was changed either by rotating it 90° in depth or by replacing it with another object from the same basic-level category (token change). Rotation change detection, in particular, required memory for the visual details of the target, since the changed target differed from the original target only in orientation. Coherence theory predicts that these object changes should not have been

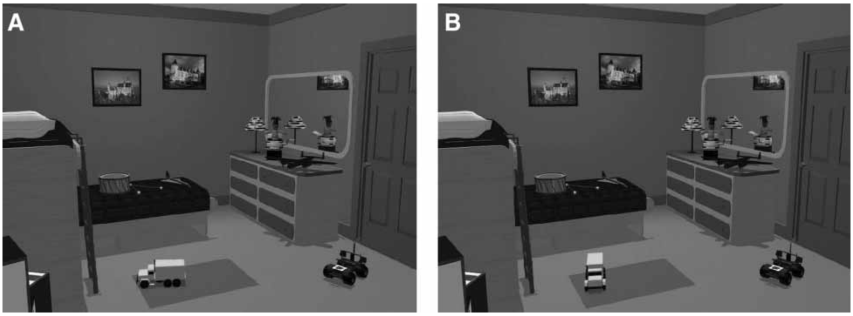

Figure 1. Sample stimuli from studies of visual memory and scene perception (Hollingworth, 2003a, 2004, 2005b; Hollingworth & Henderson, 2002). Panel A shows the initial scene. Panel B shows an object change (rotation in depth of the toy truck).

detectable, because attention had been withdrawn from the target object prior to the change. Yet, participants were able to successfully detect token and rotation changes on a significant proportion of trials, demonstrating that visual memory accumulates visual representations from previously attended objects in scenes.

In a converging experiment (Hollingworth & Henderson, 2002), a previously attended target object was masked during a saccade to a different object in the scene. Two object alternatives were then displayed sequentially within the scene. One was the original target, and the other was either a different token or different orientation distractor. Despite the fact that attention was no longer directed to the target when it was masked, participants performed the discrimination tasks at rates above 80% correct. Further, accurate discrimination performance was observed even when many fixations on other objects intervened between target fixation and test. When more than nine fixations on other objects intervened between target fixation and test, token discrimination performance was 85.3% correct and orientation discrimination performance was 92.3% correct. Memory for the visual details of previously attended objects was clearly robust across shifts of attention and of the eyes.

The experiments in Hollingworth and Henderson (2002) depended on the relationship between eye position and attention (that attention covertly precedes the eyes to a saccade target) to ensure that tested objects were not currently attended at test. Hollingworth (2003a) used a converging method to control the allocation of attention. Participants viewed a scene for 20 s. Then, a bright green dot appeared abruptly in the scene at a location different from that of the target object. The target object was then masked, and the mask was removed to reveal either the target object changed (rotation or token change) or unchanged. The task was change detection. Given evidence that abruptly appearing objects capture attention (e.g., Yantis & Jonides, 1984), attention should have been allocated to the dot onset, and not to the target, when the target was masked. A control experiment demonstrated that participants did indeed shift attention to the onset dot. Finally, a four-digit verbal working memory load minimized the possibility of verbal encoding. Despite the fact that the target was not currently attended, change detection performance was very high indeed. Percentage correct data were used to calculate A', a signal detection measure that models proportion correct in a two-alternative forced choice paradigm and varies from .5 (chance) to 1.0 (perfect sensitivity) (Grier, 1971). Mean A' was .91 for token change detection and .87 for rotation change detection, consistent with the results of Hollingworth and Henderson. Again, visual representations sufficient to make subtle judgements were retained reliably after the withdrawal of attention.

Accurate memory for the visual form of objects in these studies was almost certainly due to the retention of abstracted visual representations in VSTM, in VLTM, or in both. To examine the relative contributions of VSTM and VLTM to the online visual representation of natural scenes, Hollingworth (2004) used a serial position manipulation to control the sequence of objects fixated and attended within a scene. On each trial of this *follow-the-dot* paradigm, participants followed a small, bright green dot as it visited a series of objects in a scene, shifting gaze to fixate the object most recently visited by the dot. Each object in a scene was visited once, and the sequence of objects was designed to mimic a natural eye movement scan path on the scene. The serial position of a target object in the sequence was manipulated. The dot could appear on the target relatively early in viewing or relatively late. After the sequence was completed, the target object was masked, and the mask was then removed to reveal either the original target object or a different object token.[2] Again, the task was change detection.

If VSTM contributes to online scene representation, then one would expect the objects attended most recently before the test to be remembered most accurately, a recency effect characteristic of retention in short-term memory (Murdock, 1962; Phillips & Christie, 1977). If VLTM contributes to online scene representation, then one would expect memory for objects fixated early in viewing to be consistently above chance and to reflect retention beyond typical 3–4 object estimates of VSTM capacity. This is exactly what was found. The basic pattern of results from Hollingworth (2004) is depicted in Figure 2. Object memory was consistently superior for the two objects fixated most recently before the test. This recency advantage indicates a VSTM component to online scene representation, apparently limited to two objects.[3] Objects examined earlier than two-objects before the test were nonetheless remembered at rates well above chance ($A' = \sim.80$), and there was no evidence of further forgetting with more intervening objects. That is, performance was equivalent for objects fixated between three objects before the test and ten objects before the test. At ten objects before the test, memory capacity easily exceeded 3–4 object estimates of VSTM capacity. This robust prerecency performance therefore indicates a VLTM component to online scene representation. Irwin and Zelinsky (2002; see also Zelinsky & Loschky, 1998) have found similar effects for object position memory. Thus, VSTM appears to support memory for the visual form of the last two objects fixated and attended in a scene, with memory for objects attended earlier supported by VLTM.

