Journal of Experimental Psychology:
Human Perception and Performance
2002, Vol. 28, No. 1, 113–136

0096-1523/02/$5.00 DOI: 10.1037//0096-1523.28.1.113

Accurate Visual Memory for Previously Attended Objects in Natural Scenes

Andrew Hollingworth and John M. Henderson
Michigan State University

The nature of the information retained from previously fixated (and hence attended) objects in natural scenes was investigated. In a saccade-contingent change paradigm, participants successfully detected type and token changes (Experiment 1) or token and rotation changes (Experiment 2) to a target object when the object had been previously attended but was no longer within the focus of attention when the change occurred. In addition, participants demonstrated accurate type-, token-, and orientation-discrimination performance on subsequent long-term memory tests (Experiments 1 and 2) and during online perceptual processing of a scene (Experiment 3). These data suggest that relatively detailed visual information is retained in memory from previously attended objects in natural scenes. A model of scene perception and long-term memory is proposed.

Because of the size and complexity of the visual environments humans tend to inhabit, and because high-acuity vision is limited to a relatively small area of the visual field, detailed perceptual processing of a natural scene depends on the selection of local scene regions by movements of the eyes (for reviews, see Henderson & Hollingworth, 1998, 1999a). During scene viewing, the eyes are reoriented approximately three times each second by *saccadic eye movements* to bring the projection of a local scene region (typically a discrete object) onto the area of the retina producing the highest acuity vision (the fovea). The periods between saccades, when the eyes are relatively stationary and detailed visual information is encoded, are termed *fixations* and last an average of approximately 300 ms during scene viewing. During each brief saccadic eye movement, however, visual encoding is suppressed (Matin, 1974). Thus, the visual system is provided with what amounts to a series of snapshots (corresponding to fixations), which may vary dramatically in their visual content over a complex scene, punctuated by brief periods of blindness (corresponding to saccades).

Andrew Hollingworth and John M. Henderson, Department of Psychology and Cognitive Science Program, Michigan State University.

This research was supported by a National Science Foundation (NSF) graduate fellowship to Andrew Hollingworth and NSF Grants SBR 9617274 and ECS 9873531 to John M. Henderson. This research derives from a doctoral dissertation submitted by Andrew Hollingworth to Michigan State University. Aspects of the data were reported at the seventh annual Workshop on Object Perception and Memory, Los Angeles, CA, November 1999, and the 41st annual meeting of the Psychonomic Society, New Orleans, LA, November 2000.

We thank the dissertation committee of Tom Carr, Rose Zacks, and Richard Hall for their helpful discussions of the research and for their comments on the dissertation. We also thank Dan Simons and Dave Irwin for comments on an earlier version of the article, and Gary Schrock for his technical assistance.

Correspondence concerning this article should be sent to Andrew Hollingworth, who is now at Yale University, Department of Psychology, Box 208205, New Haven, Connecticut 06520-8205. E-mail: andrew.hollingworth@yale.edu

The selective nature of scene perception places strong constraints on the construction of an internal representation of a scene. If a detailed visual representation is to be formed, then information from separate fixations must be retained and combined over one or more saccadic eye movements as the eyes are oriented to multiple local regions. The temporal and spatial separation of eye fixations on a scene leads to two general memory problems in the construction of a scene representation. One is the short-term retention of scene information across a single saccadic eye movement, particularly from the target of the next saccade (for reviews, see Henderson & Hollingworth, 1999a; Irwin, 1992b; Pollatsek & Rayner, 1992). The second, which this study investigated, is the accumulation of scene information across longer periods of time and across multiple fixations. That is, what information is retained from previously fixated (and hence attended) regions of a scene, and how is that information used to construct a larger-scale representation of the scene as a whole, if such a representation is constructed at all?

Scene Representation as the Construction of a Composite Image

One possibility is that a sensory representation is retained and combined from previously fixated and attended regions to form a global composite image of the scene. We define *sensory representation* as a precategorical, maskable, complete, and metrically organized (i.e., iconic) representation of the properties available from early vision (such as shape, shading, texture, and color; Irwin, 1992b, Sperling, 1960). According to this composite image hypothesis, sensory representations from individual fixations are integrated within a visual buffer and organized according to the position in the world from which each was encoded (Breitmeyer, Kropfl, & Julesz, 1982; Davidson, Fox, & Dick, 1973; Feldman, 1985; McConkie & Rayner, 1976). Metaphorically, local high-resolution information is painted onto an internal canvas, producing over multiple fixations a metrically organized composite image of previously attended regions. Such a composite sensory image could then be used to support a variety of visual–cognitive tasks

and would account for the phenomenological perception of a highly detailed and stable visual world.

This possibility has proved attractive, but a large body of research demonstrates conclusively that the visual system does not integrate sensory information across saccadic eye movements (Bridgeman & Mayer, 1983; Henderson, 1997; Irwin, 1991; Irwin, Yantis, & Jonides, 1983; McConkie & Zola, 1979; O'Regan & Lévy-Schoen, 1983; Rayner & Pollatsek, 1983). For example, Irwin et al. (1983) and Rayner and Pollatsek (1983) found that participants could not integrate two dot patterns when presented in the same spatial position but on subsequent fixations, suggesting that the type of sensory fusion possible within a fixation across short interstimulus intervals (e.g., Di Lollo, 1980) does not occur across separate fixations. In addition, a precise representation of object contours does not appear to be retained across eye movements (Henderson, 1997; Pollatsek, Rayner, & Collins, 1984). If sensory representations are not retained across a single saccadic eye movement, sensory information could not be accumulated across multiple fixations to form a composite global image of a scene.

Although transsaccadic memory does not support sensory integration, visual representations are nonetheless retained across eye movements. In transsaccadic object identification studies, participants are faster to identify an object when a preview of that object has been available prior to the saccade (e.g., Henderson, Pollatsek, & Rayner, 1989), and this benefit is influenced by visual changes, such as substitution of one object with another from the same basic-level category (Henderson & Siefert, in press) and mirror reflection (Henderson & Siefert, 1999, in press). In addition, object priming across saccades appears to be governed primarily by visual similarity rather than by conceptual similarity (Pollatsek et al., 1984). Finally, structural descriptions of simple visual stimuli can be retained and integrated across eye movements (Carlson-Radvansky, 1999; Carlson-Radvansky & Irwin, 1995). Thus, integration across saccades appears to be limited to visual codes abstracted from precise sensory representation, but detailed enough to specify individual object tokens and the viewpoint at which the object was observed. Higher-level visual representations meeting these criteria have been proposed in the object recognition literature, including viewpoint-dependent structural descriptions (Bülthoff, Edelman, & Tarr, 1995) and abstract 2-D-feature representations (Riesenhuber & Poggio, 1999). It is then a possibility that such higher-level visual representations are retained from previously fixated and attended regions and accumulate within a representation of the scene.[1]

Research using natural scene stimuli has provided converging evidence that the visual system does not form a representation as detailed and complete as a composite sensory image. A number of studies have made changes to a scene during a saccadic eye movement with the logic that if a global image of the scene were constructed and retained across eye movements, changes to the scene should be detected easily. However, participants have proved rather poor at detecting scene changes across saccadic eye movements (Currie, McConkie, Carlson-Radvansky, & Irwin, 2000; Grimes, 1996; Henderson & Hollingworth, 1999b; McConkie & Currie, 1996). For example, Grimes and McConkie (Grimes, 1996; McConkie, 1991) coordinated relatively large scene changes (e.g., enlarging a child in a playground scene by 30% and moving the child forward in depth) with saccadic eye movements and found that participants detected changes at well below 50% correct. A stronger manipulation was conducted by Henderson, Hollingworth, and Subramanian (1999), in which every pixel in a scene image was changed during a saccade (a set of gray bars occluded half of the scene image; during a saccade, the occluded and visible portions were reversed). Although the pictorial content changed dramatically over the entire scene, participants detected these changes less than 3% of the time.

In addition, this phenomenon of poor change detection, or *change blindness*, is not limited to scene changes made during saccadic eye movements. Rensink, O'Regan, and Clark (1997) examined whether the apparent inability to accumulate sensory information across discrete views of a scene is a specific property of saccadic eye movements or a more general property of visual perception and memory. Rensink et al. simulated the visual events caused by moving the eyes: Initial and changed scene images were displayed in alternation for 250 ms each (roughly the duration of a fixation on a scene), and each image was separated by a brief 80-ms blank interval (corresponding to saccadic suppression). As in transsaccadic change studies, participants were often quite poor at detecting significant changes to a scene in this flicker paradigm. Subsequent research has demonstrated similar change blindness when a change occurs across many different forms of visual disruption, including film cuts (Levin & Simons, 1997), occlusion in a real-world encounter (Simons & Levin, 1998), or a blink (O'Regan, Deubel, Clark, & Rensink, 2000).

In summary, the literature on visual memory across saccades and other visual disruptions conclusively demonstrates that sensory representations are not integrated to form a composite image of a scene. If visual representations are accumulated from previously fixated and attended regions, they must be in a form significantly more abstract than sensory representation.

Localist Attention-Based Accounts

Recent proposals have abandoned the idea of a composite sensory image in favor of a view that visual scene representation is more local and more transient. Irwin (1992a, 1992b; Irwin & Andrews, 1996) has proposed an *object file theory of transsaccadic memory*, developed primarily to explain the integration of

[1] As is evident from this discussion, we use the term *visual* to refer to both lower-level sensory representations and higher-level visual representations, such as a structural description. This usage is consistent with most of the existing literature in visual cognition, object perception, and transsaccadic memory and integration. In addition, we distinguish visual representations (encoding properties such as shape and color) from conceptual representations (encoding object identity and other associative information). However, within the change blindness literature, some researchers prefer to limit the term *visual* to sensory representation (e.g., Simons, 1996). Others appear to equate visual representation with conscious visual experience (e.g., Wolfe, 1999). Yet given that higher-level visual representations form the basis of integration across saccades and are functional in visual processes such as object recognition, it does not seem appropriate to limit the term *visual* to sensory representation. In addition, whatever constitutes visual experience across saccades must necessarily be due to higher-level visual representation, as sensory information is not retained from one fixation to the next. Finally, given that much of the work of vision is unavailable to awareness, we believe it is unnecessarily constraining to equate visual representation with conscious visual experience.

information across single eye movements but with implications for scene representation, in general, and for the accumulation of scene information from previously fixated and attended regions. The allocation of visual attention, according to Irwin, rules what local visual information is and is not represented from a complex scene. When attention is directed to an object, visual features are bound into a unified object description (Treisman, 1988). In addition, a temporary representation is formed, an *object file*, that links the visual object description to a spatial position in a master map of locations (Kahneman & Treisman, 1984). Across a saccade, object files are maintained in visual short-term memory (VSTM), a relatively long-lasting, capacity-limited store maintaining visual codes abstracted from sensory representation (Irwin, 1992b). Objects files, then, are the primary content of memory across saccades, providing local continuity from one fixation to the next.

In this view, only a very small portion of the local information available in a complex natural scene is represented across a saccade, due to the strong capacity constraints on VSTM. Irwin (1992a) has provided evidence that three or four discrete object files can be retained in VSTM across a saccade. In those experiments, an array of letters was presented prior to an eye movement. The letters were removed during the saccade, and a position was cued. Participants' ability to report the identity of the letter in the cued position was consistent with the retention of three or four position-bound letter codes. As a consequence of this limited capacity in VSTM, only currently or recently attended objects are represented in any detail. Visual representations from previously fixated and attended regions should be quickly replaced as new object files are constructed. In support of this view, Irwin and Andrews (1996) used the partial report paradigm described above but allowed participants two fixations rather than one in the array prior to test. If visual information encoded during the second fixation accumulates with that encoded during the first fixation, then report performance should have been reliably higher with two fixations prior to the probe versus one. Yet report performance after two fixations was not reliably improved, suggesting little if any visual accumulation across the two fixations.

In addition, Irwin (1992b; Irwin & Andrews, 1996) has integrated the object file framework into a more general theory of scene representation and transsaccadic memory. In this view, the scene information retained across a saccadic eye movement is limited to three sources. One is active object files maintaining visual codes (abstracted from sensory representation) from currently attended or recently attended objects and, in particular, from the target of the next saccade (Currie et al., 2000). The second is position-independent activation of conceptual nodes in long-term memory (LTM) coding the identity of local objects that have been recognized (Henderson, 1994; Henderson & Anes, 1994; Pollatsek, Rayner, & Henderson, 1990). The third is schematic scene-level representations derived from scene identification, coding conceptual–semantic properties such as scene meaning or gist. However, only object files maintain a visual representation of local objects, and these structures are transient. Irwin and Andrews (1996) summarize this view as follows:

> According to object file theory, relatively little information actually *accumulates* across saccades; rather, one's mental representation of a scene consists of mental schemata and identity codes activated in long term memory and of a small number of detailed object files in short-term memory. (p. 130)

More recent proposals, drawn primarily from the change blindness literature, have placed even greater emphasis on the role of attention in scene perception and on the transience of visual representation (O'Regan, 1992; O'Regan, Rensink, & Clark, 1999; Rensink, 2000a, 2000b; Rensink et al., 1997; Simons & Levin, 1997; Wolfe, 1999). Rensink (2000a, 2000b) has provided the most detailed account of this view, termed *coherence theory*. As in Irwin's object file theory of transsaccadic memory, coherence theory claims that visual attention is necessary to bind sensory features into a coherent object representation and to maintain this representation in VSTM, which is stable across brief disruptions such as saccadic eye movements. In contrast, unattended sensory representations decay rapidly and are overwritten by new visual encoding. When visual attention is withdrawn from an object, however, the representation of that object immediately reverts to its preattentive state, becoming "unglued" (see also Wolfe, 1999).[2] Finally, initial perceptual processing of a scene activates schematic representations of scene gist and general spatial layout that are preserved across visual interruptions, providing an impression of scene continuity. According to Rensink (2000a), scene gist corresponds to a scene category label (e.g., *bedroom*), and the representation of spatial layout does not contain information about the visual properties of individual objects. Thus, visual representation is limited to the currently attended object. Because there are few, if any, representational consequences of having previously attended an object, the visual system is unable to accumulate information from previously attended regions.

