Effects of semantic consistency on eye movements during scene viewing

Article *in* Journal of Experimental Psychology Human Perception & Performance · February 1999

DOI: 10.1037/0096-1523.25.1.210

Andrew Hollingworth
University of Iowa

Journal of Experimental Psychology:
Human Perception and Performance
1999, Vol. 25, No. 1, 210–228

0096-1523/99/$3.00

The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing

John M. Henderson, Phillip A. Weeks, Jr., and Andrew Hollingworth
Michigan State University

Eye movements were recorded while participants viewed line-drawing pictures of natural scenes in preparation for a memory test (Experiment 1) or to find a target object (Experiment 2). Initial saccades in a scene were not controlled by semantic information in the visual periphery, although fixation densities and fixation durations were affected by semantic consistency. The results are compared with earlier eye-tracking studies, and a qualitative model of eye movement control in scene perception is discussed in which initial saccades in a scene are controlled by visual but not semantic analysis.

Real-world scene viewing is an active process: The viewer selects the region of the scene that will be processed most completely at any given time via saccadic eye movements. In the first direct study of eye movement patterns during scene perception, Buswell (1935) recorded the eye movements of viewers while they examined pictures of artwork, including pictures of complex scenes, buildings, and sculpture. Buswell found that eye movement patterns were highly regular and related to the information in the pictures. For example, viewers tended to concentrate their fixations on the people rather than on background regions when examining *Sunday Afternoon on the Island of La Grande Jatte* by Georges Seurat. These data thus provided some of the earliest evidence that eye movement patterns during complex scene perception are related to perceptual and cognitive processing. Yarbus (1967) replicated these effects and attempted to capture the systematicity in the eye movement record with the suggestion that the eyes tend to land on scene regions that are either in reality or in the viewer's opinion "useful or essential for perception."

The influence of the properties of local scene regions on eye movement patterns has also been demonstrated using more analytical methods. In an early study, Mackworth and Morandi (1967) divided each of two color photographs into 64 square regions. A group of participants then rated the informativeness of each region based on how easy the region would be to recognize on another occasion. A second group of participants viewed the two pictures with the task of determining which one they preferred. Mackworth and Morandi found that the density of fixations (i.e., the number of discrete fixations made) in each of the 64 regions was related to the rated informativeness of the region, with regions rated more informative receiving more fixations than those rated less informative. In addition, regions rated low in informativeness often received no fixations at all, suggesting that the scenes were filtered by peripheral vision and that uninformative regions could be rejected as potential fixation sites based on peripheral information alone. In addition, Mackworth and Morandi found that viewers were as likely to place their fixations on informative regions in the scene during the first 2 s of viewing as in other 2-s intervals, providing evidence for early visual analysis of local scene regions.

In a similar study, Antes (1974) also provided evidence for an effect of early analysis of local scene regions on eye movement behavior. Antes presented achromatic pictures (taken mostly from the Thematic Apperception Test) to participants who viewed them in a picture-preference task. Informativeness was defined as the degree to which a region contributed to the total amount of information conveyed by the whole picture, as determined by the ratings of a separate group of viewers. A viewer's first saccade within the scene tended to be to an informative region. Like Mackworth and Morandi (1967), Antes (1974) concluded that informative regions in a scene could be found using information acquired from the visual periphery and that the eyes could then be directed to those regions early in scene exploration.

In both the Mackworth and Morandi (1967) and Antes (1974) studies, region "informativeness" was based on observer ratings and so could have reflected either relatively low-level visual information (e.g., the presence of contours and other discontinuities in color, texture, or luminance) or semantic information. Particularly interesting is the idea that the eyes may be controlled early during scene viewing by the relationship between global scene meaning derived from an initial fixation (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982; Schyns & Oliva, 1994; see Henderson, 1992a, for a

John M. Henderson, Phillip A. Weeks, Jr., and Andrew Hollingworth, Department of Psychology, Michigan State University. Phillip A. Weeks, Jr., is now at AT&T Bell Laboratories.

This research was supported by grants from the U.S. Army Research Office (DAAH04-94-G-0404) and the National Science Foundation (SBR 9617274), and by a National Science Foundation graduate fellowship to Andrew Hollingworth. The contents of this article 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 Fernanda Ferreira for many insightful discussions of these issues and Keith Rayner and David Irwin for their helpful reviews.