---

[2] Memory for object orientation was also tested in a two-alternative forced-choice task and produced the same serial position effects as token change detection.

[3] This estimate is consistent with independent estimates of VSTM capacity for complex objects (Alvarez & Cavanagh, 2004).

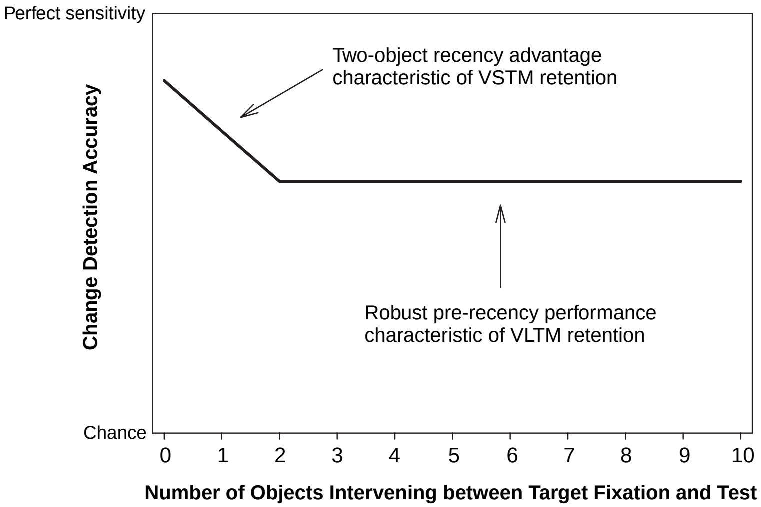

Figure 2. Stylized depiction of the serial position results from Hollingworth (2004), plotting change detection accuracy against the number of objects fixated between target fixation and test. Zero objects indicate that the last object attended in the scene was tested.

Given this significant role for VLTM in the online representation of natural scenes, how robustly are the visual details of individual objects retained in long-term memory (LTM)? In the 1960s and 1970s, studies of picture memory found that LTM could support the retention of multiple thousands of individual photographs (Nickerson, 1965; Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). The distractor pictures used in these experiments, however, were typically chosen to be highly different from studied images, making it difficult to identify the type of memory supporting recognition. Based in part on change blindness effects, recent discussions of this literature have tended to ascribe accurate long-term picture memory to retention of scene gist rather than to retention of the visual details of the photographs (Chun, 2003; Potter, Staub, & O'Connor, 2004; Simons, 1996).

To examine the capacity of LTM for the visual details of individual objects in natural scenes, Hollingworth (2004) used the follow-the-dot method but delayed the change detection test until the end of the session, after all scenes had been viewed. In this condition, more than 400 objects, on average, were examined between target examination and test. Of course, participants did not know which of these objects would be tested until the test occurred. Despite these considerable memory demands, participants performed the token change detection task at a rate well above chance ($A' = .75$), which was only moderately lower than change detection performance when object memory was tested during scene viewing. We have recently delayed object change detection tests (both token and orientation changes) for 24 hours (Hollingworth, 2005b), and change detection performance was still above chance ($A' = \sim.70$). To put this in concrete terms, after having viewed 48 different scenes and hundreds of individual objects, and after a delay of 24 hours, participants can still detect that the toy truck in the bedroom scene (Figure 1) has changed orientation. Clearly, VLTM is not limited to scene gist.

These results are in striking contrast to visual transience claims in the change blindness literature. For example, Simons and Levin (1997) speculated that only the basic meaning of a scene (the gist, such as *bedroom*) may be retained across a saccade from one eye fixation on a scene to the next, a delay of only 20–60 ms. Instead, participants are capable of retaining the visual details of hundreds of individual objects (Hollingworth, 2004) across delays of at least 24 hours (Hollingworth, 2005b).

Understanding change blindness

Evidence of robust visual memory for the visual details of individual objects in scenes naturally leads one to consider why change blindness would ever be observed in the first place. Change blindness is a relative phenomenon. Rarely are subjects entirely insensitive to changes.[4] For example, in the first demonstration of change detection failure within real-world scenes, Grimes (1996) and McConkie (1991) found that some changes were detected by only 25% of participants, whereas others were detected by as much as 80% of participants. Similarly, in a prominent example of an incidental, real-world change paradigm, Simons and Levin (1998) found that approximately 50% of participants failed to detect the replacement of one person for another.