Though clearly similar, coherence theory and the object file theory of transsaccadic memory appear to differ on three points. First, Rensink (2000a) proposes that only one object can be maintained in VSTM across visual disruptions, whereas Irwin (1992a) provides evidence that three or four objects can be maintained. Second, Rensink proposes that sensory representations can be retained in VSTM across disruptions such as saccades, whereas object file theory holds that VSTM supports the maintenance of visual representations abstracted from sensory information. Third, coherence theory holds that visual object representations disintegrate as soon as attention is withdrawn, whereas object files can remain active after attention is withdrawn (at least until replaced). The first two differences are unlikely to be critical. Although Rensink claims that VSTM is limited to one object, he leaves open the possibility that the visual system may treat a collection of three or four objects as a single entity. The second difference is signif-

[2] Although a key proposal in Wolfe (1999) is that a unified object representation dissolves when attention is withdrawn from an object, more recent work by Wolfe, Klempen, and Dahlen (2000) has modified this earlier proposal. The modified claim in Wolfe et al. (2000) is that when attention is withdrawn from an object, the link established between the visual representation of that object and corresponding LTM representations (which allows conscious identification) is dissolved. As a result, multiple objects in a scene cannot be consciously and simultaneously recognized. However, Wolfe et al. (2000) have left open the possibility that bound object representations may be retained in memory after attention is withdrawn from an object and used for subsequent change detection. Thus, Wolfe's view no longer appears consistent with the original claim that visual object representations disintegrate on the withdrawal of attention.

icant, but extant data provide conclusive evidence that sensory representations cannot be retained across disruptions such as saccades, as reviewed above. The only difference of real import, then, concerns the fate of previously attended objects: Do visual representations disintegrate immediately on the withdrawal of attention, or do they remain active until replaced by subsequent encoding? This difference in theory leads to different predictions regarding the detection of changes to natural scenes. Coherence theory predicts that only visual changes to a currently attended object should be detected, whereas object file theory predicts that changes to an unattended object could be detected if the object has been attended earlier and if its object file has not been replaced by subsequent encoding.

In summary, both theories propose that the visual representation of a scene across disruptions such as saccades is local and transient, with only currently or recently attended objects represented in any detail. Thus, we refer to these proposals as *visual transience hypotheses* of scene representation. Although the representation of visual information is proposed to be transient, these theories do allow for the retention of more abstract and stable representations, coding such properties as scene gist, the spatial layout of the scene, and the identities of recognized objects. With regard to visual representation, visual transience hypotheses are consistent with a view of perception in which the visual system does not rely heavily on memory to construct a scene representation, but instead depends on the fact that local objects in the environment can be sampled when necessary by movements of the eyes or attention. The world itself serves as an "external memory" (O'Regan, 1992; O'Regan et al., 1999). In addition, visual transience hypotheses are consistent with functionalist approaches to scene representation (Ballard, Hayhoe, Pook, & Rao, 1997; Hayhoe, 2000; Hayhoe et al., 1998), which reject the notion that the visual system creates a general purpose representation that can support a variety of tasks. Instead, the representation of local scene information is directly governed by the allocation of attention to goal-relevant objects.

Researchers initially assumed that the goal of vision was to construct a global and veridical internal representation of the visual world by integrating sensory representations from multiple local fixations. The pendulum of theory has now swung to the view that little or no visual information is retained from previously fixated and attended regions of a scene, that visual representation is transient, leaving no lasting memory.

In the next sections, we discuss two lines of evidence relevant to the visual transience claim: (a) research on LTM for scenes and (b) change detection studies that have examined visual representation after the withdrawal of attention.

Evidence From Long-Term Scene Memory

One place to look for initial evidence regarding the retention of visual information from natural scenes is the literature on LTM for pictures. Visual transience hypotheses hold that the LTM representation of a scene cannot contain specific visual information, because such information is not retained for very long after attention is withdrawn from an object. Instead, scene memory under this view is limited to gist, layout, and perhaps the abstract identities of recognized objects (Rensink, 2000b; Simons, 1996; Simons & Levin, 1997; Wolfe, 1998). The picture memory literature, however, indicates that long-term picture memory can preserve quite detailed information. Initial studies of picture memory demonstrated that humans possess a prodigious ability to remember pictures presented at study (Nickerson, 1965; Shepard, 1967; Standing, Conezio, & Haber, 1970). For example, Standing et al. (1970) tested LTM for 2,560 images, about 600 of which were classified as *city scenes*. Memory for a subset of 280 images was tested in a two-alternative, forced-choice test, with mean discrimination performance of approximately 90% correct. Thus, memory for a very large number of scenes can be quite accurate, even when some of the studied images have similar subject matter.

Although studies of memory capacity demonstrate that scene memory is specific enough to successfully discriminate between thousands of different items, they do not identify the nature of the stored information supporting this performance. However, three studies suggest that scene memory can indeed preserve specific visual information. First, Friedman (1979) presented line drawings of six common environments for 30 s each during a study session. At test, changed versions of each scene were presented, and the participant's task was to determine whether the scene was the same as the studied version. One change conducted by Friedman was to replace an object in the initial scene with another object from the same basic-level category (a token change), a manipulation that tested whether specific visual information (as opposed to purely conceptual information) was preserved in memory. Participants correctly rejected 25% of the changed scenes when the target object was very likely to appear in the scene, 39% when the target object was moderately likely to have appeared in the scene, and 60% when the target object was unlikely to have appeared in the scene. In addition, Parker (1978) found accurate correct rejection performance (above 85% correct) on a recognition memory test for token and size changes to individual objects in a scene.

Together these data suggest that visual information can be retained in memory from individual objects in a scene. One potential problem with these studies, however, is that each of a relatively small number of scenes was repeated a large number of times. In addition to the initial 30-s study period, Friedman (1979) presented each scene 12 different times in the memory test session to test different object changes. A second potential problem with these studies is that they used relatively simple stimuli. Parker's scenes contained just six discrete objects arranged on a blank background. Thus, the small number of scenes, the visual simplicity of those scenes, and the repeated presentation of each scene may have produced unrealistic estimates of the extent to which specific visual information was retained.

Converging evidence that LTM preserves specific visual information comes from the Standing et al. (1970) study. Memory for the left–right orientation of studied pictures was tested by presenting studied scenes at test, either in the same orientation as at study or in the reverse orientation. The orientation of a picture was unlikely to be encoded using a purely conceptual representation, as the meaning of the scenes did not change when the orientation was reversed. However, participants were able to correctly identify the initially viewed picture orientation 86% of the time after a 30-min retention interval. Thus, the Standing et al. study demonstrated that in addition to being accurate enough to discriminate between thousands of studied pictures, LTM for scenes is not limited to the gist of the scene or to the identities of individual objects. However, it is at least possible that left–right orientation discrimination could have been driven by an accurate representation of the layout of the

scene without the retention of visual information from local objects (Simons, 1996).

In summary, the picture memory literature converges on the conclusion that scene representation is more detailed than would be expected under visual transience hypotheses. However, no single study provides unequivocal evidence that visual representations are reliably retained in memory from previously attended objects in scenes.

Evidence From Change Detection Studies

Further evidence bearing on whether visual information is accumulated from previously fixated and attended objects comes from studies examining change detection as a function of eye position (reviewed in Henderson & Hollingworth, in press). Hollingworth, Williams, and Henderson (2001) made a token change to a target object in a line drawing of a scene during the saccade that took the eyes away from that object after it had been fixated the first time. Prior to a saccade, visual attention is automatically and exclusively directed to the target of that saccade (Deubel & Schneider, 1996; Henderson et al., 1989; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995; Rayner, McConkie, & Ehrlich, 1978; Shepherd, Findlay, & Hockey, 1986). For example, Hoffman and Subramaniam (1995) provided evidence that visual attention is allocated exclusively to the saccade target prior to the execution of a saccade, as participants could not attend to one object in the visual field when preparing a saccade to another, even when such a strategy would have been optimal. Thus, in Hollingworth et al. (2001), when a change was introduced on the saccade away from the target object, attention had been withdrawn from the target object before the change occurred. According to coherence theory, this type of change should not be detected, because the maintenance of a visual object representation depends on the continuous allocation of attention to that object (Rensink, 2000a; Rensink et al., 1997). However, participants were able to detect these changes, albeit at a fairly modest rate of 27% correct (the false alarm rate was 2%). This result has also been observed using 3-D-rendered color images of scenes and with a different type of visual change: a 90° rotation in depth (Henderson & Hollingworth, 1999b).

Coherence theory has difficulty accounting for these data, but they are not necessarily inconsistent with the object file theory of transsaccadic memory, because the latter view holds that visual object representations can be maintained briefly after attention is withdrawn. However, two pieces of evidence from the Hollingworth et al. (2001) and Henderson and Hollingworth (1999b) studies appear to be inconsistent with object file theory as well. First, in each of these experiments, detection was often delayed until refixation of the changed object, suggesting that information specific to the visual form and orientation of an object was retained in memory across multiple intervening fixations and consulted when focal attention was directed back to the changed object. Second, in Hollingworth et al. (2001), when a change was not explicitly detected, fixation duration on the changed object was significantly longer than when no change occurred, and this effect was likewise delayed over multiple intervening fixations (13.5 on average). Under object file theory, this longer-term retention of visual information should not occur, because the critical object file should have been replaced by the creation of new object files as the eyes and attention were directed to other objects in the scene. Instead, these data are consistent with the picture memory literature suggesting that detailed visual information (though clearly less detailed than a sensory image) is retained from previously attended objects.

Change Blindness Reconsidered

If relatively detailed visual information can be retained in memory from previously attended objects in a scene, as suggested by the picture memory literature and our initial change detection studies (Henderson & Hollingworth, 1999b; Hollingworth et al., 2001, in press), why would change blindness occur at all? We considered three possibilities. First, in studies demonstrating poor change detection performance, the critical change in the scene could have occurred before the target region was fixated and thus before detailed information was encoded from that region. This hypothesis is motivated by evidence that the encoding of scene information is strongly influenced by fixation position. For example, in an LTM study, Nelson and Loftus (1980) demonstrated that the encoding of object information into a scene representation is generally limited to a very local region around the current fixation position. In addition, in an online change detection paradigm, Hollingworth, Schrock, and Henderson (2001) found that fixation position played a significant role in the detection of scene changes made periodically across a blank interval, with the majority of changes detected only when the object was in foveal or near-foveal vision (see also O'Regan et al., 2000). To acquire further evidence, we reexamined data from Henderson and Hollingworth (1999b), from a control condition in which a target object was changed (deletion or 90° in-depth rotation) during a saccade to a different object in the scene. Trials were divided into those on which the target object had been fixated prior to the change and those on which the change occurred before fixation on the target object. Changes that occurred after fixation on the target were detected more accurately (39.7% correct) than changes that occurred before fixation on the target (14.2% correct), $F(1, 16) = 11.44$, $MSE = 965.5$, $p < .005$. Thus, given the likely dependence of change detection on prior target fixation, changes could sometimes go undetected in change blindness studies simply because the target region was not fixated prior to the change.

A second reason change blindness could underestimate the detail of scene representation is that in studies demonstrating poor change detection performance, information encoded from the target region might not have been retrieved to support change detection. As reviewed above, a number of studies have found that a change to an object is sometimes detected only when the changed region is refixated after the change (Henderson & Hollingworth, 1999b; Hollingworth et al., 2001; Parker, 1978). Thus, fixation and/or focal attention may sometimes be necessary to retrieve stored information about a previously fixated object. If the changed region is not refixated, then the change may go undetected, even though the stored representation of that object is sufficiently detailed to support change detection.

Finally, the standard interpretation of change detection performance in change blindness studies may be incorrect. Within the change blindness literature, the interpretation of change detection measures has tended to use the following logic. Explicit change detection directly reflects the extent to which scene information is

represented. Therefore, if a change is not detected, the information necessary to detect the change must be absent from the internal representation of the scene. However, a number of recent studies have demonstrated that for trials on which a change was not explicitly detected, effects of that change can be observed on more sensitive measures (Fernandez-Duque & Thomton, 2000; Hayhoe et al., 1998; Hollingworth et al., 2001; Williams & Simons, 2000). For example, Hollingworth et al. (2001) found that gaze duration on a changed object when the change was not detected was 250 ms longer on average than when the same object was not changed. Thus, change blindness may be observed not because the critical information is absent from the scene representation but because explicit detection is not always sensitive to the presence of that information.

This Study

The goal of this study, then, was to investigate the nature of the information retained in memory from previously attended objects in natural scenes. The study sought to resolve the apparent discrepancy between evidence of poor change detection (and visual transience hypotheses which seek to explain such change blindness) and evidence of excellent memory for pictures. This primary goal was broken down into a number of component questions. First, how specific is the representation of objects in a scene that have been previously attended but are not within the current focus of attention, both during the online perceptual processing of the scene and later, after the scene has been removed? Second, is fixation of an object important for encoding that object into a scene representation and thus for the detection of changes? Third, does refixation play a role in the retrieval of stored object information, supporting change detection? Fourth, to what extent does explicit change detection reflect the detail of the underlying representation?

Experiment 1

In Experiment 1, we combined a saccade-contingent change paradigm with an LTM paradigm to investigate the nature of the scene representation constructed during scene viewing and stored into LTM.

