

Semantic Informativeness Mediates the Detection of Changes in Natural Scenes

Andrew Hollingworth and John M. Henderson

Department of Psychology, Michigan State University, East Lansing, USA

Three experiments investigated whether the semantic informativeness of a scene region (object) influences its representation between successive views. In Experiment 1, a scene and a modified version of that scene were presented in alternation, separated by a brief retention interval. A changed object was either semantically consistent with the scene (non-informative) or inconsistent (informative). Change detection latency was shorter in the semantically inconsistent versus consistent condition. In Experiment 2, eye movements were eliminated by presenting a single cycle of the change sequence. Detection accuracy was higher for inconsistent versus consistent objects. This inconsistent object advantage was obtained when the potential strategy of selectively encoding inconsistent objects was no longer advantageous (Experiment 3). These results indicate that the semantic properties of an object influence whether the representation of that object is maintained between views of a scene, and this influence is not caused solely by the differential allocation of eye fixations to the changing region. The potential cognitive mechanisms supporting this effect are discussed.

When viewing natural environments, human beings experience a visually detailed percept of the external world. A central issue to the understanding of dynamic visual perception is the extent to which this experience accurately reflects visual representation in the brain: Do internal representations maintain a veridical image of the external world? A number of studies have indicated

Please address all correspondence to A. Hollingworth, Department of Psychology, 129 Psychology Research Building, Michigan State University, East Lansing, MI 48824-1117, USA. Email: andrew@eyelab.msu.edu

This research was supported by a National Science Foundation Graduate Fellowship to A.H. and grants from the US Army Research Office (DAAH04-94-G-0404) and the National Science Foundation (SBR9617274) to J.H. The contents of this paper are those of the authors and should not be construed as an official Department of the Army position, policy, or decision. We would like to thank Dan Simons, Ron Rensink, and an anonymous reviewer for their helpful comments on the manuscript, Phil Weeks for his work in developing the change detection paradigm in Experiments 2 and 3, and Amy Siefert for help running the experimental sessions.

that very little visual information may actually be maintained from one view to the next (Currie, McConkie, Carlson-Radvansky, & Irwin, in press; Grimes, 1996; Henderson, 1997; Henderson & Hollingworth, 1999b; Irwin, 1991; Levin & Simons, 1997; McConkie & Currie, 1996; McConkie & Zola, 1979; Rensink, O'Regan, & Clark, 1997). These studies have generally employed change detection paradigms, with the logic that the types of changes detected across a brief retention interval reflect information that is maintained by the visual system. Participants have been largely insensitive to changes in the visual form of real-world objects (Henderson, 1997) and text (McConkie & Zola, 1979) when those changes are made during a saccade. In addition, participants have been quite poor at detecting a variety of changes to natural scenes when the change occurs during a saccade (Currie et al., in press; Grimes, 1996; Henderson & Hollingworth, 1999b; McConkie & Currie, 1996), between successive views of a scene separated by a brief retention interval (Rensink et al., 1997), or across different views in a film (Levin & Simons, 1997). This phenomenon of poor change detection performance has been termed "change blindness" (Rensink et al., 1997; see Simons & Levin, 1997, for a review).

If only a small portion of the information contained in a scene is maintained from one view to the next, what factors determine which information is and is not preserved? More specifically, what is the nature of the information that is selectively maintained between views of a scene, and what are the cognitive mechanisms that support this selectivity? In a change detection study using photographs of natural scenes, Rensink et al. (1997; see also Rensink, O'Regan, & Clark, this issue) sought to identify the stimulus properties that might influence the retention of scene regions between views. Verbal descriptions of scenes were used to rate local scene regions as "high interest" (those regions often listed in verbal descriptions), "medium interest," or "low interest" (those regions rarely listed). In a subsequent change detection experiment, a number of different changes could occur to one of these regions, including changes to an object's presence, location, or colour. Change detection performance was investigated using a "flicker" paradigm. One image and a modified version of that image were presented for 240 msec each in an alternating sequence (A, A, A', A' ...), with an 80-msec gray field appearing between images. The participant's task was to detect the change between the two images, and the sequence was repeated until the participant pressed a response button. After responding, the participant was required to accurately describe the location and nature of the change. Rensink et al. found that change detection was remarkably poor, with the detection of some changes requiring more than 45 sec of viewing. Importantly, not all changes were equally difficult to detect, with changes to regions of high interest detected most quickly and changes to regions of low interest detected least quickly. Rensink et al. (1997; Rensink, this issue) proposed that areas of higher interest are preferentially attended, and changes to attended areas are easier to detect because visual

information is more likely to be preserved across the retention interval for attended relative to unattended areas (see also Scholl, this issue).

The Rensink et al. (1997) study suggests that information from interesting regions of a scene is preferentially maintained across views compared to information from less interesting regions. However, "interestingness", as operationalized by the likelihood of verbal report, may depend not only on the semantic informativeness of a scene region but also on the visual informativeness of that region. For example, it is reasonable to think that a larger, centrally located object in a scene would be more likely to be listed in verbal reports than a smaller, peripherally located object. This raises the possibility that detection performance as a function of interest level may have been influenced by lower-level visual factors, such as the physical magnitude and eccentricity of the change. Because the Rensink et al. (1997) "interestingness" rating may have reflected both semantic and visual factors, it is not possible to determine from that study whether semantic informativeness alone can lead to better change detection. One motivation for the present study, then, was to determine specifically whether the semantic informativeness of a scene region influences whether the representation generated by that region is maintained across views of a scene.

A second issue with the Rensink et al. (1997) study is that eye movements were not controlled or monitored, so it is not clear whether differential change detection as a function of region interest was due to differences in the pattern of eye fixations or to differences in the allocation of covert attention. A number of studies of eye movements in scenes have indicated that more informative regions of a scene are fixated longer and more often than less informative regions (Antes, 1974; Friedman, 1979; Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978; Mackworth & Morandi, 1967). For example, Antes (1974) had participants rate regions of a scene for "informativeness": Specifically, the degree to which that region contributed to the overall information carried by the scene. In a subsequent experiment, eye movements were monitored while participants viewed the scenes to prepare for a memory test. Antes found that the eyes fixated highly informative regions sooner, longer, and more often compared to less informative regions. It is therefore likely that in the Rensink et al. (1997) study, participants' overt attention (i.e. eye fixations) was preferentially allocated to regions of higher interest compared to regions of lower interest. If the eyes were more likely to be fixating a high interest region when a change occurred in the scene, changes to high interest regions would be detected more quickly than changes to lower interest regions.