Failures of change detection are typically juxtaposed with an ideal of error-free change detection. But what would be necessary to ensure error-free change detection? Error-free change detection across temporal disruptions such as saccades and brief ISIs requires (at least) two representations and a comparison operation. First, the initial image must be represented in visual memory across the disruption. Second, the perceptual information available in the second image must be represented after the change has been introduced. Third, a comparison process must operate to detect discrepancies between the two representations. To ensure error-free change detection performance, the memory representation would need to be a precise and complete record of the visual information available in the initial image. Similarly, the perceptual representation from the test image also would need

---

[4] See Henderson and Hollingworth (2003b) for one of the few cases in which change blindness is apparently absolute.

to be precise and complete. Finally, the comparison process would need to operate over the entire extent of both representations.

But this model of error-free change detection performance has little a priori plausibility. The human retina encodes high-resolution visual information only over a very small region of the visual field (Riggs, 1965). Consider the case in which a participant is provided a single fixation on a scene before the introduction of a change. Even if the memory representation of the initial image has perfect fidelity, with no visual information loss across the delay, a change could be missed because the changing object lay in the periphery of the visual field and information of sufficient resolution had not been available.

To rescue our model of error-free change detection, we might propose that if participants were allowed multiple fixations on the initial image prior to the change, high-resolution, sensory information from foveated regions might be integrated to form a composite representation retaining precise information across much of the visual field. This type of composite global image model was proposed in the 1970s (e.g., McConkie & Rayner, 1976), primarily to explain the phenomenology of seeing a complete and detailed visual world across eye movements. Such a model has typically been considered to predict error-free change detection (that is, change blindness has been taken as evidence against this model). However, even with the ability to construct a composite image, changes may go undetected if the changing region has not been fixated prior to the change. Since a composite image would require fixation of many local scene regions, its construction would take a significant amount of time (on the order of seconds), and many regions of a scene are not fixated even given extended viewing of 20 s or longer (Henderson & Hollingworth, 1998). And, even if the critical region had been fixated prior to the change, if the critical region is not fixated in the test image, then the change could go undetected despite the retention of precise sensory information in memory. Thus, the physical structure of the eye and the serial nature of foveal object processing makes it highly unlikely that one could ever achieve error-free change detection performance, even with the most generous model of visual memory and comparison processes.

Things get worse for change detection, because we have known for many years that the visual system does not build up precise representations of scenes by integrating high-resolution, foveal information from fixated regions. As reviewed above, precise sensory memory is fleeting (Averbach & Coriell, 1961; Di Lollo, 1980; Sperling, 1960) and simply does not last long enough to support sensory integration across multiple fixations. In the early 1980s, researchers directly tested whether visible persistence is integrated from one fixation to the next, as would be needed to construct a visible, composite sensory image of a scene (Bridgeman & Mayer, 1983; Irwin, Yantis, & Jonides, 1983; McConkie & Zola, 1979; O'Regan & Lévy-

Schoen, 1983; Rayner & Pollatsek, 1983; see also Henderson, 1997; Henderson & Hollingworth, 2003c; Irwin, 1991). For example, Irwin et al. (1983) found that participants could not integrate two complementary patterns of dots when the dots were presented in the same spatial position but on subsequent fixations, demonstrating that the type of sensory integration possible within a fixation at short SOAs (Di Lollo, 1980) does not occur across separate fixations.

Recently, Henderson and Hollingworth (2003c) sought to put the issue of sensory accumulation across eye movements to rest. Participants were shown images of common environments, with each image partially occluded by a set of vertical grey bars (as if viewing the scene from behind a picket fence). During eye movements, the occluded and visible portions of the image were reversed, so that all previously occluded regions of the scene became visible and all previously visible regions occluded. This change drastically altered the low-level content of the entire image (the value of every single pixel changed) but preserved more abstract visual information, such as the general shape of objects and the spatial relationships between objects. Participants were almost entirely insensitive to these changes, demonstrating that visual memory across eye movements is abstracted away from precise sensory information. If sensory representations are not retained and integrated across an eye movement, then sensory information could not be accumulated across multiple fixations to form a composite, global image of a scene.

Actually, one can easily demonstrate that high-resolution, foveal information is not integrated across saccades to form a visible, composite image. Choose two smallish objects (any two objects will do) and place them at two different positions about a foot apart on a nearby surface. Fixate object 1, and without removing fixation, attend to the quality of visual experience for object 2. Object 2 will be projecting to a region of the retina with relatively low resolution, so it should appear fuzzy and indistinct. Now, shift fixation to object 2 and fixate it as long as you care to. Shift fixation back to object 1, and again attend to the quality of perceptual experience for object 2. It should still appear fuzzy and indistinct, precisely as it did before it was directly fixated. The high-resolution foveal information encoded during the fixation on object 2 was not retained to support subsequent detailed perceptual experience of that object: No composite sensory image was formed.

This demonstration illustrates an additional point about visual experience and change detection. Change blindness is often thought to be surprising given the fact that we see a detailed visual world across the visual field. But, in fact, we do not see detail across the visual field, even for previously fixated objects. People see a complete and detailed visual world only in the loose sense that they are not typically aware of the fact that they are not experiencing detail across the visual field (see Dennett, 1991). But it takes only a modicum of effort to realize that peripheral vision is sketchy and indistinct. It would be more accurate to say that we perceive the visual world *as being* complete and detailed. This is a perfectly valid inference (Cohen, 2002), because the world itself is indeed complete and detailed, even if we do not see all of that detail at once.