In an initial study session, computer-rendered color images of common environments were presented to participants, whose eye movements were monitored as they viewed each image for 20 s in preparation for a later memory test. In each scene, one target object was chosen. To investigate the representation of previously attended objects during scene viewing, the target object was changed during a saccade to a different region of the scene, but only if the target object had already been fixated at least once. Because visual attention and fixation position are tightly linked during normal viewing, making the change only after the object had been fixated ensured that that object had been attended at least once prior to the change. However, because visual attention is automatically and exclusively allocated to the target of the next saccadic eye movement prior to the execution of that eye movement, as reviewed above, the target object was not within the current focus of attention when it changed: Before the initiation of the eye movement that triggered the change, visual attention had shifted to the object within the change-triggering region, and thus, participants could not have been attending the target object when the change occurred. Note that this method depends on the (uncontroversial) assumption that fixated objects have also been focally attended; however, this assumption does not entail that all attended objects are necessarily fixated.

To test the specificity of the representation of previously attended objects, the target object in each scene was changed in one of two ways: a type change, in which the target was replaced by another object from a different basic-level category, and a token change, in which the target was replaced by an object from the same basic-level category (Hollingworth et al., 2001; see also Archambault, O'Donnell, & Schyns, 1999). These conditions are illustrated in Figure 1. In the type-change condition, detection could be based on a basic-level coding of object identity. However, if participants can detect token changes, information specific to the object's visual form, as opposed to its basic-level identity, was likely to have been retained.

If detailed visual information is retained from previously attended scene regions, as suggested by the picture memory literature, participants should be able to detect both type changes and token changes. Coherence theory, however, makes a different prediction. Coherence theory holds that only changes to attended visual information, the gist, or the layout of a scene can be detected, as these are the only forms of information retained across disruptions such as saccades. The target object changes in this experiment do not alter attended visual information, as the target object was not attended when the change occurred. In addition, general layout should not be altered by these changes, as the original and changed target objects occupied the same spatial position and were matched for size. It is possible that a type change might alter the gist of the scene if that representation is detailed enough to code the identities of individual objects. The most common definition of gist is a short description capturing the identity of the scene, such as *child's bedroom* (a definition shared by Rensink, 2000a). So, although we conservatively assumed that a type change could alter the gist of the scene, in reality, most type changes would not alter this very abstract description. A token change, however, should never alter the gist of the scene, as the change does not even alter the basic-level identity of the target object itself. Thus, coherence theory makes the clear prediction that token changes should not be detected in this study. In fact, Rensink (2000a) has stated directly that information specific to object tokens can be maintained only in the presence of attention.

The object file theory of transsaccadic memory also predicts poor detection performance. If the object file coding detailed visual information from an object is replaced quickly after attention is withdrawn from that object, detection performance in the token-change condition should decrease as a function of the elapsed time between the withdrawal of attention from the target and the change. A more precise prediction depends on making a number of assumptions about the creation of object files and their replacement in VSTM. According to object file theory, an object file is formed when attention is directed to a new perceptual object. Attention precedes the eyes to the next saccade target, and thus, object file creation could be expected to be roughly one-per-saccade during scene viewing. This is an admittedly rough estimate, because attention could be allocated to more than one object within a single fixation or to the same object across more than one fixation. In addition, the length of time an object file persists after

Figure 1. Sample scene illustrating the change conditions in Experiments 1 and 2. A: Initial scene, in which the notepad is the target object; B: Type change (Experiment 1); C: Token change (Experiments 1 and 2); D: Rotation (Experiment 2). In the experiments, stimuli were presented in color.

the withdrawal of attention depends on the mode of replacement in VSTM. If replacement is first in, first out, then detection performance should decline sharply to zero if the change happens more than about three or four fixations after eyes leave the target region. It is possible, though, that replacement is a stochastic process, in which case a much more gradual, exponential decline in detection performance should be observed.

In either case, however, Irwin's object file theory predicts a significant decline in detection performance as a function of the number of intervening fixations between the last exit of the eyes from the target region prior to the change and the change itself. In keeping with Irwin and Andrews' (1996) claim that there is little accumulation of detailed information across eye movements and that replacement in VSTM is likely to be first in, first out, this view predicts that detection performance should decline to zero quite quickly, within a maximum of about four fixations. Type changes, on the other hand, might be detected successfully and in a manner independent of the number of intervening fixations if the change is significant enough to alter the gist of the scene or if an abstract identity code is retained from the target object, as Irwin's theory holds that these types of information can be maintained in a stable form across multiple eye movements.

The change-after-fixation condition was contrasted with two control conditions. In the change-before-fixation condition, the target object was changed before the first fixation on that object. In the no-change control condition, the initial scene was not changed. The change-before-fixation condition was included to test the extent to which local object encoding is dependent on fixation. If encoding is facilitated by object fixation, then change detection should be reliably poorer when the object had not been fixated prior to the change, compared with when it had been fixated. The no-change control condition was included to assess the false-alarm rate.

Finally, to investigate LTM for the target objects in the scenes, we administered a forced-choice recognition test for no-change control scenes after the study session. Participants saw two versions of each scene in succession, one containing the studied object and the other a distractor object in the same spatial position. The

distractor could be either a different type (type-discrimination condition) or a different token (token-discrimination condition). Similar predictions hold for the LTM test as for online change detection. If visual object representations are retained in memory after attention is withdrawn, as the picture memory literature implies, participants should be able to successfully discriminate between both type and token alternatives. However, if visual representation is transient and there is little accumulation of information from local scene regions, as proposed by both visual transience hypotheses, participants should not be able to accurately discriminate two token alternatives.

In addition, the LTM test in this study avoids some of the interpretative difficulties present in other scene memory paradigms. First, distractors in prior studies were often chosen to maximize discriminability, whereas studied scenes and distractors in this study differed only in the properties of a single object. Second, whereas prior studies showing the retention of token-specific information repeated each scene many times, participants viewed each scene in this study only once prior to the test. Third, prior studies often used a variety of materials from a variety of sources (e.g., mixing together color images with black and white images), whereas the similarity between studied scenes was fairly high in this study: Each scene was a 3-D-rendered color image of a common environment, many of the scenes were taken from the same large-scale model of a single house, and some scenes were created by rendering different viewpoints within a single room model. Thus, this study provides a particularly stringent test of scene memory.

Method

Participants. Twelve Michigan State University undergraduates participated in the experiment for course credit. All participants had normal vision and were naive with respect to the hypotheses under investigation.

Stimuli. Thirty-six scene images were computer-rendered from 3-D wire-frame models using 3-D graphics software. Wire-frame models were acquired commercially, donated by 3-D graphic artists, or developed in-house. Each model depicted a typical human-scaled environment (e.g., *office* or *patio*). To create each initial scene image, a target object was chosen within the model, and the scene was rendered so that this target object did not coincide with the initial experimenter-determined fixation position. To create the type-change scene images, the scene was re-rendered after the target object had been replaced by another object from a different basic-level category. To create the token-change condition, the scene was re-rendered after the target object had been replaced by another object from the same basic-level category. In the changed scenes, the 3-D graphics software automatically filled in contours that had been occluded prior to the change and corrected the lighting of the scene. All scene images subtended 15.8° × 11.9° visual angle at a viewing distance of 1.13 m. Target objects subtended 2.41° on average along the longest dimension in the picture plane. The objects used for type and token changes were chosen to be approximately the same size as the initial target object in each scene. The full set of scene stimuli is listed in the Appendix.

Apparatus. The stimuli were displayed at a resolution of 800 × 600 pixels × 15-bit color. The display monitor refresh rate was set at 144 Hz. The room was dimly illuminated by an indirect, low-intensity light source. Eye movements were monitored using a dual-Purkinje-image eyetracker (Generation 5.5, Stanford Research Institute; for more information, see Crane & Steele, 1985). A bite-bar and forehead rest were used to maintain the participant's viewing position. The position of the right eye was tracked, though viewing was binocular. Eye position was sampled at a rate of better than 1000 Hz. Button presses were collected using a button panel connected to a dedicated input–output (I/O) card. The eyetracker, display monitor, and I/O card were interfaced with a 90-MHz Pentium-based microcomputer. The computer controlled the experiment and maintained a complete record of time and eye position values over the course of each trial.

Procedure. On arriving for the experimental session, participants were given a written description of the experiment along with a set of instructions. The description informed participants that their eye movements would be monitored while they viewed images of real-world scenes on a computer monitor. Participants were informed that they would view each image to prepare for a memory test on which they would have to "distinguish the original scenes from new versions of the scenes that may differ in only a small detail of a single object." In addition to the memory test instruction, participants were instructed to monitor each scene for object changes during study and to press a button immediately after detecting a change. The two types of possible changes were demonstrated using a sample scene. These instructions were the same as in Henderson and Hollingworth (1999b) and similar to instructions in other studies demonstrating transsaccadic change blindness (e.g., Grimes, 1996). Following review of the instructions, the experimenter calibrated the eyetracker by having participants fixate four markers at the centers of the top, bottom, left, and right sides of the display. Calibration was considered accurate if the computer's estimate of the current fixation position was within ±5 min of arc of each marker. The participant then completed the experimental session. Calibration was checked every three or four trials, and the eyetracker was recalibrated when necessary. To begin each trial, the participant fixated a central box on a fixation screen. The experimenter then initiated the trial.

Scene changes were initiated on the basis of eye position, as illustrated in Figure 2. In the change-after-fixation condition, an invisible region was initially activated surrounding the target object (Region A in Figure 2). This region was 0.36° larger on each side than the smallest rectangle enclosing the target object. When the eyes had dwelled within the target region continuously for at least 90 ms, the computer activated a change-triggering region surrounding a different object on the opposite side of the scene (Region B in Figure 2). The center of this region was on average 11.0° from the center of the target region. When the eyes crossed the boundary of the change-triggering region, the change was initiated. At a refresh rate of 144 Hz, the change was completed in a maximum of 14 ms. In the control condition, the procedure was identical except that the initial scene was replaced by an identical scene image as the eyes crossed the boundary of the change-triggering region. The procedure in the change-before-fixation condition was slightly different. At the beginning of the trial, an initial 4.9° (horizontal) × 3.9° (vertical) region was activated at the center of the screen (Region C in Figure 2). The participant's initial fixation on the scene fell within this region. The change-triggering region was activated as the eyes left the central region, and as in the other conditions, the change was initiated as the eyes crossed the boundary of the change-triggering region.

In the experimental session, each participant saw all 36 scenes. Six scenes appeared in the change-after-fixation condition, 18 in the change-before-fixation condition, and 12 in the control condition. The large number of change-before-fixation trials was included because in that condition, sometimes the target object would be fixated between the point that the eyes left the central region and the point when they crossed the change-triggering boundary. Trials on which this occurred were recoded as change-after-fixation trials. In each of the change conditions, the trials were evenly divided between type-change trials and token-change trials. Across the twelve participants, each scene appeared in each condition an equal number of times. Each scene was displayed for 20 s, and the order of image presentation was randomly determined for each participant. The study session lasted approximately 20 min.

After all 36 scenes had been viewed, the LTM test was administered. There was a delay of approximately 5 min between the study and test

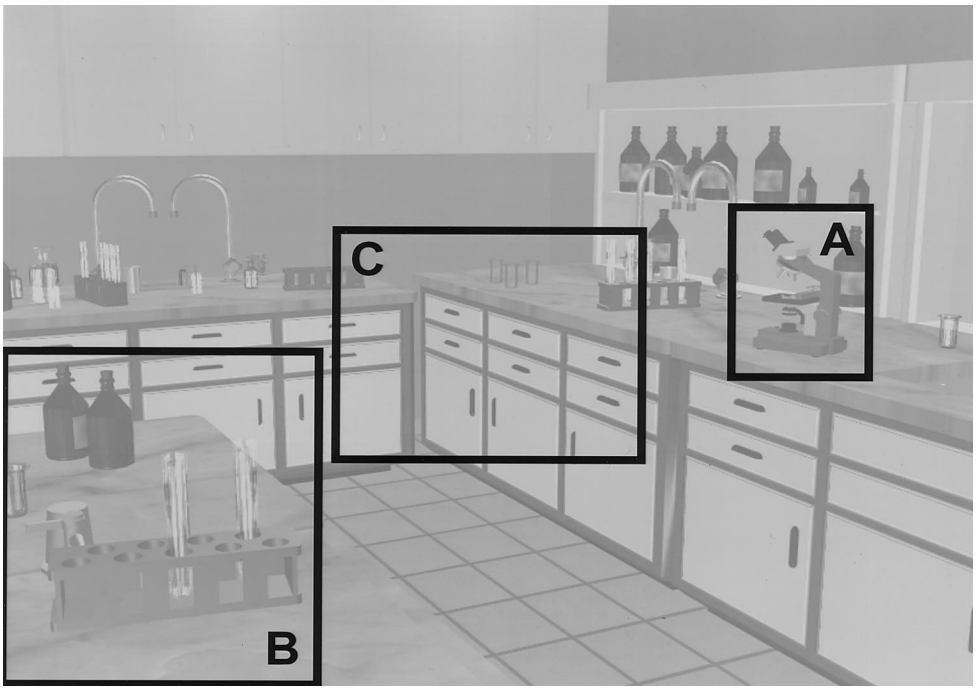

Figure 2. Sample scene (with contrast reduced) illustrating the software regions used to control scene changes in Experiment 1. Participants began by fixating the center of the screen. In the change-after-fixation condition, the computer waited until the eyes had dwelled in the target object region (A) for at least 90 ms. Then, the change-triggering region (B) was activated, and as the eye crossed the boundary to this region, the change was initiated. In the change-before-fixation condition, the computer waited until the eyes left the central region (C) before activating the change-triggering region (B), and the change was initiated as the eyes crossed the change-triggering boundary. The regions depicted here were not visible to the participants.

sessions, during which the experimenter reviewed the memory test instructions and demonstrated the paradigm using a sample scene. Thus, the retention interval for scenes varied from a minimum of about 5 min to a maximum of about 30 min. Memory was tested for the twelve scenes appearing in the control condition. Participants saw two versions of each scene sequentially: the studied scene, and a distractor scene that was identical to the studied scene except for the target object. In the type-discrimination condition, the distractor object was from a different basic-level category (identical to the changed target in the type-change condition); in the token-discrimination condition, the distractor target object was from the same basic-level category (identical to the changed target in the token-change condition). To ensure that participants based their decision on target object information, the target was marked with a small green arrow in both the studied and distractor scenes. Each version was presented for 8 s with a 1-s interstimulus interval. The order of presentation was counterbalanced. Participants were instructed to view each scene and then press one of two buttons to indicate whether the first or second version was identical to the scene studied earlier. Across participants, each scene item appeared in the type- and token-discrimination conditions an equal number of times.