If this explanation of the Rensink et al. (1997) results is correct, then the phenomenon of differential change detection performance as a function of interest level may be caused simply by the fact that participants are better at detecting changes to regions they are looking at versus changes to regions they are not looking at, and thus would not extend our current understanding of the

influence of region informativeness on scene processing. If, however, differential change detection cannot be accounted for solely by overt attentional factors, then other cognitive mechanisms would need to be proposed to explain this effect. Thus, a second motivation for this study was to test whether eye movements are a necessary condition for the differential detection of changes as a function of region informativeness.

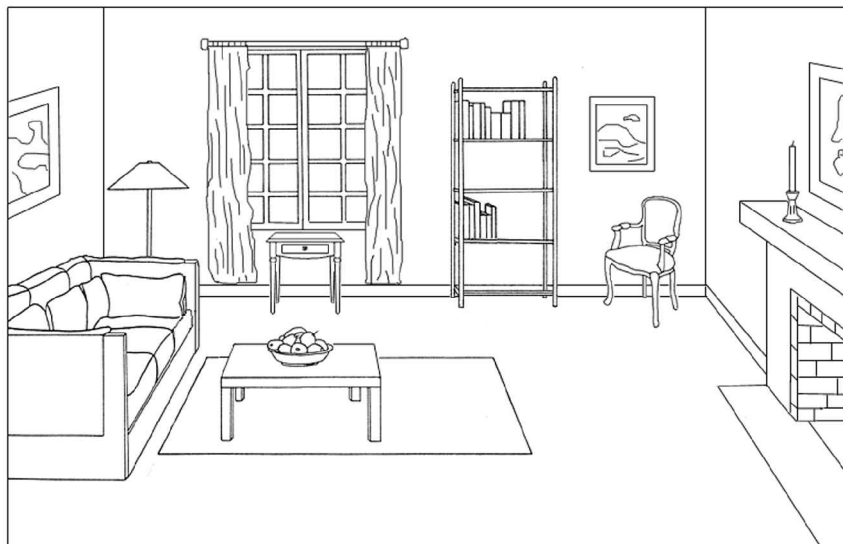
In the following experiments, the informativeness of a scene region was defined along a dimension that can be manipulated directly within natural scenes: The semantic consistency between an object and the scene in which it appears (Loftus & Mackworth, 1978). A semantically inconsistent object (e.g. a fire hydrant in a living room) is informative because it provides information that is not carried by other elements in the scene; a semantically consistent object (e.g. a chair in a living room) is less informative because it provides information that is redundant with other sources of information in the scene.¹ To address the potential confound between semantic and visual informativeness, we manipulated the semantic consistency of an object independently of its visual informativeness. Semantically consistent target objects were first chosen for each scene. The scene stimuli were then paired, and the semantically inconsistent target conditions were created by swapping objects across scenes. For example, a fire hydrant was the semantically consistent target in a street scene, and a chair was the consistent target in a living room scene. These targets were swapped across scenes so that the fire hydrant was the semantically inconsistent target in the living room scene, and the chair was the inconsistent target in the street scene. Thus, each target object appeared in both the semantically consistent and inconsistent conditions, controlling for visual factors such as object size and complexity. In addition, consistent and inconsistent target objects were placed in the same position in each scene, controlling for eccentricity. Figure 1 shows a sample scene and an example of the semantic consistency manipulation.

EXPERIMENT 1

In Experiment 1, the effect of semantic informativeness on the maintenance of the representation of a scene region between views of a scene was investigated

¹The term "informativeness" has been used differently by different researchers. The original studies of the influence of region informativeness on scene perception operationally defined informativeness as the extent to which a region contributed to the overall meaning of the scene (e.g. Antes, 1974). Loftus and Mackworth (1978), however, defined informativeness as the extent to which a scene region contained information that was inconsistent with other information in the scene. The precise relationship between these two types of informativeness has not been investigated. However, highly informative regions of a scene (established by either method) are fixated longer and more often than less informative regions (see Henderson & Hollingworth, 1998).

(a)



(b)

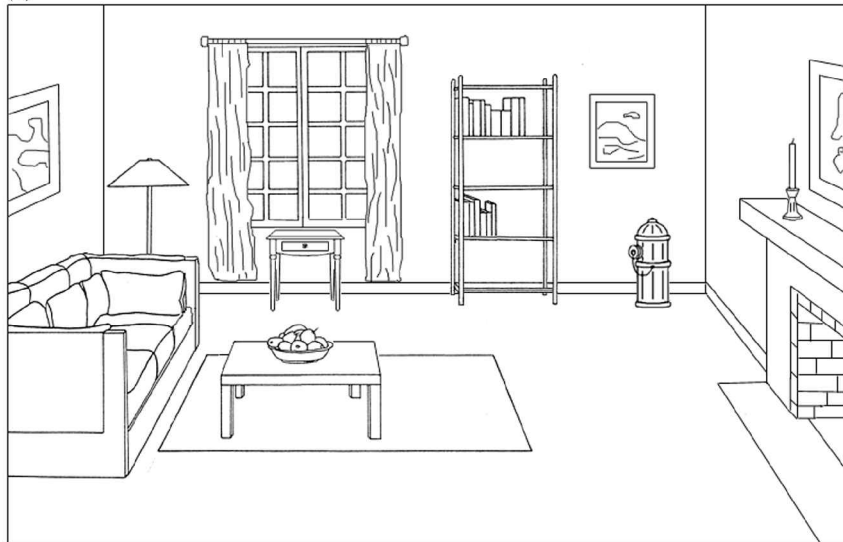


FIG. 1. An example of the type of scene used and the target object semantic consistency manipulation. (a) contains a semantically consistent target object (chair), whereas (b) contains a semantically inconsistent target object (fire hydrant). This living room scene was paired with a street scene in which the fire hydrant was consistent and the chair inconsistent.

using a version of the flicker paradigm (Rensink et al., 1997), illustrated in Fig. 2. Informative target objects were semantically inconsistent with the scene in which they appeared, and non-informative target objects were semantically consistent with the scene in which they appeared. Three change conditions were used. In the deletion–addition condition, the target object was present in Scene A but absent in Scene A'. Across repetitions of the flicker sequence, the target object disappeared, then reappeared, and so on, until response. The deletion–addition factor was included to investigate the maintenance of information about object presence. In the left–right orientation change condition, the target object was presented in one orientation in Scene A and in a mirror-reversed orientation in Scene A'. Across repetitions of this sequence, the object alternated between the two mirror images. The left–right orientation condition was included to investigate the maintenance of visual information