To summarize, there is simply no plausible model of visual memory and comparison that would produce error-free change detection in change blindness paradigms. Some degree of change blindness is inevitable. But that still leaves open the question of why, given evidence of robust memory for the visual form of objects (Hollingworth, 2003a, 2004, 2005b; Hollingworth & Henderson, 2002), change detection can be as poor as it often is. For example, the appearance and disappearance of the engine on an aeroplane in a flicker paradigm should certainly be detectable by the retention of a higher level visual representation in VSTM or VLTM. Why, then, do participants often fail to notice this sort of change?

Failures of encoding. Change blindness may occur in many circumstances because the local information from the target region has yet to be encoded when a change occurs. Hollingworth and Henderson (2002; see also Hollingworth, Schrock, & Henderson, 2001a) examined change detection performance as a function of whether the target object had been fixated prior to the change. Changes to previously fixated objects were detected at rates well above chance. However, changes to unfixated objects were detected at a rate no higher than the false alarm rate, suggesting that without direct fixation, information sufficient to detect a change was rarely encoded. It may take participants many seconds to fixate each of the potentially changing objects in a scene, explaining delays in detection of repeating changes, such as those in the flicker paradigm. These observations are consistent with the general claim that attention is important for change detection (Rensink et al., 1997), especially when one considers that fixation position and the spatial allocation of visual attention are tightly linked. Although attention may be critical for forming a visual memory representation sufficient to detect most changes, that does not imply that visual representations come unbound after the withdrawal of attention, however.

Failures of retrieval and comparison. Even if one forms a visual memory representation of sufficient accuracy to detect a change, it may be no trivial matter retrieving that representation and comparing it with current perceptual information in order to detect a change. Early change blindness studies assumed that explicit change detection provided an exhaustive measure of visual memory. A number of converging sources of evidence, however, demonstrate that explicit change detection significantly underestimates visual memory and that retrieval and comparison failures are a significant cause of change blindness.

First, changes may go undetected despite accurate memory, because the changed object is not attended or fixated after the change. In Hollingworth and Henderson (2002), single changes to objects in scenes often went undetected until the object happened to be refixated later in viewing. Such a delay, if it had been observed in a flicker paradigm, would have been considered extended change blindness, yet the ultimate detection of the change demonstrated that participants had a memory representation of the relevant object; that representation just was not retrieved and compared to current perceptual information until attention and the eyes were directed back to the target.

Hollingworth (2003a) directly examined the role of retrieval and comparison failure in change blindness using a change detection paradigm in which a scene was viewed for 20 s, followed by a brief mask and a test scene. The target object in the test scene was either the same, rotated, or replaced by a different token. In addition, the target object in the test scene was either postcued by a green arrow or not postcued. The latter method is typical of change blindness experiments. Without a postcue, participants had to decide whether any object in the scene had changed. With a postcue, participants only needed to determine whether the cued object had changed. If change blindness is caused, at least in part, by failed retrieval and comparison, then change detection should be improved when retrieval and comparison demands are minimized by the postcue. This was indeed the case, with significantly higher change detection performance in the postcue condition. In addition, with the benefit of a postcue, change detection performance approached ceiling, both for token and orientation change detection; change blindness was largely eliminated.

Converging evidence that explicit change detection underestimates visual memory comes from three studies conducted by Simons, Levin, and colleagues. In Simons, Chabris, Schnur, and Levin (2002), a naïve participant engaged in conversation with a person carrying a basketball. The basketball was covertly removed during a disruption, and the participant was then asked to report any odd events or changes. If such general questions did not yield report of the removed basketball, the participant was asked specifically about the basketball. With a direct retrieval cue, participants could then often report specific perceptual details of the basketball, even though they did not notice that it had been removed. Similarly, Angelone, Levin, and Simons (2003) found that when participants failed to detect the replacement of one person for another, they could still choose the original person in a forced-choice test at levels above chance. Mitroff, Simons, and Levin (2004), using a computer-based object change detection task, found that on some miss trials, participants had sufficient pre- and postchange information to detect a change but had not adequately compared those representations.

Finally, evidence of preserved memory in the face of change blindness is observed using measures more sensitive than explicit report of change (Fernandez-Duque & Thornton, 2000; see Thornton & Fernandez-Duque, 2002, for a review). When a change is not reported, participants are slower to incorrectly report "same" for an object that changed than to correctly report "same" for an object that did not change (Williams & Simons, 2000). And when a change is not reported, fixation durations on a changed object are longer than on the same object when it has not changed (Hayhoe, Bensinger, & Ballard, 1998; Henderson & Hollingworth, 2003a; Hollingworth et al., 2001b; Ryan, Althoff, Whitlow, & Cohen, 2000). There is currently debate about whether these effects indicate that there are multiple change detection mechanisms (i.e., an implicit mechanism and an explicit mechanism), or whether the data can be accounted for by a single change detection mechanism (Fernandez-Duque, Grossi, Thornton, & Neville, 2003; Fernandez-Duque & Thornton, 2003; Mitroff, Simons, & Franconeri, 2002). Regardless of the resolution of this debate, effects of change on indirect measures (such as RT and fixation duration) in the absence of explicit report of change demonstrate that explicit change detection underestimates visual memory.