Results

Online change detection performance. Eye movement data files consisted of time and position values for each eyetracker sample. Saccades were defined as changes in eye position greater than 8 pixels (about 8.8 min of arc) in 15 ms or less. Samples that did not fall within a saccade were considered part of a fixation. The position of each fixation was calculated as the mean of the position samples (weighted by the duration of time at each position) that fell between consecutive saccades (see Henderson, McClure, Pierce, & Schrock, 1997). Fixation duration was calculated as the elapsed time between consecutive saccades. Fixations less than 90 ms and greater than 2,000 ms were eliminated as outliers. Trials were eliminated if the eyetracker lost track of eye position prior to the change or if the change was not completed before the beginning of the next fixation on the scene. Eliminated trials accounted for 2.1% of the data. In addition, in the change-before-fixation condition, the target object was fixated before the change on 57.0% of the trials. These were recoded as change-after-fixation trials.

Mean percentage correct detection data are reported in Figure 3. When a change occurred after target fixation, we observed 51.1% correct type-change detection and 28.4% correct token-change detection, which were reliably different, $F(1, 11) = 8.66$, $MSE = 357.2$, $p < .05$. Performance in each of these conditions was reliably higher than the false alarm rate of 9.1% in the no-change control condition: type change versus false alarms, $F(1, 11) = 85.63$, $MSE = 123.7$, $p < .001$; token change versus false alarms, $F(1, 11) = 7.47$, $MSE = 299.5$, $p < .05$. When the change occurred before target fixation, we observed 8.8% correct detection in the type-change condition and 4.7% correct detection in the token-change condition, which did not differ ($F < 1$). Performance

in the change-before-fixation condition did not differ from the false alarm rate: type change versus false alarms, $F < 1$; token change versus false alarms, $F(1, 11) = 1.32$, $MSE = 85.2$, $p = .28$. Finally, comparison of the change-after-fixation condition with the change-before-fixation condition demonstrated that performance was reliably higher in the former compared with the latter, both for type changes, $F(1, 11) = 32.72$, $MSE = 327.0$, $p < .001$, and token changes, $F(1, 11) = 8.11$, $MSE = 413.2$, $p < .05$.

One potential explanation for poor detection performance in the change-before-fixation condition is that on average, changes occurred earlier in these trials compared with those in the change-after-fixation condition. Figure 4 plots detection performance as a function of the elapsed time to the change, both for the change-before-fixation and change-after-fixation conditions, collapsing across type and token change. There was a positive correlation between change detection and elapsed time to the change in the change-before-fixation condition, which approached reliability, $r_{pb} = .21$, $t(79) = 1.89$, $p = .065$.[3] This marginally positive correlation suggests that there may have been some encoding of target information without direct fixation, but even in the fourth quartile of the elapsed time distribution in this condition, detection performance (13.0%) was not much above the false alarm rate (9.1%). In addition, the elapsed time distributions overlapped for change before and after fixation. In the region of overlap, change detection after fixation on the target object was still clearly higher than when the change occurred before fixation on that object. Finally, there appeared to be little effect of elapsed time to change on detection in the change-after-fixation condition, $r_{pb} = .07$, $t(171) = 0.91$, $p = .36$. Thus, prior fixation of the target object clearly played a significant role in change detection.

Further evidence that target fixation plays a significant role in subsequent change detection comes from an analysis of fixation time on the target object prior to the change. In the change-after-fixation condition, mean total time fixating the target object prior to the change was 568 ms in the type-change condition and 622 ms

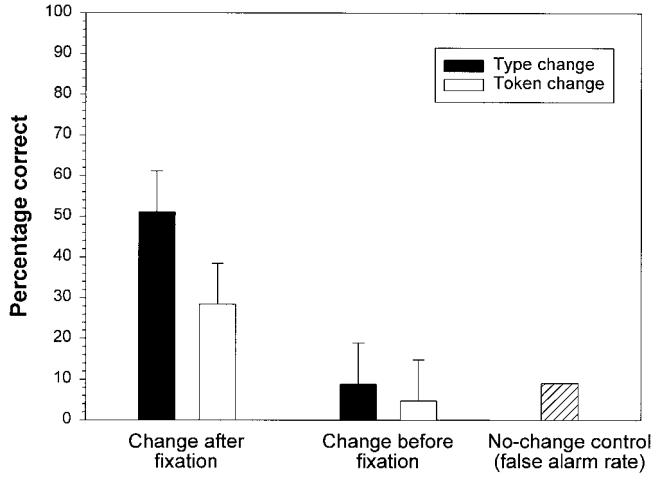

Figure 3. Experiment 1: Mean percentage correct change detection for each change condition and mean false alarms for the no-change control condition. Error bars are 95% confidence intervals, based on error term for the interaction between change condition (token or type) and eye position (change before or after fixation).

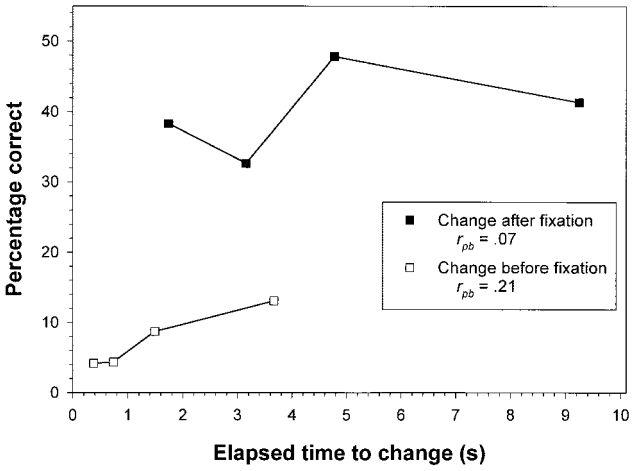

Figure 4. Experiment 1: Mean percentage correct change detection as a function of the elapsed time from the beginning of the trial to the change, for the change-before-fixation and change-after-fixation conditions (collapsing across type and token changes). In each condition, the mean of each elapsed time quartile is plotted against mean percentage detections in that quartile.

in the token-change condition. Figure 5 plots detection performance as a function of fixation time on the target prior to the change. The correlation between fixation time and detection performance was reliable in the token-change condition, $r_{pb} = .35$, $t(76) = 3.25$, $p < .005$, and approached reliability in the type-change condition, $r_{pb} = .18$, $t(83) = 1.64$, $p = .10$. Thus, at least for token changes, detection depended not only on whether the target was fixated prior to the change but also on the length of time the target was fixated.

The ability of participants to detect changes in this experiment, particularly token changes, is inconsistent with coherence theory, as the target object was not attended when the change occurred. However, object file theory could account for the change detection results if changes occurred soon enough after the object had been attended that the relevant object file had not been replaced by subsequent encoding. Thus, we examined detection performance in the change-after-fixation condition as a function of the number of fixations that intervened between the last exit of the eyes from the target region prior to the change and the change itself. There was an average of 4.7 fixations between the last exit from the target region and the change. Figure 6 plots detection performance as a function of the number of intervening fixations. Zero intervening fixations indicates that the saccade leaving the target object region crossed the boundary of the change-triggering region, triggering the change. However, contrary to the object file theory prediction, there was no evidence of decreasing detection perfor-

[3] In this and subsequent regression analyses, we regressed a predictor variable of interest (such as elapsed time to change) against the dichotomous detection variable, yielding a point-biserial coefficient. Each trial was treated as an observation. Because each participant contributed more than one sample to the analysis, variation due to differences in participant means was removed by including participant as a categorical factor (implemented as a dummy variable) in the model.

mance with an increasing number of intervening fixations: type change, $r_{pb} = -.07$, $t(83) = -0.61$, $p = .54$; token change, $r_{pb} = -.05$, $t(76) = -0.48$, $p = .64$.

For correct detections in the change-after-fixation condition, we examined the position of the eyes when the change was detected. The vast majority of detections came on refixation of the target object. On 93.2% of the trials, the detection button was pressed when the participant was refixating the target object after the change or within one eye movement after refixation. In addition, these detections tended to occur quite a long time after the change occurred. Mean detection latency in the change-after-fixation condition was 5.7 s.[4]

We were also interested in whether there might be effects of change not reflected in the explicit detection measure. For trials on which a change was not explicitly detected, we examined gaze duration (the sum of all fixation durations on an object region, from entry to exit from that region) on the target object for the first entry after the change. Miss trials in the change-after-fixation condition were compared with the equivalent entry in the no-change control condition. There was no difference between mean gaze duration on the changed object for miss trials in the type-change condition (477 ms) and the no-change control (479 ms; $F < 1$). For token changes, there was a trend toward elevated gaze duration for miss trials compared with the no-change control, with mean gaze duration of 649 ms for token-change misses versus 479 ms in the no-change control, $F(1, 11) = 2.40$, $MSE = 72{,}263$, $p = .15$.

Long-term memory performance. Mean percentage correct for the forced-choice memory test was calculated for type-discrimination and token-discrimination conditions. Contrary to the predictions of both coherence theory and object file theory, discrimination performance was well above the chance level of 50% correct, both for the type-discrimination condition (93.1%) and the token-discrimination condition (80.6%), which were reliably different, $F(1, 11) = 6.05$, $MSE = 154.5$, $p < .05$.

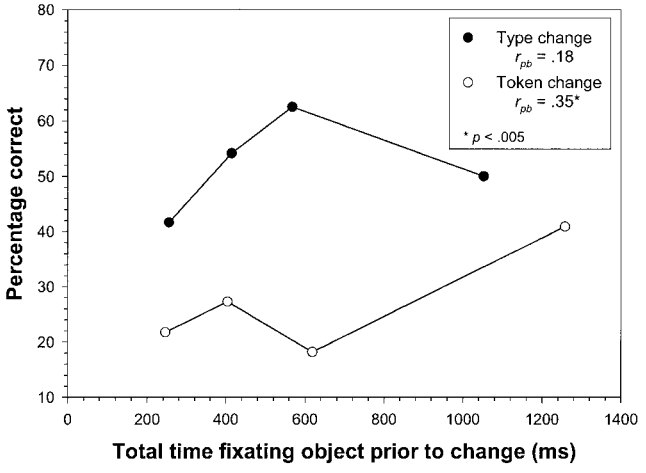

Figure 5. Experiment 1: Mean percentage correct change detection in the change-after-fixation condition, as a function of the total fixation time on the target object prior to the change. In each change type condition, the mean of each fixation time quartile is plotted against mean percentage detections in that quartile.

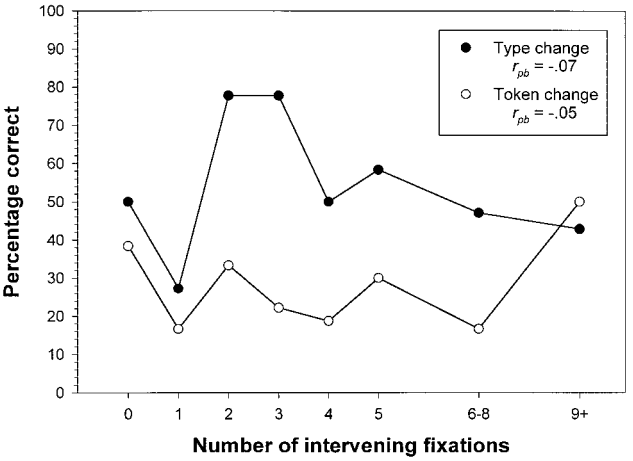

Figure 6. Experiment 1: Mean percentage correct change detection in the change-after-fixation condition, as a function of the number of intervening fixations between the last exit from the target region prior to the change and the change itself. Zero intervening fixations indicates that the saccade leaving the target object region crossed the change-triggering boundary, triggering the change.

Discussion

The principal issue in Experiment 1 was whether visual object representations persist after attention is withdrawn from an object, or whether such representations are transient, consistent with recent proposals in the transsaccadic memory and change blindness literatures. The data support the former view. Participants were able to detect both type and token changes when the changed object had been previously fixated and attended but was no longer within the focus of attention when the change occurred. Coherence theory (Rensink, 2000a) would appear unable to account for these data, particularly in the token-change condition, as that theory holds that coherent visual object representations disintegrate as soon as attention is withdrawn. The results are also inconsistent with object file theory (Irwin & Andrews, 1996), because detection often occurred many fixations after the last fixation on the target object, after the object file for the target should have been replaced by subsequent encoding. In addition, detection was significantly delayed after the change, more than 5 s on average, and typically until the target object had been refixated. This result suggests that visual information was often retained for a relatively long period of time and was consulted when focal attention and the eyes were directed back to the changed object. In summary, there appears to be significant accumulation of local scene information across multiple eye fixations on a scene.