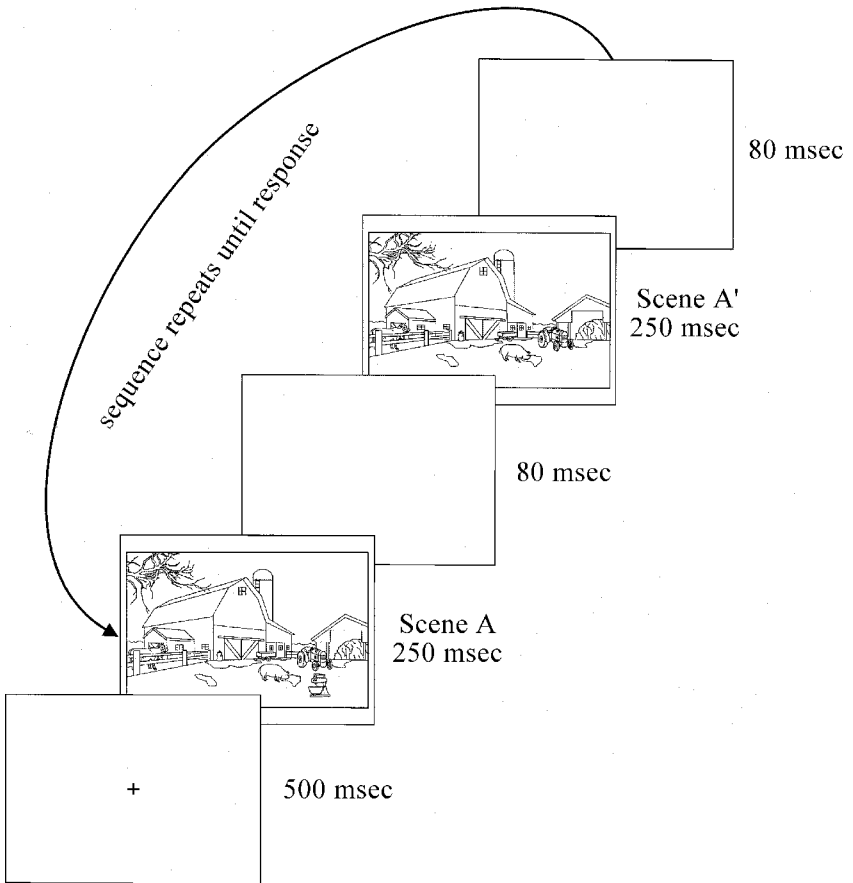


FIG. 2. Schematic illustration of a trial in Experiment 1.

independent of object identity. In the same condition, the two scenes were identical.

If the cognitive mechanisms supporting the selective maintenance of scene information from one view to the next are sensitive to the semantic properties of a scene region, changes to semantically inconsistent objects should be detected more rapidly than changes to semantically consistent objects. If, however, selectivity is driven by lower-level visual factors (such as object size and eccentricity), no difference in detection time should be observed as a function of the semantic consistency of the changing object.

Method

Participants. Twelve Michigan State University undergraduate students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision and were naïve with respect to the hypotheses under investigation.

Stimuli. Twenty scenes and twenty target objects were used as stimuli. The stimuli were generated from photographs of natural scenes. Fourteen scenes were modified from those used by van Diepen and De Graef (1994), and the other six scenes were generated from photographs taken in the East Lansing, Michigan area. In both cases, the main contours of the scenes were traced using Adobe Photoshop 4.0 to create gray-scale line drawings. The images generated from the two sources were not distinguishable. Semantically consistent target objects for each scene were also created by digitally tracing scanned images. The 20 scenes were paired, and the semantically inconsistent target conditions were created by swapping objects across scenes. Both target objects appeared in the same position in each scene, which did not coincide with the experimenter-determined initial fixation position.

One potential concern with swapping target objects between scenes is that the visual informativeness of the target may depend on the scene in which it appears. For example, if a target object were drawn together with the scene in which it was consistent, stylistic factors particular to the drawing of the scene could cause that object to be visually incongruous in the paired scene. In addition, if a natural object drawn in a setting containing other natural objects (e.g. a bird in a garden) were swapped with an artifact drawn in a setting containing other artifacts (e.g. a candlestick in a living room), the visual differences between natural objects and artifacts could make each target object more visually informative in the paired scene than in the original scene. To avoid these problems, the target objects in this experiment were created separately from the scenes in which they appeared. In addition, of the 20 target objects, 19 were artifacts, controlling (at least generally) for visual differences between natural

objects and artifacts. This ratio of artifacts to natural objects was not uncharacteristic of the general composition of the scene stimuli; most objects portrayed in the scenes were artifacts.

The scenes subtended a visual angle of 23 degrees (width) by 15 degrees (height) at a viewing distance of 64 cm. Target objects subtended about 2.75 degrees on average (range = 1.25 to 4.92 degrees). All images were displayed as gray-scale contours on a white background at a resolution of 800 by 600 pixels by 16 levels of gray. Gray-scale was used for anti-aliasing so that the contours appeared smooth and sharp.

Apparatus. The stimuli were displayed on a NEC XE15, SVGA monitor operating at 100 Hz. Participants used a button box to start each trial and to make their response. Stimulus presentation and response collection was controlled by a 486-66 microcomputer.

Procedure. Participants were tested individually. The experimenter explained that the task was to determine if a change was occurring between successive presentations of a scene. The experimenter also described the nature of the possible changes and explained that each type of change could occur for any of the objects in the scene. Participants were then seated in front of the computer monitor, with one hand resting on each button of the button box. Viewing distance was maintained by a forehead rest.

During each trial, participants saw a fixation cross and a prompt instructing them to press a pacing button to begin the trial. Once the participant pressed the button, the fixation cross remained on the screen for an additional 500 msec. Scene A was then presented for 250 msec, followed by a blank (white) screen for 80 msec, followed by Scene A' for 250 msec, followed by another blank screen for 80 msec. This sequence was repeated until the participant responded. The participant was instructed to press a button labelled "yes" as soon as a change was detected or a button labelled "no" if no change was evident.

A practice block of 16 trials was initially presented. The two scenes used in the practice block were not used in the experimental trials. After the practice session, the experimenter answered any questions the participant might have about the procedure before continuing to the experimental session. Participants then completed an experimental session of 160 trials produced by a within-participant factorial combination of 2 semantic consistency conditions \times 4 change conditions (deletion-addition, left-right orientation change, and two levels of same) \times 20 scenes. Two levels of the same condition were included to equate the number of trials on which a change did or did not occur, and were combined for the purpose of statistical analysis. The trials were presented in a random order generated independently for each participant. The entire session lasted approximately 40 minutes.