Effects of unreported change on indirect measures may be generated by threshold mechanisms for signalling change in the world. Dynamic vision often introduces perceptual discrepancies that could be attributed either to internal error or to external change. For example, when making a saccade to an object, the eyes often fall short of the target of the eye movement. After the completion of such an eye movement, the saccade target object does not lie at the centre of gaze. This circumstance could be due to the inaccuracy of the eye movement (internal error), but it could also be due to the movement of the target object during the saccade (external change). By actually shifting saccade targets during saccades, researchers have revealed that the visual system sets a threshold for attributing position discrepancy to a change in the world. If the displacement of the saccade target is greater than approximately one-third of the distance of the saccade, participants are likely to perceive the target to have moved, attributing the discrepancy to change in the world (e.g., Bridgeman, Hendry, & Stark, 1975; McConkie & Currie, 1996). Below that threshold, the visual system remains sensitive to the displacement (a corrective saccade is executed to bring the saccade target onto the fovea), but participants are rarely aware of the displacement or of the corrective saccade. Thus, for small discrepancies that are likely to have been caused by motor error, the visual system does not attribute the discrepancy to a change in the world, and participants do not perceive the target object to have moved. Despite sensitivity to the discrepancy and appropriate correction, participants are “blind” to the change. Further, if external change is made clearly evident by blanking the target briefly after the eye movement, participants are more likely to attribute the discrepancy to a shift of the target, and explicit awareness of the shift is dramatically improved (Deubel, Schneider, & Bridgeman, 1996).

Threshold mechanisms can also be observed in the phenomenon of insensitivity to incremental change (Hollingworth & Henderson, 2004). In a version of the flicker paradigm, Hollingworth and Henderson gradually rotated an entire scene, with each image incremented by 1 degree of orientation. Participants were remarkably insensitive to these gradual changes, often coming to treat significantly different views of a room (i.e., ones in which many of the original objects had rotated out of view) as unchanged continuations of the initial view. Despite failure to detect the incremental rotation, memory was nevertheless sensitive to the difference between views. With incremental rotation, scene memory came to reflect the recent, changed state of the environment rather than the initial state. This implicit updating of memory to reflect the most recent state of the environment meant that comparison typically operated over similar representations: The currently visible image was compared to memory for the most recent image(s). Thus, the discrepancy between perceptual information and memory tended to be very small, falling below threshold for explicit detection of change, despite the fact that both representations were highly different from the initial image. Even though memory was sensitive to the fact that the image had changed, individual comparisons rarely exceeded threshold for explicit awareness of change, yielding change blindness.

The original explanations for change blindness were highly attractive in their parsimony: Changes were missed because coherent visual representations disintegrate upon the withdrawal of attention (Rensink et al., 1997). It is clear that visual sensory memory is indeed transient (Sperling, 1960), and participants would certainly detect changes more accurately if sensory information was retained and integrated across disruptions such as eye movements, but we have long known that such integration does not occur (e.g., Irwin et al., 1983). Although sensory memory is transient, higher level visual representations are retained robustly in VSTM and in VLTM (Hollingworth, 2003a, 2004, 2005b; Hollingworth & Henderson, 2002): Coherent visual representations do not necessarily disintegrate upon the withdrawal of attention. Recent evidence suggests that despite the ability to accumulate visual representations in VSTM and VLTM, participants fail to detect changes (1) because they have not fixated and attended the changing object prior to the change and thus have not had an opportunity to encode information sufficient to detect a change, (2) because they have not retrieved or adequately compared a memory representation to current perceptual information, and (3) because, for many comparisons, evidence of discrepancy falls below threshold for signalling a change in the world.

VISUAL SEARCH AND THE REPRESENTATION OF NATURAL SCENES

Simultaneously with work on change blindness, Horowitz and Wolfe (1998) reported a phenomenon they termed *memory-free* search, which led to a theoretical account of perception, attention, and memory very similar to Rensink's (2000) coherence theory. Horowitz and Wolfe used a search task in which they either kept the positions of search elements static or scrambled the locations of search elements every 111 ms. If search has a memory component, e.g., one that keeps track of which objects and locations have already been examined (Klein, 1988), then search should have been less efficient when scrambling eliminated the utility of memory. Yet, Horowitz and Wolfe found that search efficiency, as measured by the slope relating reaction time to set size, was no different for scrambled search and static search. They concluded that visual search does not rely on memory. Converging evidence came from a paradigm in which participants repeatedly searched for a different target over a static search array (Wolfe, Klempen, & Dahlen, 2000). Wolfe et al. found that search efficiency did not improve with array repetition, suggesting that participants did not form a memory representation of the array that could influence dynamic visual search.