Further evidence that visual information accumulates from previously fixated and attended regions of a scene comes from accurate discrimination performance on the LTM test. Discrimination

[4] Given the strong relationship between detection and refixation, we examined percentage correct in the change-after-fixation condition, eliminating trials on which the target was not refixated after the change. On only 4.4% of the trials did the participant fail to refixate the changed target object, so detection performance was raised only slightly with their elimination (type change, 53.6% correct; token change, 29.2% correct).

performance in both the type- and token-discrimination conditions was above 80% correct. These results are inconsistent with visual transience hypotheses but correspond with the picture memory literature (Friedman, 1979; Nelson & Loftus, 1980; Parker, 1978). Scene memory is clearly not limited to the gist or layout of the scene, or even to the identities of individual objects, because token-discrimination performance was quite accurate.

A puzzling issue given these results is why Irwin and Andrews (1996) found little evidence of visual accumulation across multiple eye movements. In that study, two fixations within an array of letters did not produce reliably better partial report performance than one. Although this result is consistent with the object file theory of transsaccadic memory, another aspect of Irwin and Andrews' data was not. Object file theory predicts that information from the most recently attended region of the array should be most often retained, as object files created earlier should be rapidly replaced. However, Irwin and Andrews found that partial report performance was better for array positions near the first saccade target rather than the second, suggesting that visual information from the region of the array attended earlier was preferentially retained over information from the region attended later. This complicates the interpretation of Irwin and Andrews' results considerably. In addition, the many methodological differences between our study and Irwin and Andrews make pinpointing the source of the discrepancy difficult: In Irwin and Andrews, stimuli consisted of letter arrays rather than natural scenes, letters were not directly fixated, and fixation durations and saccade targets were controlled by the experimenter. Whatever the source of the difference, the data from our study demonstrate that for free viewing of natural scenes, type- and token-specific information reliably accumulates from previously attended regions.

Our experiment also sought to shed light on the relationship between fixation position and change detection. The first issue was whether change detection depends on prior fixation of the target object. This was clearly the case, as change detection without prior target fixation was no higher than the false alarm rate. In addition, change detection performance increased with the length of time spent fixating the target prior to the change. Thus, in studies demonstrating change blindness, poor detection performance may have occurred, in part, because target regions were not always fixated prior to the change. The second issue was whether refixation of the target object plays an important role in change detection. The vast majority of detections came on refixation of the changed object, suggesting that refixation may cue the retrieval of stored information about a previously fixated and attended object. In studies demonstrating change blindness, then, poor detection performance could have also occurred because target regions were not always refixated after the change.

However, these potential explanations cannot fully account for change blindness phenomena. In this experiment, even when the target object was fixated before the change and again after the change, detection performance was still only modest, with 53.6% correct for type changes and 29.6% correct for token changes. It is important to note however, that visual transience theories cannot account for even this modest detection performance when attention has been withdrawn. One reason for an intermediate level of change detection performance may be that the change detection measure itself is not particularly sensitive to the detail of the scene representation. This possibility finds support in evidence from other studies demonstrating effects of change on trials without explicit detection (e.g., Hollingworth et al., 2001). In addition, the finding that forced-choice discrimination performance on the LTM test was apparently superior to performance on the online change detection test suggests the latter may not have reflected in full the information retained from previously attended objects. This issue was addressed in Experiment 3.

Exactly what is the nature of the information supporting detection and discrimination performance in this experiment? The possibility that sensory information was retained from previously fixated and attended objects can be ruled out, as prior research shows that such information is not retained across a single saccadic eye movement. Thus, it is likely that higher-level visual representations, abstracted away from sensory information, are retained across multiple fixations after attention is withdrawn and are ultimately stored in LTM. A large body of research indicates that such visual representations can be retained across eye movements (Carlson-Radvansky, 1999; Carlson-Radvansky & Irwin, 1995; Henderson, 1997; Henderson & Siefert, 1999; Pollatsek et al., 1984; Pollatsek et al., 1990). It is tempting to speculate that the advantage for type-change detection and type discrimination compared with token-change detection and token discrimination indicates that qualitatively different information was used to support performance in each case. For example, it is possible that for type-change detection and type discrimination, both information about the visual form of the object and also basic-level identity codes could have been used. Though plausible, it is difficult to conclude this was the case given that the visual difference between initial and changed objects in the two conditions was not controlled. In general, objects from the same category are more visually similar than objects from different categories. Thus, the possibility that visual information was solely functional in change detection and discrimination cannot be ruled out.

Experiment 2

The purpose of Experiment 2 was to strengthen the evidence that visual representations persist after attention is withdrawn from an object and are ultimately stored into LTM. In Experiment 1, this conclusion depended primarily on evidence from token manipulations. However, it is possible that the representations underlying this performance could have been conceptual in nature rather than visual. For example, if participants were to have encoded object identity at a subordinate category level, an identity code of *legal notebook* could have been sufficient to discriminate the original target from the changed target (a spiral notebook) in the office scene illustrated in Figure 1. Thus, in Experiment 2, a rotation manipulation was used (see Figure 1). The changed target object was created by rotating the initial target object 90° in depth (Henderson & Hollingworth, 1999b). In this condition, the identity of the target object was not changed at all, yet the visual appearance of the object was modified. If participants are able to successfully detect the rotation of a previously attended object and discriminate between two orientations of the same object on the subsequent LTM test, this would provide strong evidence that specifically visual information had been retained in memory.

For the online change detection task, in addition to the rotation manipulation, the token-change and control trials were retained from Experiment 1. The change-before-fixation condition was

eliminated; on all trials, the target object was changed only after it had been directly fixated at least once. Otherwise, Experiment 2 was identical to Experiment 1.

Method

Participants. Twelve Michigan State University undergraduates participated in the experiment for course credit. All participants had normal vision, were naive with respect to the hypotheses under investigation, and had not participated in Experiment 1.

Stimuli. To create the changed images in the rotation condition, the initial scene model was rendered after the target object model had been rotated 90° in depth. In addition, three scenes were modified slightly to accommodate the rotation condition. Two of these were minor modifications to target objects whose original appearances did not change significantly on rotation. The third change was to replace the book target object in a bedroom scene (which did not change much on rotation) with an alarm clock target.

Apparatus and procedure. The apparatus was the same as in Experiment 1. The procedure was the same as in Experiment 1, except that the type-change condition was replaced by a rotation condition. In addition, the change-before-fixation condition was eliminated. Twelve scene items appeared in each of the three change conditions: token change, rotation, and no-change control. The 12 control scenes served as the basis of the LTM test. Six of these scenes appeared in the token-discrimination condition and six in the orientation-discrimination condition. Across participants, each scene item appeared in each condition an equal number of times.

Results

Online change detection performance. Trials were eliminated if the eyetracker lost track of eye position prior to the change or if the change was not completed before the beginning of the next fixation on the scene. These accounted for 5.3% of the data. Fixations shorter than 90 ms or longer than 2,000 ms were eliminated as outliers.

Mean percentage correct detection data are reported in Figure 7. In all change trials, the change was made after the target object had been fixated at least once, equivalent to the change-after-fixation condition of Experiment 1. Detection performance was 26.0% correct in the token-change condition and 29.2% correct in the rotation condition, which did not differ ($F < 1$). Performance in each of these conditions was reliably higher than the false alarm rate of 4.2% in the no-change control condition: token change versus false alarms, $F(1, 11) = 11.32$, $MSE = 186.7$, $p < .005$; rotation versus false alarms, $F(1, 11) = 20.29$, $MSE = 185.8$, $p < .005$. Replicating Experiment 1, the vast majority of detections came on refixation of the target object (89.2%) and detection was significantly delayed, with mean detection latency of 4.6 s in the token-change condition and 4.5 s in the rotation condition.

As in Experiment 1, detection performance was influenced by the length of time the target object was fixated prior to the change. Mean total time fixating the target object region prior to the change was 768 ms in the token-change condition and 760 ms in the rotation condition. Figure 8 plots detection performance as a function of the length of time spent fixating the target object prior to the change. There was a reliable positive correlation between fixation time and percentage correct detection in both the token-change condition, $r_{pb} = .21$, $t(115) = 2.35$, $p < .05$, and the rotation condition, $r_{pb} = .26$, $t(124) = 2.96$, $p < .05$.

Above floor change detection for previously attended objects is not consistent with coherence theory. To test object file theory,

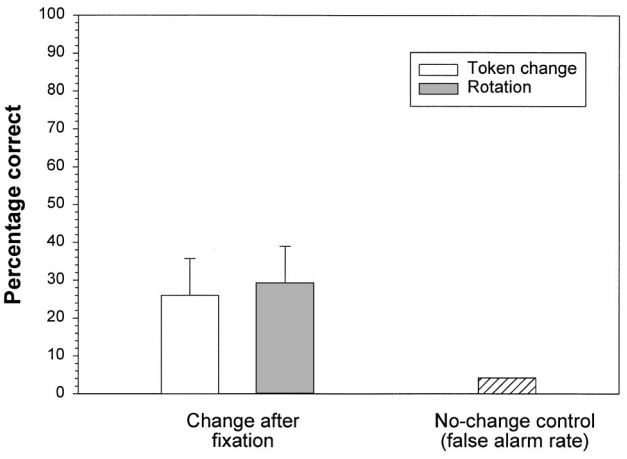

Figure 7. Experiment 2: Mean percentage correct change detection for each change condition and mean false alarms for the no-change control condition. Error bars are 95% confidence intervals, based on the error term of token-rotation contrast.

however, we again examined detection performance in the change conditions as a function of the number of fixations that intervened between the last exit of the eyes from the target region prior to the change and the change itself. There was an average of 4.8 fixations between the last exit from the target region and the change. Figure 9 plots detection performance as a function of the number of intervening fixations. Unlike Experiment 1, however, there was some evidence that detection performance fell with the number of intervening fixations: The rotation condition produced a marginally reliable negative correlation between the number of intervening fixations and detection performance, $r_{pb} = -.17$, $t(124) = -1.87$, $p = .06$, though no such effect was observed in the token-change condition, $r_{pb} = -.09$, $t(115) = -.97$, $p = .33$.

Finally, we examined whether there might be effects of change not reflected in the explicit detection measure. Gaze duration was calculated for the first entry of the eyes into the target region after the change. Miss trials in the change conditions were compared with the equivalent entry in the no-change control condition. For rotations, there was no reliable difference between mean gaze duration for miss trials (586 ms) compared with that for the no-change control (535 ms; $F < 1$). For token changes, there was again a trend toward elevated gaze duration for miss trials compared with that for the no-change control, with mean gaze duration of 655 ms for token-change misses versus 535 ms for the no-change control, $F(1, 11) = 2.48$, $MSE = 34{,}378$, $p = .14$. Because these analyses consulted only a subset of the data and had relatively little power, we combined the token-change and control data from Experiments 1 and 2 in a mixed analysis of variance (ANOVA) design, with experiment treated as a between-subjects factor. The combined analysis revealed a reliable 145-ms difference between gaze duration on changed objects for token-change misses (652 ms) compared with that for the no-change control (507 ms), $F(1, 22) = 4.71$, $MSE = 53{,}321$, $p < .05$. This implicit effect of token change on gaze duration has since been replicated (Hollingworth et al., 2001).

One potential alternative explanation for these gaze duration results needs to be examined.[5] In Experiments 1 and 2, detection performance was positively correlated with fixation time on the object prior to the change. Thus, detection trials eliminated from the above analysis of postchange gaze duration were more likely to have been trials on which the target had been fixated for a relatively long period of time prior to the change. If one entertains the additional assumption that there may be a baseline tendency for short fixation times on an object to be followed by longer fixation times on that object and vice versa (i.e., a negative correlation between early fixation times and later fixation times), then one might find elevated postchange gaze duration for token-change misses (which occur later in a trial) simply because more trials with longer initial fixation times had been eliminated from the analysis. To test whether there is a baseline tendency for early fixation times to be negatively correlated with later fixation times, we examined the no-change control condition for Experiments 1 and 2. The control condition allowed us to test this assumption independently of object changes. In Experiment 1, there was actually a reliable positive correlation between fixation time on the target prior to the change (a change in this condition was the substitution of an identical image) and gaze duration on the target for the first entry of the eyes after the change, $r_{pb} = .22$, $t(103) = 2.26$, $p < .05$. In Experiment 2, there was no observed correlation between these variables, $r_{pb} = .04$, $t(109) = 0.43$, $p = .67$. The tendency, at least in Experiment 1, for longer initial fixation times to be followed by longer fixation times later in viewing is consistent with the earlier finding that elevated gaze duration on semantically incongruous objects is observed not only for the first entry of the eyes into that object region but also for the second entry (Henderson, Weeks, & Hollingworth, 1999). Thus, the generally positive baseline relationship between initial and later fixation times on an object, combined with the elimination of more trials with longer initial fixation times in the analysis of token-change misses, would serve to underestimate mean postchange gaze duration for miss trials and would work against our finding of a significant implicit effect of token change on gaze duration.

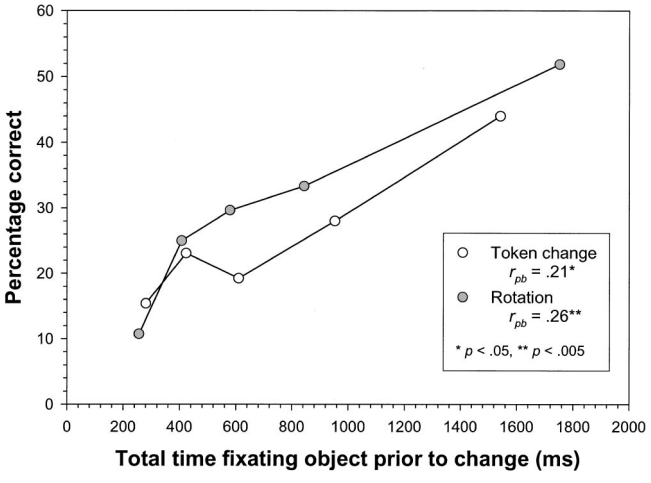

Figure 8. Experiment 2: Mean percentage correct change detection as a function of the total fixation time on the target object prior to the change. In each change type condition, the mean of each fixation time quintile is plotted against mean percentage detections in that quintile.