Results

Reaction Time Analysis. Trials with responses under 330 msec were eliminated, because the changed scene did not appear until 330 msec after the onset of the initial scene. Only one trial was eliminated across all participants. Mean reaction time as a function of semantic consistency and change condition is presented in Table 1. There was a main effect of semantic consistency, $F(1,11) = 7.47$, $MSe = 7169$, $p < .05$. Overall, responses to scenes containing an inconsistent target object were faster (1622 msec) than responses to scenes containing a consistent target object (1676 msec). There was also a reliable effect of change condition, $F(2,22) = 41.46$, $MSe = 169396$, $p < .001$, with mean response times of 2258 msec in the same condition, 1225 msec in the deletion–addition condition, and 1464 msec in the left–right orientation condition. Semantic consistency and change condition did not interact, $F(2,22) = 1.34$, $MSe = 4083$, $p > .25$.

A planned analysis of the deletion–addition and left–right orientation change conditions (i.e. excluding the same condition) was of particular interest. There was a main effect of semantic consistency, $F(1,11) = 11.36$, $MSe = 5471$, $p < .01$, with faster detection of changes to semantically inconsistent objects (1309 msec) than consistent objects (1380 msec). In addition, there was a main effect of change condition, $F(1,11) = 83.58$, $MSe = 8148$, $p < .001$; deletion–addition changes were detected more quickly than left–right orientation changes. These factors did not interact, ($F < 1$).

A final, simple effects test was conducted on the two levels of the same condition. There was no effect of semantic consistency, ($F < 1$).

Percentage Correct Analysis. Percentage correct performance was analysed to determine if differences in reaction time were mediated by differences in accuracy. The results are summarized in Table 1. First, there was no main effect of semantic consistency, ($F < 1$), suggesting that the differences in reaction time as a function of semantic consistency were independent of response accuracy. There was, however, a main effect of change condition, $F(2,22) = 17.18$, $MSe = .0063$, $p < .001$, with 98.0% performance in the same

TABLE 1
Mean reaction time (msec) as a function of target object semantic consistency and change condition, Experiment 1

Target Object Consistency	Change Condition		
	Same	Deletion–addition	Left–right Orientation Change
Consistent	2268 (98.3)	1261 (93.3)	1500 (85.4)
Inconsistent	2245 (97.7)	1190 (95.0)	1427 (84.6)

Percentage correct performance in parentheses.

condition, 94.2% performance in the deletion–addition condition, and 85.0% performance in the left–right orientation condition. Given that participants responded almost 1 sec slower in the same condition than in the deletion–addition and left–right orientation conditions, it is not surprising that percentage correct performance was higher for same responses. The difference between the deletion–addition and left–right orientation change conditions suggests that deletion–addition changes were more salient than left–right orientation changes, because the former were detected both more quickly and more accurately. Semantic consistency and change condition did not interact ($F < 1$).

An analysis of the deletion–addition and left–right orientation change conditions (i.e. excluding the same condition) revealed a similar pattern. There was no main effect of semantic consistency ($F < 1$), a reliable effect of change condition, $F(1,11) = 10.65$, $MSe = .0095$, $p < .01$, and no interaction between these factors, ($F < 1$). Finally, there was no effect of semantic consistency when comparing the two levels of the same condition ($F < 1$).

Discussion

The main result of Experiment 1 was that changes to semantically informative (inconsistent) objects were detected more rapidly than changes to less semantically informative (consistent) objects. The results of this experiment cannot be explained by differences in the visual informativeness of the changed objects, because each target object appeared in both the semantically consistent and inconsistent conditions, and target object eccentricity was controlled. These data demonstrate that the cognitive mechanisms supporting the selective maintenance of scene information from one view to the next are sensitive to the semantic properties of a scene region. In addition, the inconsistent object change detection advantage held both for deletion–addition and left–right orientation changes, suggesting that information about object presence and orientation are both preferentially maintained for semantically informative versus less informative objects. Finally, deletion–addition changes were detected more quickly and more accurately than left–right orientation changes. This result is not particularly surprising, however, because deletion–addition is a relatively large physical change, including changes both to the target object and to the background contours, which repeatedly become occluded and unoccluded.²

²Change detection latency was much shorter in this experiment than in Rensink et al. (1997). The difference may have arisen because our line drawings were less visually complex than the photographs used by Rensink et al., because the rate of change in Experiment 1 was approximately twice as fast as that in the Rensink et al. study, because the number of possible changes in Experiment 1 were relatively constrained compared to Rensink et al., and because our changes were always to discrete objects, whereas many of Rensink et al.'s changes were to object details or background elements.

One possible explanation for the difference in reaction time between consistent and inconsistent change conditions is that participants actively searched for the inconsistent objects in the scenes. This strategy is possible given that in the inconsistent target condition, only one object was inconsistent with the scene, and it changed on half the trials. In the consistent target condition, the target object was one of a number of consistent objects in the scene. Thus, participants may have learned that the status of the inconsistent object in the scene determined whether a change was or was not occurring, and so may have adopted a strategy of searching for inconsistent objects.

We have a number of reasons to believe, however, that participants did not adopt this strategy. First, if participants learned that the status of the inconsistent object determined the correct response, they should have been faster to respond that no change was occurring in the inconsistent target same condition versus the consistent target same condition. The data do not support this prediction. Although responses were slightly faster in the inconsistent target same condition than in the consistent target same condition, this difference did not approach reliability ($F < 1$). Second, if participants actively searched for inconsistent objects, not only should reaction times be shorter in the inconsistent change conditions but accuracy should also be higher. The data do not support this prediction, as there was no difference in accuracy as a function of target object consistency. Third, if participants adopted a strategy of searching for inconsistent objects, and if such a strategy accounted for the response time difference for consistent versus inconsistent object change detection, that difference should more prevalent later in the experiment (after participants had an opportunity to observe the change contingencies) than earlier in the experiment. To test this possibility, we calculated response time for change trials as a function of object consistency and first half versus second half of the trials. There was a main effect of first versus second half of the trials, $F(1,11) = 15.97$, $MSe = 27222$, $p < .005$; performance improved with practice from 1406 msec in the first half to 1272 msec in the second. Critically, the response time advantage for inconsistent versus consistent objects was actually larger in the first half of the experiment (122 msec) than the second half (38 msec), though the interaction between half of the trials and semantic consistency was not reliable, $F(1,11) = 1.24$, $MSe = 34311$, $p > .25$. A similar analysis of each quarter of the experimental trials indicated the same pattern, with the largest advantage for inconsistent objects appearing in the first quarter of the trials (143 msec).³ These data do not support the hypothesis that the development of an

³Given that each target object in each scene changed twice (once in the deletion condition and once in the left–right orientation condition), the improvement in performance with practice may have been caused by participants learning which objects changed in each scene and at what location. Such learning may have masked effects of semantic consistency, accounting for the smaller consistency effect later compared to earlier in the experiment.

inconsistent object search strategy caused the response time advantage for inconsistent object change detection. As a final test of the strategic encoding hypothesis, we modified the paradigm in Experiment 3 to assure that a strategy of selectively encoding inconsistent objects would be unprofitable.