To account for these findings, Wolfe (1999) proposed that early vision produces loose assemblages of visual features (which Wolfe & Bennett, 1997, termed *preattentive object files* and which appear to be essentially the same concept as Rensink's proto-objects). Attention serves to bind features into a coherent object representation (Wolfe, 1999, p. 77):

> When the eyes first open on a new scene, preattentive processes extract features and assign them, loosely, to preattentive objects. Typically, attention will be deployed to one object. The act of attention allows the features of the object to be organized and processed in a way that permits object recognition. The attended object is perceived differently than the not-yet-attended objects in the scene. Assuming this is to be the case, what happens when attention is to be deployed to the next object? Does the visual representation have a *memory* for the work of attention?

Based on the evidence from search paradigms, Wolfe concluded, as did Rensink (2000), that visual representations dissolve into their elementary features after the withdrawal of attention.

Wolfe's (1999) proposals primarily concern conscious vision. Whatever effects attention has on conscious perception are lost as soon as attention is withdrawn from an object. Put in slightly different terms, Wolfe proposes that the visual system does not accumulate *visible*, coherent object representations as attention is oriented from one object to another. But is there a plausible model that could produce visible accumulation across shifts of attention? The only visual memory store that preserves visible representations is visible persistence, and visible persistence decays within 80–100 ms after the onset of a stimulus, providing no plausible basis for accumulation. Thus, the Wolfe claim appears entirely consistent with early research demonstrating that visible, sensory memory is transient (Averbach & Coriell, 1961; Di Lollo, 1980; Sperling, 1960) and that visible representations do not accumulate during viewing (e.g., Irwin et al., 1983).[5]

Leaving issues of phenomenology aside, the more general claim that memory plays no role in search (Horowitz & Wolfe, 1998) has generated a great deal of research demonstrating that memory does indeed play an important role in visual search paradigms. Memory supports search both within a trial, as the visual system keeps track of which objects have been examined (Gibson, Li, Skow, Brown, & Cooke, 2000; Klein, 1988; Klein & MacInnes, 1999; Kristjánsson, 2000; Müller & von Mühlenen, 2000; Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Takeda & Yagi, 2000; von Mühlenen, Müller, & Müller, 2003), and across trials, as memory for previous searches guides attention to a target object (Chun & Jiang, 1998, 1999). I will not discuss this work in detail, as it has been reviewed comprehensively elsewhere (Shore & Klein, 2000; Woodman & Chun, 2006 this issue). I will, however, briefly review studies conducted in my laboratory designed to investigate the role of memory in real-world search, linking research on visual search to research on scene perception and change detection.

To provide a direct test of the role of memory in search over natural scenes, I developed a search paradigm in which previous exposure to a search scene was controlled (Hollingworth, 2003b). Participants either saw a preview of a scene prior to search through that scene, or they did not see a preview. Figure 3 shows the key events in a trial. Each trial in the preview condition began with a preview display of a real-world scene for 10 s

---

[5] If visible object representations were to accumulate as attention shifts from object to object in a scene, then the following should occur. When first gazing upon a new environment, perceptual experience should be quite impoverished, since few objects would yet have been attended. However, visual experience should get progressively richer as more objects are attended and visible representations are accumulated. The fact that this does not happen—the world looks equivalently rich whether one has been looking at a scene for a few hundred ms or a few minutes—provides further, intuitive evidence that visible information does not accumulate during viewing, whether across shifts of attention or shifts of the eyes.

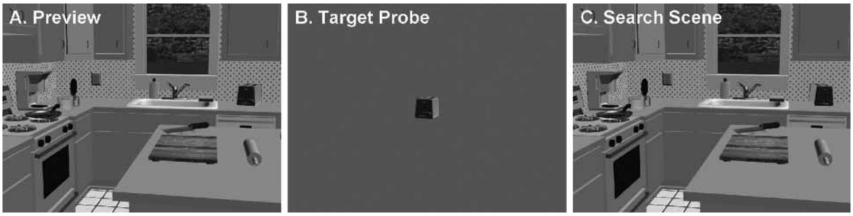

Figure 3. Key stimuli in a preview condition trial of Hollingworth (2003b). In the search scene, the target object has been mirror reversed.

(Figure 3, panel A). The object that would later be the search target was present in the preview, but participants did not know which of the objects in the preview would be the target. The target was then displayed in isolation in the centre of the screen (the *target probe*, Figure 3, panel B) to indicate which object should be found in the search scene. The target probe was identical to the target object that had appeared in the preview scene. Next, a search scene was displayed (Figure 3, panel C). The target object was always present in the search scene in the same location as it had appeared in the preview. However, the target in the search scene was either identical to the target in the preview scene (and thus identical to the target probe), or it was mirror reversed. Participants' task was to find the target in the search scene and respond to indicate whether it was the same as the target probe or mirror reversed. Note that memory for the orientation of the target in the preview could not facilitate search, since the target orientations in the preview and search scenes were uncorrelated. Only memory for spatial properties of the scene (such as target position) could facilitate search. The no-preview condition was identical to the preview condition, except no scene preview was displayed. Search efficiency was assessed by collecting reaction time data and by monitoring eye movements.