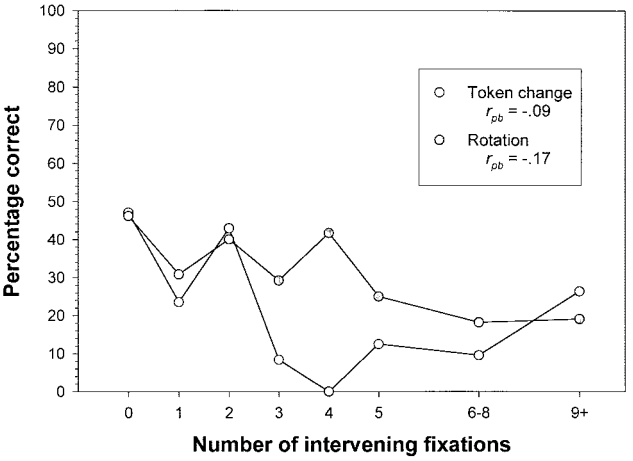

Figure 9. Experiment 2: Mean percentage correct change detection as a function of the number of intervening fixations between the last exit from the target region prior to the change and the change itself.

Long-term memory performance. Mean forced-choice discrimination performance was calculated for the token-discrimination and orientation-discrimination conditions. Contrary to the predictions of both coherence theory and object file theory, discrimination performance was well above chance performance of 50% correct, for both the token-discrimination condition (80.6%) and the orientation-discrimination condition (81.9%), which did not differ ($F < 1$).

Discussion

In Experiment 2, a rotation condition was included in which the visual form, but not the identity, of the target object changed between the initial and changed scene images. Contrary to the prediction derived from coherence theory, participants were able to detect rotations and token changes even though the object was not within the current focus of attention when the change occurred. Participants' ability to detect rotations provides converging evidence that specifically visual, as opposed to conceptual, representations were retained after attention was withdrawn. Unlike in Experiment 1, however, there was some evidence that detection performance fell as a function of the number of intervening fixations between the last exit of the eyes from the target object region and the change, consistent with the prediction of object file theory. This relationship was observed for rotations but not for token changes. Thus, the explicit detection data do not support coherence theory but are consistent, to some degree, with object file theory. The LTM test results, however, support neither the attention hypotheses nor object file theory. Both token- and orientation-discrimination performance was above 80% correct. Thus, although there appeared to be some decay of information relevant to the detection of rotation changes, token- and orientation-specific information was reliably retained in memory long after object file theory predicts such information should have been replaced.

[5] We thank Dan Simons for suggesting this alternative account.

In addition, Experiment 2 provided converging evidence that refixation serves as a strong cue to retrieve stored information from previous fixations. In replication of Experiment 1, change detection was delayed, on average, about 4.5 s after the change and typically until refixation of the changed object. In addition, change detection performance was influenced by the amount of time spent fixating the target prior to the change, supporting the conclusion drawn from Experiment 1 that change detection depends on prior target fixation.

Although rotation detection and orientation-discrimination performance in Experiment 2 cannot be attributed to an abstract coding of object identity, it is still possible that performance was mediated by the maintenance of nonvisual representations. Specifically, participants may have produced an abstract, verbal description of the visual properties of the target object (e.g., *yellow, lined, rectangular notebook with writing on the page, a black spine, and oriented so that the longer side is roughly parallel to the nearest edge of the table* would describe the notebook in Panel A of Figure 1 fairly well). If this were so, object memory may not have been visual in the sense that it was not based on representations in a visual format (though a verbal description of this sort would still preserve visual content, coding visual properties such as shape or color). Though possible, verbal encoding does not appear to provide a plausible account of performance in Experiments 1 and 2. First, participants could not have known beforehand which features would be critical to differentiating between the original target and the changed target. In addition, token and orientation trials were mixed together, so participants could not have known which type of task they would have to perform when encoding information from the scene. Thus, to support successful performance, and discrimination performance in particular, verbal descriptions would have had to have been quite detailed, encoding enough features from the original target so that a critical feature would happen to be encoded. Second, participants could not have known beforehand which of the objects in the scene was the target. Thus, they would have had to have produced a highly detailed verbal description of each of the objects in the scene. Third, a detailed verbal description must have been produced in a relatively short amount of time. In Experiments 1 and 2, participants fixated the target object for approximately 750 ms prior to the change and for approximately 1,500 ms prior to the memory test. In addition, as described in the results of Experiment 3, participants demonstrated token- and orientation-discrimination performance above 80% correct after having fixated the target object for only 702 ms on average prior to the test. Although a verbal description hypothesis cannot be definitively ruled out (in theory, a verbal description of unlimited specificity could be produced with enough time and enough words), it seems highly unlikely that participants could produce verbal descriptions for each of the objects in a scene, with each description detailed enough to perform accurate token and orientation discrimination, and do this within approximately 700 ms per object.

Experiment 3

Accurate discrimination performance in the LTM tests of Experiments 1 and 2 provides strong evidence that visual scene information is retained in LTM. However, the fact that performance in the online change detection task was approximately 30% correct for token and rotation changes doesn't allow the very strongest conclusion that the representation formed during online scene perception contains visual information from previously attended objects. One could reasonably argue that accurate LTM performance could not occur unless the information supporting that performance had been present during the online perceptual processing of the scene. In addition, any evidence of above-floor detection performance in the absence of sustained attention is inconsistent with visual transience theories, in general, and with coherence theory, in particular. Nevertheless, it remains the case that modest change detection performance is typically interpreted as evidence for the absence of representation.

There are a number of reasons, however, why online change detection performance may have underestimated the specificity of the scene representation, compared with, in particular, the forced-choice task used in the LTM tests. First, the online change detection task was performed concurrently with the task of studying for the memory test. Thus, change detection may have underestimated the detail of the scene representation, because participants could not devote their full attention to monitoring for object changes. Second, in the forced-choice discrimination test, the target object was specified with a green arrow. Thus, participants could limit retrieval to information about the target object. However, such focused analysis was not possible in the online change detection task, because the target was not specified. Finally, explicit change detection, regardless of other task demands, may not be very sensitive to visual representation (as reviewed above with regard to implicit effects), especially if participants adopt a fairly high criterion for change detection. By forcing participants to make a choice between two alternatives, information unavailable or insufficient for explicit detection could nevertheless influence performance. In support of these points, there is direct evidence from Experiments 1 and 2 that explicit change detection did not reflect the full detail of the scene representation constructed online, as gaze duration on the changed object for token-change miss trials was reliably longer compared with the same entry when no change had occurred.

In Experiment 3, then, we used a forced-choice discrimination procedure to test the representation of previously attended objects during the online perceptual processing of a scene. Figure 10 illustrates the sequence of events in a trial in Experiment 3. As in the change-after-fixation conditions of Experiments 1 and 2, the computer waited until the participant had fixated the target object, at which point a second region was activated around another object in the scene. When the eyes crossed the boundary to this second region, instead of changing the target object, the target object was masked by a speckled, green rectangular field slightly larger than the object itself. Participants were instructed to fixate this mask and press a button to continue. As in the LTM tests of Experiments 1 and 2, participants were then shown two object alternatives in sequence, one of which was identical to the initial target. The distractor was either a different token (token-discrimination condition) or the same object rotated 90° in depth (orientation-discrimination condition). Participants indicated whether the first or second object alternative was the same as the one initially present in the scene.

This paradigm replicates the encoding conditions of the change detection trials in Experiments 1 and 2, yet uses a forced-choice

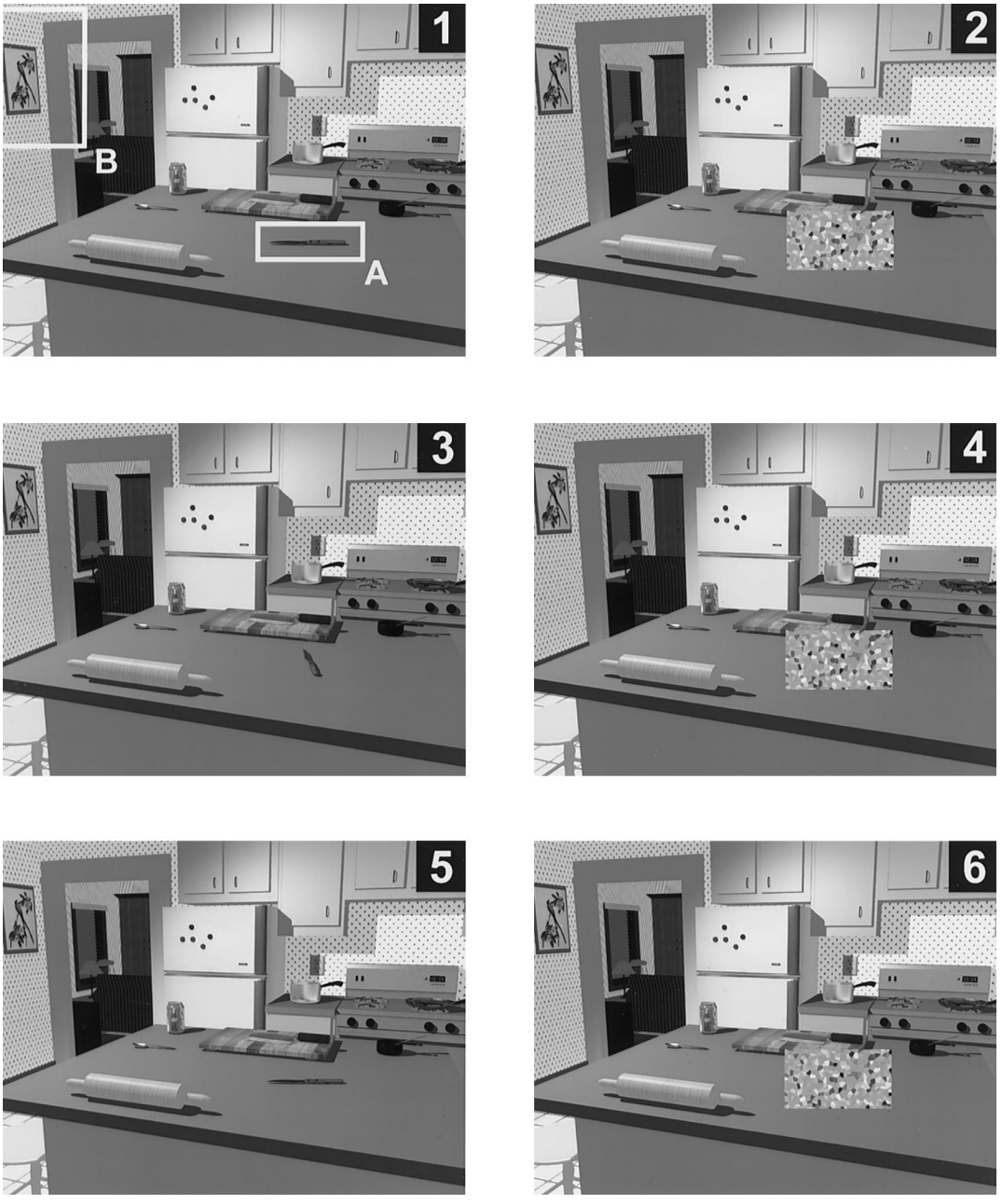

Figure 10. Sequence of events in an orientation-discrimination trial in Experiment 3. Panel 1: Initial scene image (the regions depicted in the figure were not visible to participants). Participants began by fixating the center of the screen. The computer waited until the eyes had dwelled within the target object region (Region A) for at least 90 ms. Then, a second region (Region B) was activated around a different object in the scene. As the eye crossed the boundary to Region B, the target object was occluded by a salient mask (Panel 2). The mask remained visible until the participant pressed a button to begin the forced-choice test. After a delay of 500 ms, the first target object alternative was displayed for 4 s (Panel 3), followed by the target object mask for 1 s (Panel 4), the second target object alternative for 4 s (Panel 5), and the target object mask (Panel 6), which remained visible until response.

discrimination procedure similar to that used in the LTM tests. This method should eliminate the factors that could have caused the change detection tasks of Experiments 1 and 2 to underestimate the detail of scene representation. First, the instruction to study for an LTM test was eliminated, so participants had only one task to perform, the discrimination task. Second, the critical object was specified (by the mask), so participants could limit analysis to the target. Third, the potentially more sensitive forced-choice proce-

dure was used. Given the results of the first two experiments, participants should be able to perform this task very accurately (i.e., above 80% correct), which would provide strong evidence that visual information is retained from previously attended objects during online scene perception. In contrast, visual transience hypotheses predict poor discrimination performance. Coherence theory predicts 50% discrimination performance (i.e., chance), because attention is withdrawn from the target object prior to the onset of the mask. Object file theory predicts that discrimination performance should fall to chance as the eyes and attention are oriented to new objects and the object file from the target is replaced.

Method

Participants. Twelve Michigan State University undergraduates participated in the experiment for course credit. All participants had normal vision, were naive with respect to the hypotheses under investigation, and had not participated in Experiments 1 or 2.

Stimuli. The stimuli were the same as in Experiment 2, with minor modifications to three of the scene items. In these scenes, a few more objects were added, and the target object was moved closer to the center of the scene. These modifications were part of an effort to improve the scene stimuli and were not related to any experimental manipulation. The green mask in each scene 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.

Apparatus. The apparatus was the same as in Experiment 1.

Procedure. Participants were informed that their eye movements would be monitored while they viewed images of real-world scenes on a computer monitor. They were instructed that at some point during the viewing of each scene, a bright green, speckled box would appear, concealing an object in the scene. When they saw the box, they were to look directly at it and press a button to continue. After a brief delay, two objects were displayed in succession at that position, only one of which was identical to the original object. Participants were instructed that after presentation of the two alternatives, they were to press the left-hand button on the button box if the first alternative was identical to the original object, or the right-hand button if the second alternative was identical to the original. The two types of possible distractors were described using a sample scene. Following review of the instructions, the experimenter calibrated the eyetracker as described in Experiment 1.