Correspondence concerning this article should be addressed to John M. Henderson, Department of Psychology, Michigan State University, East Lansing, Michigan 48824-1117. Electronic mail may be sent to john@eyelab.msu.edu.

review) and the semantic properties of scene regions viewed in the visual periphery.

Semantic Information and Eye Movements

In the first study designed to explore directly the influence of semantic informativeness on eye movements during scene perception, Loftus and Mackworth (1978) presented viewers with line drawings of scenes in which a target object was either semantically inconsistent (informative) or consistent (uninformative) with the global gist of the scene. For example, both a farm scene and an underwater scene could contain either a tractor or an octopus. Participants began each trial by fixating a location 30° below the area that would contain the scenes. A scene was then presented. Participants moved their eyes to the scene and then freely explored it in preparation for a later recognition test. Loftus and Mackworth found that viewers tended to fixate the inconsistent objects earlier than the consistent objects over the course of scene viewing. Most strikingly, viewers were more likely to move their eyes to a semantically inconsistent than a consistent object immediately after the first fixation on the scene. Loftus and Mackworth argued that these data could not be explained by the hypothesis that the objects were being identified in near-foveal vision because the distance of the saccade to the target objects averaged over 7° of visual angle. Thus, these data suggested that viewers could determine in a single fixation the semantic consistency of an object with its scene based on peripheral visual information and that semantic consistency could then exert an immediate effect on eye movement control.

De Graef, Christiaens, and d'Ydewalle (1990) investigated the influence of object semantics on eye movement patterns during scene viewing using a visual search task rather than a memorization task. In this task, viewers searched line drawings of scenes for the presence of objectlike figures that were not associated with any identifiable real-world object ("nonobjects"). The data of interest derived from target objects that were embedded in the scenes. These target objects were either semantically consistent or inconsistent with the scene. (Other types of inconsistencies were manipulated as well, but we focus on semantic consistency here.) In contrast to the data reported by Loftus and Mackworth (1978), De Graef et al. (1990) found no evidence that semantically inconsistent objects were fixated earlier than consistent objects during scene viewing.

One potential explanation for the difference between the results of Loftus and Mackworth (1978) and De Graef et al. (1990) is the nature of the task used in the two studies. Although participants in Loftus and Mackworth's study attempted to memorize the scenes, participants in the De Graef et al. study searched for nonobjects. It could be that scene semantics, and thus the relationship between scene semantics and the semantics of local scene regions, were analyzed by participants in Loftus and Mackworth's study but not in the De Graef et al. study. If scene semantics were not typically analyzed by participants in the latter study, then no effect of semantics would be expected. Another potential explanation for the difference in results rests on differences in the nature of the scene stimuli used in the two studies. Loftus and Mackworth used line drawings that were simple and visually sparse, consisting only of a small number of discrete objects and a great deal of empty space. In contrast, De Graef et al. used line drawings that were derived from photographs of natural scenes and so were relatively complex and realistic. It could be that the eyes are not controlled by the relationship between visually peripheral scene regions and overall scene semantics when scenes are visually complex and realistic regardless of viewing instructions.

The Present Research

The present research was designed to investigate the influence of semantic factors on eye movement patterns during the free viewing of complex, natural scenes. Our main focus was on the extent to which initial fixation placements in a scene would be determined by the semantic characteristics of peripheral scene regions. This issue is critical for understanding the role of cognitive factors in eye movement control during scene viewing.

We were also interested in two subsidiary questions. First, although there is good evidence that visual informativeness affects the density of fixations in a given scene region during scene viewing (Antes, 1974; Buswell, 1935; Mackworth & Morandi, 1967; Yarbus, 1967; see Henderson & Hollingworth, 1998, for a review), the role of the semantic informativeness of a scene region on fixation density is less clear. Loftus and Mackworth (1978) found greater fixation densities for semantically inconsistent than consistent objects in scenes. However, Friedman (1979) did not replicate this effect. In the latter study, participants viewed line drawings of real-world scenes in preparation for a memory test. Each scene contained objects that had been rated for their likelihood within the scene by an independent group of viewers. Friedman (1979; Friedman & Liebelt, 1981) found no effect of semantic likelihood on fixation density. The influence of scene semantics on the probability and number of region refixations is important because it speaks to the degree to which the placement of individual fixations in a scene is guided by the semantic informativeness of a region once that region has already been fixated. It may be that, although initial fixation placement is determined by visual factors, later fixation placement can be controlled by the meaning of a previously fixated region. Alternatively, fixation placement could be controlled by purely visual factors throughout the duration of scene viewing.