If search were more efficient with a scene preview, this would provide a straightforward demonstration that memory for the spatial structure of a scene can dynamically influence search. First, mean correct RT was reliably faster with a preview (1232 ms) than without (1487 ms), a difference of 255 ms. Second, mean elapsed time from the onset of the search scene to the first eye fixation on the target object was shorter in the preview condition (374 ms) than in the no preview condition (586 ms). With a 10 s preview, participants came to fixate the target object only 374 ms after the onset of the search scene. On most trials in the preview condition, there was only one or two fixations intervening between search onset and target fixation. In these cases, either the very first saccade or the second saccade on the search scene brought the eyes to the target object. These data actually underestimate how efficiently the eyes were oriented to the target; for many of the trials with two intervening fixations, the first saccade during search was directed to the target but landed just short, leading to a quick fixation and a corrective saccade. Memory typically guided attention directly to the target.

These results are consistent with work by Chun and Jiang (1998) showing benefits for search when the spatial configuration of elements is repeated. Chun and Jiang used a search task with randomly configured simple stimuli (rotated "T"s and "L"s). Throughout the session, some spatial configurations of search elements were repeated. Specifically, the locations of distractors and the target were held constant, but the identities of the distractors and target were randomly varied. After just a handful of repetitions, search over repeated configurations was faster than search over novel configurations, suggesting that participants had learned that a particular spatial configuration predicted a particular target location and had used this knowledge to guide attention efficiently to that target location. This learning appeared to be implicit, because participants could not recognize repeated configurations at the end of the session.

The scene preview results described above (Hollingworth, 2003b) complement the Chun and Jiang (1998) findings and extend our knowledge of how memory influences search. First, repeated exposures to a search environment are not necessary to produce memory effects; memory can guide search even after a single exposure. Second, repeated search and target localization are not necessary to produce memory effects; search was facilitated by memory even though participants had never searched for the target object before. Finally, memory representations supporting search need not be implicit in nature. After a 10 s preview, participants can explicitly recall the locations of objects in scenes and can estimate target position quite accurately (Hollingworth, 2005a).

What type of spatial memory supported search in the scene preview experiment? There appear to be two main possibilities. First, memory for the spatial configuration of contextual surfaces and objects in the scene could have guided attention to a location where the target object was likely to be found. For example, if one remembered the spatial position of a kitchen counter from the preview image, followed by a toaster target, one could bias search towards the likely location of the toaster on the counter (see Henderson, Weeks, & Hollingworth, 1999; Oliva, Torralba, Castelhano, & Henderson, 2003). Second, participants might remember the specific location of the target object and direct attention to that remembered location. Memory for the configuration of contextual surfaces and objects is broadly consistent with visual transience theories of scene representation (Rensink, 2000; Simons, 1996), because these views hold that abstract spatial layout is encoded and retained robustly without attention. However, memory for the specific locations of objects requires maintaining binding between object representations and scene locations (Irwin & Zelinsky, 2002).

Evidence that search can be facilitated by memory for specific target location would bolster the claim that visual memory supports the retention and accumulation of local object information in scenes (Hollingworth, 2003a, 2004, 2005b; Hollingworth & Henderson, 2002) and would extend those claims to the domain of spatial memory.

To examine the contributions of memory for contextual information and memory for specific target location, Hollingworth (2003b) added a third preview condition to the scene preview paradigm. In this target-absent preview condition, the scene preview was identical to the standard preview condition, except the target object was not present in the preview. If search were more efficient in the standard preview condition than in the target-absent preview condition, this would demonstrate that memory for the specific position of the target object facilitates search, as the two conditions differed only in the presence of the target in the preview scene. If search were more efficient in the target-absent preview condition than in the no preview condition, this would demonstrate that memory for the layout of contextual objects and surfaces facilitates search. Indeed, search was more efficient in the standard preview condition than in the target-absent preview condition, which in turn was more efficient than search in the no-preview condition, both for correct RT and for elapsed time to target fixation. In summary, scene memory can exert a strong influence on dynamic visual search, guiding attention and the eyes to a search target. Scene representations supporting search include memory for the layout of contextual objects and surfaces and memory for the positions of individual objects.

CONCLUSION

When looking upon a complex scene, visual sensory representations are generated across the visual field. If the scene is removed or perceptual processing otherwise interrupted (e.g., across an eye movement), sensory persistence decays very quickly (Averbach & Coriell, 1961; Di Lollo, 1980; Sperling, 1960) and is not integrated from one view of the scene to the next (Henderson & Hollingworth, 2003c; Irwin et al., 1983). However, directing attention to an object allows the formation of a coherent visual representation (Treisman, 1988) and the consolidation of that representation into more stable VSTM (Averbach & Coriell, 1961; Irwin, 1992a; Schmidt, Vogel, Woodman, & Luck, 2002; Sperling, 1960), which maintains visual representations abstracted away from precise sensory information (Irwin, 1991; Phillips, 1974). After attention is withdrawn from an object, abstracted visual representations persist (Hollingworth, 2003a; Hollingworth & Henderson, 2002), and they accumulate in memory as attention and the eyes are oriented from object-to-object within a scene (Hollingworth, 2004), supported both by VSTM (for the last two objects attended) and by VLTM (for objects attended earlier) (Hollingworth, 2004). VLTM then supports the retention of scores of individual object representations over relatively long periods of time (Hollingworth, 2004, 2005b). Scene representations retain information not only about the visual form of individual objects but also about the locations of objects and the configuration of objects and surfaces within a scene (Hollingworth, 2005a). Memory for the spatial properties of a scene can interact dynamically with perceptual processing during visual search, efficiently guiding attention and the eyes to target locations (Chun & Jiang, 1998; Hollingworth, 2003b).