Each trial began with the participant fixating the center of the screen. The computer waited until the eyes had dwelled in the target object region for at least 90 ms. Then, the second region (the change-triggering region in Experiments 1 and 2) was activated around a different object in the scene. As the eye crossed the boundary to this region, the target object was masked. When the button was pressed to begin the discrimination test, there was a delay of 500 ms, followed by the first object alternative display for 4 s, the target mask for 1 s, the second object alternative for 4 s, and the target mask, which remained visible until response. To avoid exceedingly long trials, if the mask had not appeared by 20 s into viewing, it was displayed regardless of eye position at that point.

Participants first completed a practice session of four trials, two in each of the discrimination conditions. Participants then completed the experimental session, in which they viewed all 36 scenes, 18 in each of the discrimination conditions. The original target was the first alternative on half the trials and the second alternative on the other half. The assignment of scene items to conditions was counterbalanced between participant groups. The order of image presentation was determined randomly for each participant. The entire session lasted approximately 20 min.

Results

On 25 trials (5.8%), the test had not been initiated by 20 s into viewing, and the target object was masked at that point. On one of these trials, the participant was fixating the target object when the mask appeared. This trial was eliminated, along with trials on which the target was not fixated for at least 90 ms prior to the onset of the mask. A total of 3.5% of the trials was removed. Eye fixations shorter than 90 ms or longer than 2,500 ms were eliminated as outliers.

Consistent with results from the LTM tests of Experiments 1 and 2, forced-choice discrimination performance was quite accurate, with 86.9% correct in the token-discrimination condition and 81.9% correct in the orientation-discrimination condition. The trend toward superior token-discrimination performance was not reliable, $F(1, 11) = 2.54$, $MSE = 119.8$, $p = .14$. There was, however, a reliable and unanticipated interaction between discrimination condition and the order of target–distractor presentation in the forced-choice test, $F(1, 11) = 12.05$, $MSE = 124.0$, $p < .01$. For token discrimination, there was little difference between the target-first condition (88.8% correct) and target-second condition (85.1% correct). However, for orientation discrimination, there was a large difference between target first (72.6% correct) and target second (91.2% correct). In the orientation-discrimination condition, participants apparently were biased to respond "second," but such a bias does not compromise the main finding of accurate performance in both the token- and orientation-discrimination conditions.

In addition, we investigated whether performance was influenced by the length of time spent fixating the target object prior to test. Mean total time fixating the target object prior to test was 725 ms in the token-discrimination condition and 678 ms in the orientation-discrimination condition. These values are roughly equivalent to the amount of time spent fixating the target object prior to the change in Experiments 1 and 2, suggesting that the encoding conditions of the online change detection task were successfully replicated. Figure 11 plots discrimination performance as a function of the length of time spent fixating the target object prior to the test. There was a reliable positive correlation between fixation time and performance in the orientation-discrimination condition, $r_{pb} = .15$, $t(193) = 2.08$, $p < .05$, but no reliable correlation in the token-discrimination condition, $r_{pb} = .04$, $t(196) = 0.60$, $p = .55$.

We also examined discrimination performance as a function of the number of fixations that intervened between the last exit of the eyes from the target region prior to the onset of the mask, and the mask's onset. There was an average of 4.6 fixations between the last exit from the target region and the onset of the mask. Figure 12 plots detection performance as a function of the number of intervening fixations. Contrary to the prediction of object file theory, there was no evidence that discrimination performance fell as the number of intervening fixations increased: token discrimination, $r_{pb} = .00$, $t(196) = -0.05$, $p = .96$; orientation discrimination, $r_{pb} = .11$, $t(193) = 1.59$, $p = .11$. In the token-discrimination condition, when nine or more fixations intervened between last exit and the onset of the mask (range = 9 to 42 fixations; $M = 15.3$ fixations), performance was 85.3% correct. In the orientation-discrimination condition, when nine or more fixations intervened between last exit and the onset of the mask

(range = 9–58 fixations; M = 16.7 fixations), performance was 92.3% correct.

Discussion

Experiment 3 used a forced-choice procedure to test the online representation of previously attended objects in natural scenes. During viewing, after the target object had been fixated, it was masked as the eyes and focal attention were directed to a different object in the scene. Memory for the target object was then tested using a forced-choice procedure. Participants demonstrated accurate token- and orientation-discrimination performance, above 80% correct in each condition, even though the target object was not attended when the test was initiated. This result provides strong evidence against the claim of coherence theory that coherent visual representations disintegrate as soon as attention is withdrawn from an object. If this were the case, then performance on the discrimination task should have been at chance. In addition, these results do not support the object file theory of transsaccadic memory (Irwin, 1992a), as there was no evidence of decreasing discrimination performance with the number of intervening fixations between the last exit of the eyes from the target object region and the onset of the mask. Instead, these data support a view of scene perception in which visual representations accumulate in memory from fixated and attended regions of a scene.

General Discussion

The three experiments reported in this study were designed to investigate the nature of the information retained from previously fixated and attended objects in natural scenes. The principal question was whether visual information is retained from previously attended objects, consistent with evidence from the picture memory literature (e.g., Friedman, 1979; Standing et al., 1970), or whether visual object representations decay rapidly after attention is withdrawn from an object, as proposed by visual transience hypotheses of scene representation (e.g., Irwin & Andrews, 1996; Rensink, 2000a). In Experiment 1, target objects in natural scenes were changed during a saccade to another object in the scene, but only after the target had been fixated directly at least once. The target was replaced with another object from a different basic-level category (type change) or from the same basic-level category (token change). In addition, LTM for target objects in the scenes was tested using a forced-choice procedure. Participants successfully detected both type and token changes on a significant proportion of trials, even though the target object was not attended when the change occurred. In addition, participants accurately discriminated between original targets and distractor objects that differed either at the level of type or token. In Experiment 2, a rotation condition was included as a more stringent test of whether visual representations persist after attention is withdrawn. Participants not only detected the rotation of previously attended target objects, but also accurately discriminated between two orientations of the same object on the LTM test. In Experiment 3, a forced-choice procedure was used to test the online representation of previously attended objects in natural scenes. During scene viewing, participants were asked to discriminate between the original target object and a different-token or different-orientation distractor. Discrimination performance for previously attended objects was quite accurate, above 80% correct.

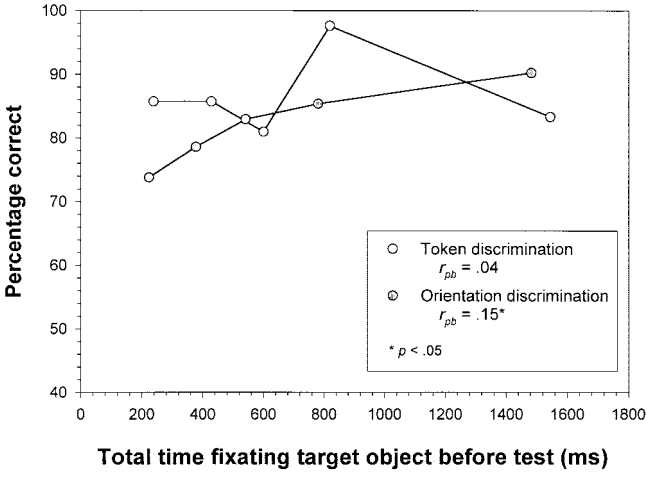

Figure 11. Experiment 3: Mean percentage correct discrimination performance as a function of the total fixation time on the target object prior to test. In each discrimination condition, the mean of each fixation time quintile is plotted against mean percentage correct in that quintile.

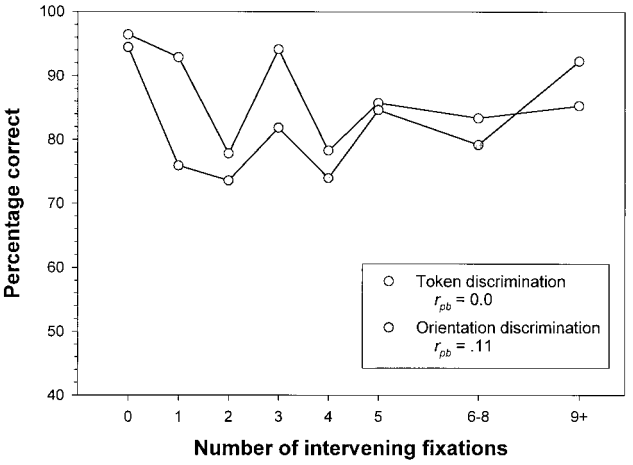

Figure 12. Experiment 3: Mean percentage correct discrimination performance as a function of the number of intervening fixations between the last exit from the target region prior to the test and the onset of the target object mask.

These results are not consistent with the proposal that visual representation is limited to the currently attended object (O'Regan, 1992; O'Regan et al., 1999; Rensink, 2000a, 2000b; Rensink et al., 1997; Simons & Levin, 1997; Wolfe, 1999). This view predicts that token and rotation changes should not be detected in the absence of attention and, additionally, that forced-choice discrimination should be at chance if attention was not allocated to the critical object when it was masked. The current change detection experiments initiated a change during a saccade to a completely different object in the scene. We have obtained converging data in experiments in which a change was made during a saccade that took the eyes away from the target object after it had been fixated the first time (Henderson & Hollingworth, 1999b; Hollingworth et al., 2001). Because attention precedes the eyes to the next fixation

position, the target object was not attended when the change occurred, as in the current study. Yet token and rotation changes were detected at rates significantly above floor. Thus, the proposal that sustained attention to a changing object is necessary to detect that change (Rensink, 2000a, 2000b; Rensink et al., 1997; Simons, 2000; Simons & Levin, 1997) is disconfirmed by this study and by converging data from related studies.

In addition, these results are inconsistent with portions of the object file theory of transsaccadic memory (Irwin, 1992a, 1992b; Irwin & Andrews, 1996). This theory holds that three or four object files, which maintain detailed visual information from attended objects, can be retained in VSTM but are quickly replaced as attention and the eyes are directed to new perceptual objects. This view predicts that detection and discrimination performance should fall quickly to zero or chance as the number of intervening fixations increases between the last exit of the eyes from the target object region and the change (in the change detection paradigm) or the onset of the mask (in the forced-choice discrimination paradigm). Although there was a reliable drop in change detection performance as a function of the number of intervening fixations for rotations in Experiment 2, the remaining five analyses showed no such effect. In particular, the more sensitive forced-choice discrimination measure used in Experiment 3 appeared to be entirely independent of the number of intervening fixations. In addition, Irwin's object file theory cannot account for successful discrimination performance on the LTM tests, as object files could not have been retained in VSTM from study to test. Thus, although there may be some decay of visual information encoded from previously attended objects, visual object representations are nonetheless reliably and stably retained from previously attended objects. However, these results are only inconsistent with the portion of Irwin's object file theory dealing with the representational fate of previously attended objects. The bulk of the theory, which concerns the retention and integration of information across single eye movements, and particularly from the attended saccade target, is not compromised by the findings of this study. In fact, the model we describe subsequently draws heavily from object file theory yet provides a different account of visual representation after the withdrawal of attention.

The LTM tests provide strong converging evidence that visual object representations are retained after attention is withdrawn and suggest that such representations are quite stable over the course of a 5–30-min retention interval. These results provide a bridge between the literature on LTM for pictures and the literature on scene memory across saccades and other visual disruptions. The LTM data from this study are consistent with prior evidence showing accurate memory for the visual form of whole scenes (Standing et al., 1970) and of individual objects in scenes (Friedman, 1979; Parker, 1978). In addition, the current study provides a stronger test of long-term scene memory compared with previous studies, because the scenes themselves were relatively complex, participants viewed each scene only once, between-item similarity was high for studied scenes, and distractors in the forced-choice discrimination test differed from targets in only the properties of a single object. One of the objectives of this study was to resolve the discrepancy between evidence of excellent picture memory and recent proposals, derived from change detection studies, that visual object representations are transient. The discrepancy appears to be resolved: Visual object representations are reliably and stably retained from previously attended objects during online scene perception and are stored into LTM. Visual object representation is not transient.

In addition to the main question of the representation of previously attended objects, we investigated three secondary questions. First, we sought to determine whether change detection depends on the prior fixation of the target object. This was indeed the case. In Experiment 1, a condition in which the target was changed before it was directly fixated produced detection performance that did not differ from the false alarm rate and was reliably poorer than detection performance when the target had been fixated prior to the change. In addition, a positive relationship between fixation time on the object prior to the change and detection performance was observed in Experiments 1 and 2. Thus, the encoding of scene information appears to be a strongly influenced fixation position, consistent with prior reports (Henderson & Hollingworth, 1999b; Hollingworth et al., 2001; Nelson & Loftus, 1980). Second, we examined the role of refixation of a changed object in the detection of that change. The vast majority of correct detections in Experiments 1 and 2 came on refixation of the changed target. Thus, refixation appears to play an important role in the retrieval of a stored object representation and the comparison of that representation to current perceptual information (see also Henderson & Hollingworth, 1999b; Hollingworth et al., 2001; Parker, 1978). Finally, we were interested in whether explicit change detection performance provides an accurate measure of the detail of the visual scene representation. In Experiments 1 and 2, when a token change was not explicitly detected, gaze duration on the changed object was reliably longer than when no change occurred, an effect that has subsequently been replicated (Hollingworth et al., 2001). Thus, the current data provide evidence that explicit change detection performance underestimates the detail of the visual scene representation.