A more plausible explanation for the results of Experiment 1 can be drawn from eye-movement studies of scene processing. As mentioned earlier, semantically informative objects are generally fixated longer and more often than less informative objects. Therefore, differences in detection performance could be explained by differences in the allocation of overt attention to an object as a function of its semantic consistency with the scene. If participants were more likely to be fixating a semantically inconsistent versus consistent object when a change occurred, changes to inconsistent objects should be detected faster than changes to consistent objects. Given the current evidence, differences in the allocation of overt attention as a function of scene region informativeness seem sufficient to explain differences in change detection latency in Experiment 1 and in Rensink et al. (1997).

EXPERIMENT 2

In Experiment 2, we sought to test the hypothesis that the inconsistent object change detection advantage in Experiment 1 was caused by the differential allocation of overt attention to the region of change. To do this, we limited the flicker paradigm to a single cycle of the change sequence. Scene A was presented once for 250 msec, eliminating the possibility that the target could be fixated prior to the change in Scene A'. If differences in change detection performance as a function of semantic informativeness are caused solely by the fact that inconsistent objects are more likely to be fixated when a change occurs in the scene, no difference should be found in the accuracy of change detection as a function of target object semantic consistency in this experiment.

An additional change from Experiment 1 to 2 was that we limited the change conditions to left–right orientation change and same, because the left–right orientation condition yielded the larger consistency effect in Experiment 1.

Method

Participants. Twenty-four Michigan State University undergraduate students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision, were naïve with respect to the hypotheses under investigation, and had not participated in Experiment 1.

Stimuli and Apparatus. The scene stimuli were the same as in Experiment 1. The pattern mask presented between the two scenes consisted of overlapping line segments, curves, and angles, and was slightly larger than the scene

stimuli. The scenes were completely obliterated when presented simultaneously with the pattern mask. The apparatus was the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, with the following modifications. Before the experiment, the experimenter described the nature of the possible change and explained that the task would be to determine if two scenes presented in succession were the same or different. Once the participant initiated a trial, a fixation cross was presented at the centre of the screen for 500 msec. Participants then saw a presentation of Scene A for 250 msec, followed by a pattern mask for 30 msec, followed by Scene A'. The retention interval was changed from 80 to 30 msec to reduce the potential influence of memory processes that might act differently on consistent versus inconsistent objects (e.g. Friedman, 1979; see General Discussion). Because the 30 msec retention interval was relatively brief, the pattern mask was presented during the retention interval to prevent the perception of apparent motion. Scene A' remained in view until the participant pressed the left button to indicate that the two scenes were identical or the right button to indicate that the scene had changed.

A practice block of eight trials was initially presented. The two scenes used in the practice block were not used in the experimental trials. Participants then completed an experimental session of 160 trials produced by presenting twice the within-participant factorial combination of 2 semantic consistency conditions \times 2 change conditions (left–right orientation change and same) \times 20 scenes. The 160 trials were distributed into 8 blocks to control for item presentation order. Each block contained all 20 scenes, and within each block 5 scenes were presented in each of the 4 conditions. Trial order was randomized independently for each participant. A group of eight participants generated a completely counterbalanced design. The entire session lasted approximately 40 minutes.

Results

Because each trial consisted of only two presentations of each scene, the dependent measure in this study was percentage correct detection of changes. Mean percentage correct performance as a function of semantic consistency and change condition is reported in Table 2. First, there was a main effect of block, $F(7,161) = 4.94$, $MSe = .0361$, $p < .001$, with better performance in later blocks than in earlier blocks. Block did not interact with semantic consistency ($F < 1.5$), but block produced a marginally reliable interaction with change condition, $F(7,161) = 1.94$, $MSe = .0432$, $p = .07$, caused by greater improvement across blocks in the left–right orientation change condition than in the same condition. Because block did not interact with semantic consistency, all subsequent analyses collapsed across the blocking factor.

TABLE 2
 Percentage correct performance as a function of target
 object semantic consistency and change condition,
 Experiment 2

Target Object Consistency	Change Condition	
	Same	Left-right Orientation Change
Consistent	78.2	49.5
Inconsistent	78.1	54.7

There was a reliable effect of semantic consistency, $F(1,23) = 6.07$, $MSe = .0026$, $p < .05$; performance was better in the semantically inconsistent target condition (66.4%) than in the consistent condition (63.9%). In addition, there was a reliable effect of change condition, $F(1,23) = 56.5$, $MSe = .0289$, $p < .001$, with better performance in the same condition (78.2%) than in the left-right orientation change condition (52.1%). Finally, there was a reliable interaction between semantic consistency and change condition, $F(1,23) = 5.04$, $MSe = .0034$, $p < .05$. This interaction was due to an advantage for inconsistent target objects in the left-right orientation change condition (5% difference) compared to no difference in the same condition.

To determine whether response biases were present as a function of target object consistency, we calculated B'' , a nonparametric measure of bias, using percentage correct in left-right orientation change trials as the hit rate and 1 minus percentage correct in the same trials as the false alarm rate (see Grier, 1971). This analysis indicated that there were no differences in bias between the two consistency conditions ($F < 1$). Thus, percentage correct in the left-right orientation change condition provided a valid measure of sensitivity to the scene changes. Orientation changes to inconsistent objects were detected more accurately (54.7%) than changes to consistent objects (49.5%), $F(1,23) = 10.70$, $MSe = .0030$, $p < .005$.⁴

Discussion

The main result from Experiment 2 was that detection of changes to semantically inconsistent objects was more accurate than detection of changes to

⁴The reaction time data exhibited the same pattern as percent correct performance. Correct change detection responses were made reliably more quickly when the target object was inconsistent (1321 ms) versus consistent (1468 msec) with the scene, $F(1,23) = 11.11$, $MSe = 23291$, $p < .005$.

semantically consistent objects, despite the fact that the initial scene was presented too briefly to allow fixation of the target object. These data suggest that differences in the fixation pattern on the changing region cannot explain entirely the inconsistent object change detection advantage. The preferential maintenance of information from semantically inconsistent objects between views of a scene must therefore depend, at least in part, on cognitive factors other than the differential allocation of overt attention. We will discuss potential explanations for this effect in the General Discussion.