REFERENCES

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.

Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception*, *32*, 947–962.

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

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

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*, 719–722.

Bridgeman, B., & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society*, *21*, 285–286.

Chun, M. M. (2003). Scene perception and memory. *Psychology of Learning and Motivation*, *42*, 79–108.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365.

Cohen, J. (2002). The grand grand illusion illusion. *Journal of Consciousness Studies*, *9*, 141–157.

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

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown.

Deubel, H., Schneider, W. X., & Bridgeman, B. (1996). Post-saccadic target blanking prevents saccadic suppression of image displacement. *Vision Research, 36,* 985–996.

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

Fernandez-Duque, D., Grossi, G., Thornton, I. M., & Neville, H. J. (2003). Representation of change: Separate electrophysiological markers of attention, awareness, and implicit processing. *Journal of Cognitive Neuroscience*, *15*, 1–17.

Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition*, *7*, 323–344.

Fernandez-Duque, D., & Thornton, I. M. (2003). Explicit mechanisms do not account for implicit localization and identification of change: A reply to Mitroff et al. (2002). *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 846–858.

Gibson, B. S., Li, L., Skow, E., Brown, K., & Cooke, L. (2000). Searching for one versus two identical targets: When visual search has memory. *Psychological Science*, *11*, 324–327.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Vancouver studies in cognitive science: Vol. 5. Perception* (pp. 89–110). Oxford, UK: Oxford University Press.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*, 125–137.

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

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics*, *57*, 787–795.

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). *Visual memory and the online representation of complex scenes*. Paper presented at the Munich Visual Search symposium, Munich, Germany.

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

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

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

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 113–136.

Hollingworth, A., & Henderson, J. M. (2004). Sustained change blindness to incremental scene rotation: A dissociation between explicit change detection and visual memory. *Perception and Psychophysics, 66,* 800–807.

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001a). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory and Cognition*, *29*, 296–304.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001b). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin and Review*, *8*, 761–768.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*, 575–577.

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 and 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 and Psychophysics*, *64*, 882–895.

Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*, 430–431.

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

Kristjánsson, A. (2000). In search of remembrance: Evidence for memory in visual search. *Psychological Science*, *11*, 328–332.

Luck, S. J., & Vogel, E. K. (1997). 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.

McConkie, G. W. (1991). *Where vision and cognition meet*. Paper presented at the Human Frontier Science Program workshop on Object and Scene Perception, Leuven, Belgium.

McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 563–581.

McConkie, G. W., & Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer & R. B Ruddell (Eds.), *Theoretical models and processes in reading* (pp. 137–162). Newark, DE: International Reading Association.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, *25*, 221–224.

Mitroff, S. R., Simons, D. J., & Franconeri, S. L. (2002). The Siren Song of implicit change detection. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 798–815.

Mitroff, S. R., & Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness results from failures to compare retained information. *Perception and Psychophysics, 66,* 1268–1281.

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

Müller, H. J., & von Mühlenen, A. (2000). Probing distracter inhibition in visual search: Inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1591–1605.

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, *19*, 155–160.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing*, *1*, 253–256.

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., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, *23*, 765–768.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–1011.

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

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, *12*, 287–292.

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

Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, *29*, 117–133.

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.

Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception and Psychophysics*, *34*, 39–48.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

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

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science*, *8*, 368–373.

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

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.

Shore, D. I., & Klein, R. M. (2000). On the manifestations of memory in visual search. *Spatial Vision*, *14*, 59–75.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, *7*, 301–305.

Simons, D. J., Chabris, C. F., Schnur, T. T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition*, *11*, 78–97.

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

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review*, *5*, 644–649.

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

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*, 73–74.

Takeda, Y., & Yagi, A. (2000). Inhibitory tagging in visual search can be found if search stimuli remain visible. *Perception and Psychophysics*, *62*, 927–934.

Thornton, I. M., & Fernandez-Duque, D. (2002). Converging evidence for the detection of change without awareness. In J. Hyönä, et al. (Eds.), *The brain's eyes: Neurobiological and clinical aspects of oculomotor research* (pp. 99–118). Amsterdam: Elsevier Science.

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

Von Mühlenen, A., Müller, H. J., & Müller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science*, *14*, 309–314.

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

Williams, P., & Simons, D. J. (2000). Detecting changes in novel 3D objects: Effects of change magnitude, spatiotemporal continuity, and stimulus familiarity. *Visual Cognition*, *7*, 297–322.

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology*, *8*, R303–R304.

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

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, *37*, 25–44.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 693–716.

Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, *14*, 808–830.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 121–138.

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.