Together these data provide an explanation for why change blindness may occur, despite strong evidence from this study that visual representations persist after the withdrawal of attention. First, in studies demonstrating change blindness, eye movements have rarely been monitored. Thus, changes may be missed simply because the target object was not fixated prior to the change. If detailed information had not been encoded from a target object, it is hardly surprising that a change to that object would not be detected. Providing further support for this idea, Hollingworth et al. (2001) monitored eye movements during a flicker paradigm (see Rensink et al., 1997) using scenes similar to those in this study. Over 70% of object deletions and over 90% of object rotations were detected only when the changing object was in foveal or near-foveal vision. Second, even if the object representation is detailed enough to discriminate between the initial and changed targets, it may not be reliably retrieved to support change detection. Our results demonstrate that changes are often detected only when the changed object is refixated after the change (see also Henderson & Hollingworth, 1999b; Hollingworth et al., 2001). Again, because most change detection paradigms do not monitor eye movements, changes may be missed because the changed region is not refixated. Finally, even if a target object is fixated before and after the change, changes may go undetected not because the relevant information is absent from the scene representation, but because the explicit detection measure is not sensitive to the presence of that information, as has been amply dem-

onstrated by studies such as this one, showing implicit effects of change (Fernandez-Duque & Thornton, 2000; Hayhoe et al., 1998; Hollingworth et al., 2001; Williams & Simons, 2000). In summary, it has been known since the early 1980s that a global sensory image is not constructed by the visual system and retained across visual disruptions, such as eye movements. However, poor change detection performance does not necessarily indicate the absence of visual representation.

A Descriptive Model of Scene Perception and Memory

If visual object representations are retained in memory after attention is withdrawn from an object, in what type of memory store is this information maintained? Clearly, the LTM tests demonstrate that fairly detailed information is retained in LTM, but what accounts for online change detection performance in Experiments 1 and 2 and online discrimination performance in Experiment 3? Three strands of evidence suggest that performance was, to a large degree, supported by the maintenance of visual object representations in LTM during the online perceptual processing of the scene, rather than in VSTM. First, if current estimates of the capacity of VSTM are correct, it is unlikely that target object information could have been retained in VSTM during the interval between the last fixation on the target object and the change or discrimination test. In Experiment 3, discrimination performance was highly accurate, even when more than nine separate fixations intervened between the last exit and the onset of the target object mask. Second, the similarity between online discrimination (Experiment 3) and long-term discrimination (Experiments 1 and 2) suggests that performance in each was supported by a similar set of processes. Third, Hollingworth et al. (2001) found that online change detection performance was strongly influenced by the semantic consistency between that target object and the scene in which it appeared, a variable known to influence the LTM representation of an object (e.g., Friedman, 1979). It therefore appears that LTM plays an important role in online scene perception (see also Chun & Nakayama, 2000). Given the amount of visual information available for analysis from a natural scene and the length of time that we may be present in the same visual environment, the visual system takes advantage of the capacity of LTM to store potentially relevant information for future analysis, such as the detection of changes to the environment.

The data from this study can be accommodated by the following model. It takes as its foundation current theories of episodic object representation (e.g., Henderson, 1994; Kahneman, Treisman, & Gibbs, 1992), and is broadly consistent with the object file theory of transsaccadic memory (Irwin, 1992a), but proposes a large role for LTM in the online construction of a scene representation. As discussed in the introduction, dynamic scene perception faces two memory problems: (a) the short-term retention and integration of scene information across single saccadic eye movements, particularly from the attended saccade target, and (b) the longer-term retention and potential integration of information from previously fixated and attended objects. The model proposed here is limited in scope to the second issue. In particular, it concerns the nature of the representations produced when attention and the eyes are oriented to an object, the retention of object information when attention and the eyes are withdrawn, the integration of object information within a scene-level representation, and the subsequent retrieval of that information. It rests on the following assumptions.

First, when attention and the eyes are oriented to a local object in a scene, in addition to low-level sensory processing, visual processing leads to the construction of representations at higher levels of analysis. These may include a visual description of the attended object, abstracted from low-level sensory properties, and conceptual representations of object identity and meaning. Higher-level visual representations can code quite detailed information about the visual form of an object, specific to the viewpoint at which the object was observed (Riesenhuber & Poggio, 1999; Tarr, Williams, Hayward, & Gauthier, 1998), and viewpoint-specific object representations can be retained across eye movements (Henderson & Siefert, 1999, in press).

Second, these abstracted representations are indexed to a position in a map coding the spatial layout of the scene, forming an object file (Kahneman & Treisman, 1984; Kahneman et al., 1992). This view of object files (described in detail in Henderson, 1994; Henderson & Anes, 1994; Henderson & Siefert, 1999) differs from earlier proposals (e.g., Kahneman et al., 1992) in that object files preserve abstracted visual representations rather than sensory information and also support the short-term retention of conceptual codes. Thus, objects files instantiate not only VSTM but also conceptual short-term memory (CSTM; see Potter, 1999).

Third, processing of abstracted visual and conceptual representations in short-term memory and the indexing of these codes to a particular spatial position leads to their consolidation in LTM. The LTM codes for an object are likewise indexed to the spatial position in the scene map from which the object information was encoded, forming what we term a *long-term memory object file*.

Fourth, when attention is withdrawn from an object, the short-term memory representations decay quite rapidly, leaving only the spatially indexed, long-term memory object files, which are relatively stable. Whether short-term memory decay is immediate or whether short-term memory information persists until replaced by subsequent encoding is not central to our proposal. However, the fact that changes to objects on the saccade away from that object are often detected immediately (Henderson & Hollingworth, 1999b; Hollingworth et al., 2001) suggests that visual object representations can be retained in VSTM at least briefly after attention is withdrawn from an object, consistent with Irwin's (1992a) view of VSTM.

Thus, over multiple fixations on a scene, local object information accumulates in LTM from previously fixated and attended regions and is indexed within the scene map, forming a detailed representation of the scene as a whole (though clearly less detailed than a sensory image, as the visual representations stored from local regions are abstracted away from sensory properties such as precise metric organization). In contrast to high-capacity LTM storage, only a small portion of the visual information in a scene is actively maintained in short-term stores, and the moment-by-moment content of VSTM and CSTM is dictated by the allocation of attention.

Fifth, the retrieval of LTM codes for previously attended objects and the comparison of this information with current perceptual representations is strongly influenced by the allocation of visual attention and thus by fixation position. Access to the contents of an object file in VSTM is proposed to be dependent on attending to the spatial position at which the file is indexed, a proposal that is

supported by evidence that preview effects are mediated by spatiotemporal continuity (Kahneman et al., 1992). Evidence for spatially mediated LTM retrieval in this study comes from that fact that changes were detected on refixation of the target object. In addition, fixating the changed object led to change detection in the type- and token-change conditions, even though the original object was no longer present and could not act as a retrieval cue. Moreover, in Henderson & Hollingworth (1999b), object deletions were sometimes detected only when the participant fixated the spatial position in the scene where the object had originally appeared. Clearly, the original object could not serve as a retrieval cue in this paradigm, as it had been deleted, suggesting that attending to the original spatial position of the target led to the retrieval of its long-term memory object file and subsequent change detection.

Sixth, the retrieval from LTM of higher-level visual codes specific to the viewed orientation of a previously attended object accounts for participants' ability to detect token and rotation changes and to perform accurately on token- and orientation-discrimination tests.

Finally, when the scene is removed, the LTM representation consists of the scene map with indexed local object codes. During subsequent perceptual episodes with the scene, the scene map is retrieved, and local object information can be retrieved by attending to the position in the scene at which information about that object was originally encoded, leading to successful performance on the LTM tests. How the correct scene map is selected is an interesting question, the answer to which lies beyond the scope of this model.

In summary, the model holds that a relatively detailed representation of a scene is constructed as the eyes and attention are directed to multiple local regions. In addition, encoding into and retrieval from this representation are controlled by the allocation of visual attention and thus by fixation position, given the tight coupling between attention and the eyes during normal scene viewing. The principal difference between this model of scene perception and visual transience hypotheses is the proposal that visual representations persist after attention is withdrawn, are stored in LTM, and form the basis of a fairly detailed scene-level representation. However, our model is consistent with the proposal of visual transience hypotheses that object representations in VSTM decay quickly once attention is withdrawn. In fact, this model is consistent with object file theory except for an additional form of representation, long-term memory object files, as object file theory has no mechanism for long-term storage. This additional representation, however, has significant implications for the nature of the representation constructed from a scene. Thus, our model describes a means by which relevant visual information can be stored and retrieved to support such processes as perceptual comparison, motor interaction, navigation, or scene recognition, while retaining the view that active visual representation is essentially local and transient, governed by the allocation of attention.

References

Archambault, A., O'Donnell, C., & Schyns, P. G. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science, 10,* 249–255.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral & Brain Sciences, 20,* 723–767.

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

Brietmeyer, B. G., Kropfl, W., & Julesz, B. (1982). The existence and role of retinotopic and spatiotopic forms of visual persistence. *Acta Psychologica, 52,* 175–196.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex, 3,* 247–260.

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.

Chun, M. M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across views. *Visual Cognition, 7,* 65–81.

Crane, H. D., & Steele, C. M. (1985). Generation-V dual-Purkinje-image eyetracker. *Applied Optics, 24,* 527–537.

Currie, C., McConkie, G., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception & Psychophysics, 62,* 673–683.

Davidson, M. L., Fox, M. J., & Dick, A. O. (1973). Effect of eye movements on backward masking and perceived location. *Perception & Psychophysics, 14,* 110–116.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36,* 1827–1837.

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

Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavioral & Brain Sciences, 8,* 265–289.

Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition, 7,* 324–344.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316–355.

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

Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7,* 43–64.

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

Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General, 123,* 410–426.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8,* 51–55.

Henderson, J. M., & Anes, M. D. (1994). Effects of object-file review and type priming on visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 826–839.

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. (1999a). High-level scene perception. *Annual Review of Psychology, 50,* 243–271.

Henderson, J. M., & Hollingworth, A. (1999b). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 10,* 438–443.

Henderson, J. M., & Hollingworth, A. (in press). Eye movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes

(Eds.), *Analytic and holistic processes in the perception of faces, objects, and scenes.* New York: JAI/Ablex.

Henderson, J. M., Hollingworth, A., & Subramanian, A. N. (1999, November). *The retention and integration of scene information across saccades: A global change blindness effect.* Paper presented at the annual meeting of the Psychonomic Society, Los Angeles, CA.

Henderson, J. M., McClure, K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics, 59,* 323–346.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & Psychophysics, 45,* 196–208.

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

Henderson, J. M., & Siefert, A. B. C. (in press). Types and tokens in transsaccadic object integration. *Psychonomic Bulletin & Review.*

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 & Psychophysics, 57,* 787–795.

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

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

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23,* 420–456.

Irwin, D. E. (1992a). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 307–317.

Irwin, D. E. (1992b). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146–165). New York: Springer-Verlag.

Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.

Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics, 34,* 35–46.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

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

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35,* 1897–1916.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4,* 501–506.

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 & Psychophysics, 25,* 221–224.

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.

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., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition, 7,* 191–212.

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., Rensink, R. A., & Clark, J. J. (1999, March 4). Change-blindness as a result of "mudsplashes." *Nature, 398,* 34.

Parker, R. E. (1978). Picture processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 284–293.

Pollatsek, A., & Rayner, K. (1992). What is integrated across fixations? In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 166–191). New York: Springer-Verlag.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113,* 426–442.

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 199–210.

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting memories* (pp. 13–46). Cambridge, MA: MIT Press.

Rayner, K., McConkie, G. W., & Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 529–544.

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

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

Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision Research, 40,* 1469–1487.

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.

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

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38(A),* 475–491.

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

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1–16.

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

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

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

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

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1,* 275–277.

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

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

(*Appendix follows*)

Appendix

Scene and Target Object Stimuli

Scene	Original target object	Type-change target
Art gallery (exterior)	Trash container	Mailbox
Attic A	Crib	Crate
Attic B	Stool	Filing cabinet
Bar	Ashtray	Bowl of nuts
Bathroom A	Hair dryer	Tissue box
Bathroom B	Spray bottle	Shampoo container
Bedroom 1A	Book	Alarm clock
Bedroom 1B	Lamp	Flowers in vase
Bedroom 2 (child's)	Toy truck	Gumball machine
Computer workstation	Pen	Pencil
Dining room	Candelabra	Flowering plant
Family Room A	Watch	Coasters
Family Room B	Eyeglasses	Remote control
Family Room C	Briefcase	Wastebasket
Front yard	Watering can	Bucket
Indoor Pool A	Drinking glass	Soda can
Indoor Pool B	Deck chair	Side table
Kitchen 1A	Teapot	Pot
Kitchen 1B	Coffee maker	Blender
Kitchen 2A	Knife	Fork
Kitchen 2B	Toaster	Canister
Kitchen 3	Coffee cup	Apple
Laboratory A	Microscope	Flask
Laboratory B	Cell phone	Stapler
Laundry room	Iron	Aerosol can
Living Room 1A	Clock	Picture in frame
Living Room 1B	Magazine	Serving tray
Living Room 2	Television	Aquarium
Living Room 3	Chandelier	Ceiling fan
Loft	Pool table	Piano
Office A	Notebook	Computer disk
Office B	Telephone	Binder
Patio	Barbeque grill	Trash can
Restaurant	Flower in vase	Candle
Stage	Guitar	Audio speaker
Staircase	Chair	Fern

Note. A short description of each scene item is listed in the first column. Multiple examples of certain scene types were used. Some of these were created from different 3-D wire frame models and are differentiated by number; some were different views within the same model and are differentiated by letter. The second column lists the original target object in each scene. The third column lists the object substituted for the target in the type-change condition of Experiment 1. Changed targets in the token-change condition were different examples of the same type of object described in the second column.

Received July 24, 2000
Revision received May 10, 2001
Accepted May 11, 2001 ■