The second subsidiary issue we wanted to investigate was the extent to which the semantics of a scene region would be reflected by different measures of eye movement behavior. Investigators who have examined the issue have found that the amount of time viewers direct their fixations at a scene region is influenced by the semantics of that region (De Graef et al., 1990; Friedman, 1979; Loftus & Mackworth, 1978). However, there are other eye movement measures of region processing that have not been examined. In the present research, we follow the lead of investigators in the recent reading literature and report a relatively large number of measures of eye movement behavior (Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). The use of these additional measures provides the basis for determining which aspects of eye movement behavior reflect semantic processing and so provides a database from which to construct a model of eye movement control during complex cognitive tasks. The use of multiple measures also facilitates the comparison of eye movement behavior in scene processing with the study of eye movement behavior in reading, where relatively sophisticated eye movement control models exist (Legge, Klitz, & Tjan, 1997; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reilly & O'Regan, 1998). Ultimately, the use of multiple measures allows one to gain a more complete understanding of the role of semantic consistency on eye movement control during complex scene viewing.

To investigate these issues, we manipulated the semantic consistency of a particular region of a complex, natural scene by changing a specific target object in that region (De Graef et al., 1990; Loftus & Mackworth, 1978). This manipulation allowed us to examine the effect of semantic consistency while controlling other factors such as the overall meaning and visual complexity of the regions. In each stimulus scene, we chose a single target object whose presence was relatively predictable given the meaning of the scene (the consistent context condition). We then exchanged target objects across pairs of scenes to create scenes in which the target object was highly unexpected (the inconsistent context condition). In Experiment 1, participants viewed the scenes in preparation for a memory test. In Experiment 2, participants searched for prespecified target objects. In both experiments, a detailed record of viewers' eye movements was recorded.

Experiment 1

Method

Participants

Eighteen Michigan State University undergraduates took part in this experiment. All participants had normal or corrected-to-normal vision, were naive with respect to the purposes of the research, and received partial credit for their introductory psychology courses or remuneration for their participation.

Stimuli

Twenty-four scenes modified from those constructed by De Graef et al. (1990; see van Diepen & De Graef, 1994) were used as stimuli. These scenes were originally created by photographing 24 real-world scenes in the environment, projecting the photographs onto a screen, and tracing the primary contours of the images to produce line drawings (De Graef et al., 1990; van Deipen & De Graef, 1994). The line drawings were then digitized using a commercial scanner. Although the line drawings did not preserve all of the contours of the original photographs, we have demonstrated that eye movement patterns while viewing these line drawings are very similar to those while viewing full-color photographs of real-world scenes (Henderson & Hollingworth, 1998). Thus, these stimuli appear to provide a reasonable approximation to the level of complexity found in natural scenes.

Contextually consistent target objects were drawn independently for each of these scenes and were digitized in the same manner as the scenes. Scenes were then paired so that the consistent object from a given scene, when placed in its paired scene, was contextually inconsistent. An example of the contextual manipulation is shown in Figure 1. In this barroom scene, a cocktail appeared as the consistent target object and a microscope appeared as the inconsistent target object. The barroom scene was paired with a laboratory scene, in which the microscope was consistent and the cocktail inconsistent. Contextually consistent and inconsistent target objects for the paired scenes were matched for real-world size and were placed in the same position within a scene. The scenes subtended a visual angle of 10° (height) × 14.5° (width). Target objects subtended about 2.1° on average, measured along their longest axis.