One theoretically uninteresting explanation for the inconsistent object advantage observed in Experiments 1 and 2 is that participants were strategically encoding the inconsistent objects. As mentioned in the discussion of Experiment 1, participants might have developed such a strategy, either consciously or unconsciously, based on the contingencies of the changes. Given that an inconsistent object was present in a scene, it was the object that changed 50% of the time. Also, when an inconsistent object was present in a scene, a consistent object never changed. The purpose of Experiment 3 was to test this strategic encoding explanation of the inconsistent object advantage.

EXPERIMENT 3

Experiment 3 replicated the basic design of Experiment 2. However, two modifications were made to test the strategic encoding hypothesis. Both modifications reduced the probability that an inconsistent object presented in a scene was the object that changed left-right orientation. First, a condition was added in which an object other than the target object changed orientation between scenes A and A'. One consistent object was chosen in each scene to play this consistent distractor role. This object was changed once in the consistent target condition and once in the inconsistent target condition. Thus, there were four conditions in which an object changed: Two target change conditions (as in Experiment 2) and two distractor change conditions. (These conditions are illustrated in Fig. 3.) In only one of these conditions was an inconsistent object changed. Four same conditions (two in the consistent target condition and two in the inconsistent target condition) were included to equate the number of trials on which an object did or did not change. The second modification was that an inconsistent object (inconsistent bystander) was added to each scene in the consistent target condition so that every scene contained one inconsistent object (either an inconsistent target or an inconsistent bystander). The inconsistent bystanders never changed. As a result of these modifications, an inconsistent object was present in all eight conditions but was changed in only one condition (12.5% of the time compared to 50% of the time in Experiments 1 and 2). If participants were to develop a strategy based on the contingencies of the types of changes that occurred in the scenes, their best strategy would be to ignore the inconsistent objects and attend to a consistent object. If the

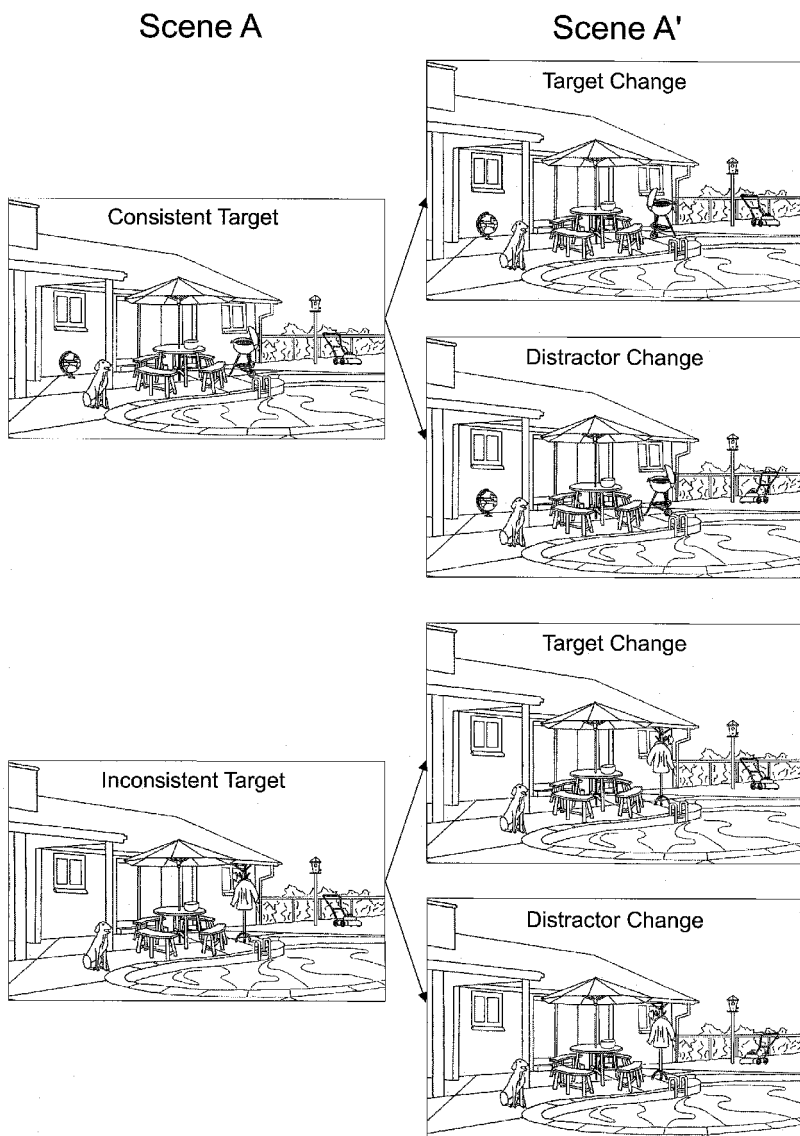


FIG. 3. Illustration of the target and distractor change conditions in Experiment 3 for the patio scene. In the target change conditions, either the consistent target (grill) or inconsistent target (coat rack) was changed between scene A and A'. In the distractor change conditions, a consistent distractor object (lawnmower) was changed both when the target was consistent and inconsistent with the scene. In addition, an inconsistent bystander object (globe) was added to the scenes in the consistent target condition, so that each scene contained one inconsistent object. Four same conditions (not pictured) were included to equate the number of trials on which a change did or did not occur.

inconsistent object advantage was due to strategic encoding, then it should be reduced or eliminated under these conditions.

Method

Participants. Twenty-four Michigan State University undergraduate students participated in the experiment for course credit. Three of the original 24 participants were replaced because their performance in the target change and same conditions was below 60%. All participants had normal or corrected-to-normal vision, were naïve with respect to the hypotheses under investigation, and had not participated in Experiment 1 or 2.

Stimuli and Apparatus. The stimuli were the same as in Experiments 2, except that one inconsistent bystander object was added to each scene in the consistent target condition. In addition, the target object position was changed in four scenes to accommodate the bystander object. An equivalent change in target position was made for each of these four scenes in the inconsistent target condition. The apparatus was the same as in Experiments 1 and 2.

Procedure. The procedure was the same as in Experiment 2. A practice block of 16 trials was initially presented. The two scenes used in the practice block were not used in the experimental trials. Each participant then completed an experimental session of 160 trials that were produced by a within-participant factorial combination of 2 target object semantic consistency conditions \times 4 change conditions (left–right orientation change to a target, left–right orientation change to a distractor, and two levels of same) \times 20 scenes. The blocking factor used in Experiment 2 was eliminated. The two levels of same in each target object consistency condition were combined for the purpose of statistical analysis. The trials were presented in a random order generated independently for each participant. The entire session lasted approximately 40 minutes.

Results

Percentage Correct Analysis. The two levels of the semantic consistency factor and the two theoretically relevant levels of the change factor (left–right orientation change to a target and same) were entered into an ANOVA. Mean percentage correct as a function of semantic consistency and change condition is reported in Table 3. There was a reliable effect of target object semantic consistency, $F(1,23) = 7.96$, $MSe = .0077$, $p < .01$; performance was better in the semantically inconsistent target condition (65.0%) than in the consistent condition (59.9%). In addition, there was a reliable effect of change condition, $F(1,23) = 134.09$, $MSe = .0342$, $p < .001$, with better performance in the same

TABLE 3
 Percentage correct and A' performance as a function
 of target object semantic consistency and change
 condition, Experiment 3

Target Object Consistency	Target Change Condition	
	Same	Left-right Orientation Change
Consistent		
Percent correct	83.6	36.3
A'		.678
Inconsistent		
Percent correct	85.0	45.0
A'		.755

condition (84.3%) than in the left-right orientation change condition (40.6%). Finally, there was a marginally reliable interaction between semantic consistency and change condition, $F(1,23) = 3.46$, $MSe = .0095$, $p = .07$. This interaction was due to a relatively small advantage for inconsistent targets in the same condition (1.4%) compared to a larger advantage for inconsistent targets in the left-right orientation change condition (8.7%). This latter difference was reliable, $F(1,23) = 6.64$, $MSe = .0138$, $p < .05$.

To determine whether response biases were present in the percentage correct data, we calculated B'' as a function of target object consistency. Participants were reliably more biased to respond that a change occurred in the scene when the target object was consistent with the scene ($B'' = .263$) than when the target object was inconsistent with the scene ($B'' = .364$), $F(1,23) = 10.64$, $MSe = .0115$, $p < .005$. To obtain an unbiased measure of change detection performance, we employed A' , a non-parametric measure of sensitivity that approximates performance in a forced-choice paradigm (see Grier, 1971). These data are reported in Table 3. Change detection performance was reliably higher for inconsistent target objects ($A' = .76$) than for consistent target objects ($A' = .68$), $F(1,23) = 10.33$, $MSe = .0069$, $p < .005$.⁵

Discussion

In Experiment 3, a semantically inconsistent object was added to the scenes that contained consistent targets, and a condition was included in which a

⁵The reaction time data exhibited the same pattern as percentage correct and A' . Correct detection responses for target changes were made faster when the target object was inconsistent (1422 msec) versus consistent (1535 msec) with the scene, though this difference was not reliable ($F < 1.2$).

consistent distractor changed rather than the target object. Thus, every scene contained an inconsistent object, but an inconsistent object was changed on only 12.5% of the trials, making a strategy of selectively encoding the inconsistent object sub-optimal. Nevertheless, the results of Experiment 3 replicated those of Experiments 1 and 2, with change detection clearly facilitated for the semantically inconsistent targets over the semantically consistent targets. These results do not support the hypothesis that the inconsistent object advantage is due to strategic encoding of inconsistent target objects.

In Experiment 3, the scenes containing the semantically consistent targets also contained one more object than did the scenes containing the semantically inconsistent targets, because the former scenes also contained an inconsistent bystander. Thus, it is possible that the poorer change detection performance for the semantically consistent target objects in Experiment 3 was due to the presence of the additional object (the inconsistent bystander) in these scenes. In order to test this possibility, we examined change detection for the consistent distractor objects. In the consistent target condition, consistent distractors changed when an inconsistent bystander was present in the scene, whereas in the inconsistent target condition, consistent distractors changed with no inconsistent bystander present. Note that in this comparison, the same consistent distractor objects are being compared given that an inconsistent bystander is and is not present. Percentage correct change detection was 33% and 29% for consistent distractors in the consistent and inconsistent target conditions ($F < 1$). This comparison shows that the presence of an extra object did not lead to a general decrement in change detection performance. Thus, these results support the conclusion that the inconsistent object advantage in Experiment 3 for the target objects was not caused by the addition of the inconsistent bystander objects in the scenes containing the consistent targets.

GENERAL DISCUSSION

The purpose of this study was to determine whether the semantic properties of a scene region influence the retention of that region from one view of a scene to the next. Rensink et al. (1997) used a flicker paradigm to show that “interesting” or informative regions may be preferentially maintained across views, but this study may have confounded the semantic informativeness of a scene region with its visual informativeness. In addition, eye movements were not controlled in Rensink et al. (1997), so it is not possible to determine from that study whether differential change detection as a function of region interest was due to differences in the pattern of eye fixations or to differences in the allocation of covert attention.

To address these issues, we manipulated the semantic informativeness of a scene region independently of its visual informativeness both in the flicker paradigm and in a paradigm that prevented fixation of the target object prior to the

change in the scene. The central finding of this study was that semantically informative (inconsistent) regions of a scene are preferentially represented between views of the scene compared to less informative (consistent) regions. In addition, this effect cannot be explained entirely by the differential allocation of overt attention (i.e. eye fixations), as the inconsistent object advantage remained when eye movements were eliminated from the paradigm. These results suggest that the cognitive mechanisms supporting the selective maintenance of scene information from one view to the next are sensitive to the semantic properties of a scene region (see also Werner & Thies, this issue). We now turn to the potential cognitive mechanisms that could support the preferential representation of semantically inconsistent information between views of a scene. We have identified three potential hypotheses. None of these hypotheses receives unambiguous support from extant data. However, we are presenting them as logical possibilities to guide further investigation.

One possible explanation for the inconsistent object advantage is the *memory schema hypothesis*. In this view, object representations that are semantically consistent with a memory schema for that type of scene are normalized to default values in the schema, whereas semantically inconsistent object representations are retained in a more veridical form. In Experiment 2, Scene A was presented for 250 msec, and Scene A' appeared only 30 msec later. Thus, a schema normalization process would need to act quite rapidly in order to produce the inconsistent object advantage. The memory schema hypothesis is supported by studies demonstrating better long-term memory for inconsistent versus consistent objects in a scene after an initial free-viewing session (Friedman, 1979; Pedzek, Whetstone, Reynolds, Askari, & Dougherty, 1989). For example, Friedman (1979) showed that changes to objects in scenes were better detected for objects that were inconsistent with the scene in the free-viewing session. However, there is a potential confound in these experiments that makes it difficult to take them as clear evidence for the memory schema hypothesis. Specifically, in free-viewing paradigms, semantically inconsistent objects are foveally fixated more often and for a longer duration than consistent objects (Friedman, 1979; Henderson et al., 1999; Loftus & Mackworth, 1978). Thus, better long-term memory for inconsistent objects in a scene could be due to increased foveal processing time and not to differences in memory processing *per se* (Biederman, Mezzanotte, & Rabinowitz, 1982; Henderson, 1992).

We have conducted two studies to investigate the memory schema hypothesis as an explanation for the inconsistent object advantage (Hollingworth & Henderson, submitted). Both experiments used the paradigm employed in Experiment 2. In the first, we manipulated whether the pattern mask appearing between scenes A and A' was presented for 30 or 400 msec. If the inconsistent object advantage derives from memory schema normalization, the advantage should become larger when the retention interval between the two scenes is

increased. The results did not support this prediction, as the magnitude of the inconsistent object advantage did not vary as a function of mask duration. In the second experiment, we presented either a pattern mask or a conceptual mask (i.e. a scene of a different conceptual type) for 250 msec between the presentation of scenes A and A'. Potter (1976) found that the conceptual processing of a scene was disrupted by the immediate presentation of a second scene. If the inconsistent object advantage is due to conceptual-level schema effects, then the presentation of a conceptual mask between scene A and A' should eliminate or attenuate that advantage. Contrary to this prediction the inconsistent object advantage was just as large in the conceptual mask condition as in the pattern mask condition.

A second potential explanation for the inconsistent object advantage is the *attentional attraction hypothesis*. According to this hypothesis covert attention is drawn to an object when there is difficulty reconciling that object's identity with the overall meaning of the scene. For example, covert attention may be drawn to a fire hydrant in a living room because the hydrant is odd or interesting in that scene context. The role of attention may be to make sure that a perceptual mistake wasn't made (is that really a fire hydrant?), or to check for further details that could help reconcile the conceptual discrepancy. The additional attentional resources may then produce a more complete or better remembered perceptual description, facilitating change detection performance compared to consistent objects.

Although many studies have demonstrated that covert visual attention can be dissociated from the current fixation position (e.g. Posner, Snyder, & Davidson, 1980; see Egeth & Yantis, 1997), covert attention appears to be tightly linked to eye movements under more natural viewing conditions (Deubel & Schneider, 1996; Henderson, Pollatsek, & Rayner, 1989; Shepard, Findlay, & Hockey, 1986). As a result, studies of eye movements in scenes can provide some indication of the allocation of covert attention during scene processing. The attentional attraction hypothesis is therefore supported by evidence from eye-movement studies that semantically inconsistent objects are fixated earlier in scene viewing than semantically consistent objects (Loftus & Mackworth, 1978), but this result has not been replicated (De Graef, Christiaens, & d' Ydewalle, 1990; Henderson et al., 1999).

Finally, the inconsistent object advantage could be explained by an *attentional disengagement hypothesis*. In this view, covert attention is initially deployed to scene regions based on stimulus factors such as discontinuity in colour, contrast, and depth (Antes, 1974; Henderson & Hollingworth, 1998, 1999a; Henderson et al., 1999; Mackworth & Morandi, 1967). Although covert attention is not initially drawn to regions of conceptual difficulty, once such a region has been attended, attention may be captured by conceptual processing difficulty. This hypothesis is consistent with results showing that the eyes are not initially drawn to semantically inconsistent objects in a scene (De Graef et

al., 1990; Henderson et al., 1999), but that once such a region has been fixated, the eyes tend to remain longer on these objects (De Graef et al., 1990; Friedman, 1979; Henderson et al., 1999; Loftus & Mackworth, 1978). Again, if attention is captured by semantic inconsistency, then a more complete representation of inconsistent objects would be generated and encoded into memory.

CONCLUSION

In summary, we reported evidence from three change-detection experiments showing that a change to an object is more easily detected when that object is semantically inconsistent with its scene than when it is semantically consistent. These results demonstrate that semantic properties of a scene region influence whether the representation of that region is or is not retained across views of the scene, and thus that the internal representation generated from a complex scene is not a veridical copy of that scene.

REFERENCES

- Antes, J.R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103*, 62–70.
- Biederman, I., Mezzanotte, R.J., & Rabinowitz, J.C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.
- Currie, C., McConkie, G.W., Carlson-Radvansky, L.A., & Irwin, D.E. (in press). The role of the saccade target object in the perception of a visually stable world. *Perception & Psychophysics*.
- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effect of scene context on object identification. *Psychological Research*, *52*, 317–329.
- Deubel, H., & Schneider, W.X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827–1837.
- Egeth, H.E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316–355.
- Grier, J.B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Vancouver studies in cognitive science: Vol. 5. Perception*. (pp. 89–110). New York: Oxford University Press.
- Henderson, J.M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology*, *46*, 319–341.
- Henderson, J.M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science*, *8*, 51–55.
- Henderson, J.M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–283). Oxford, UK: Elsevier.
- Henderson, J.M., & Hollingworth, A. (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., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception and Psychophysics*, *45*, 196–208.
- Henderson, J.M., Weeks, P.A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.
- Hollingworth, A., & Henderson, J.M. (submitted). Schema normalization fails to account for semantic informativeness effects on change detection in scenes.
- Irwin, D.E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*, 420–456.
- Levin, D.T., & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin and Review*, *4*, 501–506.
- Loftus, G.R., & Mackworth, N.H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 565–572.
- Mackworth, N.H., & Morandi, A.J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, *2*, 547–552.
- McConkie, G.W., & Currie, C.B. (1996). Visual stability while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 563–581.
- McConkie, G.W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, *25*, 221–224.
- Pedzek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 587–595.
- Posner, M.I., Snyder, C.R.R., & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174.
- Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509–522.
- Rensink, R.A. (this issue). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.
- 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.
- Rensink, R.A., O'Regan, J.K., & Clark, J.J. (this issue). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, *7*, 127–145.
- Scholl, B.J. (this issue). Attenuated change blindness for exogenously attended items in a flicker program. *Visual Cognition*, *7*, 377–396.
- Shepard, M., Findlay, J.M., & Hockey, R.J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology*, *38A*, 475–491.
- Simons, D.J., & Levin, D.T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*, 261–267.
- van Diepen, P.M.J., & De Graef, P. (1994). *Line-drawing library and software toolbox* (Psych. Rep. No. 165). Leuven, Belgium: University of Leuven, Laboratory of Experimental Psychology.
- Werner, S., & Thies, B. (this issue). Is “change blindness” attenuated by domain-specific expertise? An expert–novices comparison of change detection in football images. *Visual Cognition*, *7*, 163–173.