To verify that the target objects were sufficiently constrained in the consistent context condition and were sufficiently unexpected in the inconsistent context condition, we had a separate group of 16 participants drawn from the same pool as those who took part in the main experiment rate the scenes. Each of these participants saw each of the 24 scenes twice, once with the consistent target object and once with the inconsistent target object. The order of scene presentation was randomly determined for each participant. The task was to determine whether all of the objects fit in the scene. Participants were instructed to press the "yes" button on a response panel if they believed that all of the objects in the present scene "fit

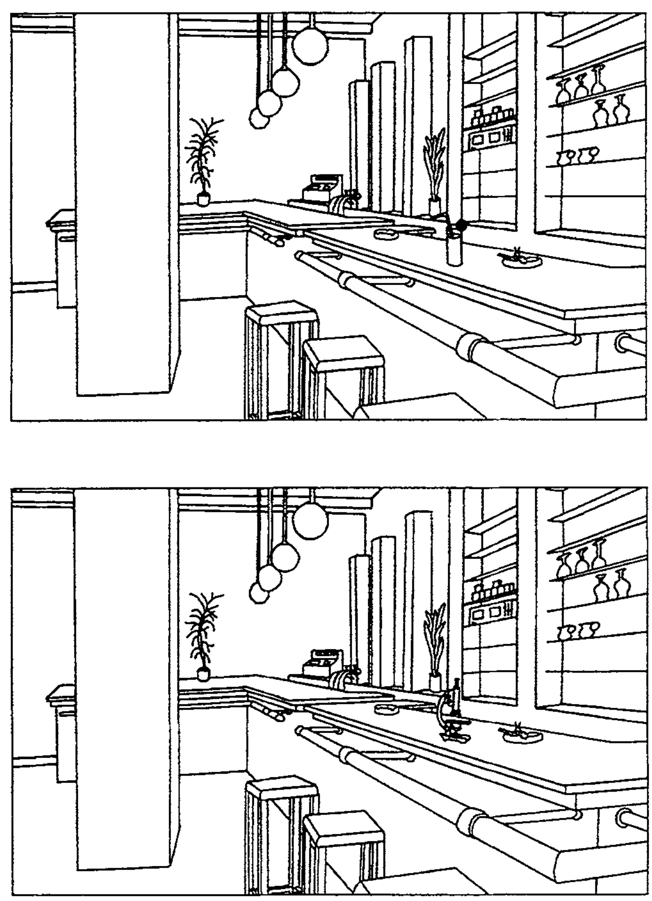

Figure 1. An example of the types of scenes used in Experiments 1 and 2. The top panel shows a consistent target object (cocktail), and the bottom panel an inconsistent target object (microscope), in a barroom scene.

the scene" and to press the "no" button if they believed that one or more of the objects "did not fit the scene for any reason." Participants were free to examine the scenes at their own pace. The entire session lasted approximately 15 min. Participants responded yes 89.3% of the time for the scenes containing consistent targets and and no 89.5% of the time for the scenes containing the inconsistent targets. An analysis of variance (ANOVA) conducted on the percentage of yes responses in the two conditions showed that the effect of the manipulation was reliable, $F(1, 15) = 779.18$, $MSE = 0.01, p < .001$. These data indicate that scene context was manipulated adequately in the construction of the scenes used in the main experiment.

Procedure

Participants first read a description of the experiment along with a set of instructions. The instructions indicated that the experiment involved scene memory: The participants were told that their eye movements would be monitored while they looked at scenes that they would later be asked to recognize. They were also told that during the recognition test, they would be required to distinguish between the original scenes and new scenes in which, for example, only a small detail of a particular object may have been changed (Friedman, 1979). The memory test was never actually given. Participants were also informed that they would be shown each scene for 15 s.

After the instructions were reviewed, the eye-tracking apparatus was calibrated. Calibration consisted of having the participant fixate four calibration markers at the top, bottom, left, and right sides of the display area. Calibration was checked by displaying a calibration screen consisting of six test positions and a fixation marker that indicated the computer's estimate of the current fixation position. The participant fixated the test positions, and if the fixation marker was within ±5 min arc of each, calibration was considered accurate.

Once calibrated, the participant took part in two practice trials, one in each of the context conditions, using scenes that were not included in the experimental stimuli. After the practice trials, any questions the participant had were answered. The eyetracker was then recalibrated following the above procedure, and the participant was shown 24 experimental trials, 12 trials in each of the two context conditions.

A trial consisted of the following events. First, the calibration screen was shown and calibration was checked. The eyetracker was recalibrated whenever calibration was deemed inaccurate. After the calibration check, the participant fixated the center marker to indicate that he or she was ready for the trial to begin. The experimenter then started the trial. The fixation display was replaced by the trial scene. The scene remained visible for 15 s, and was then replaced by the calibration screen.

A given participant saw all 24 scenes, 12 in which the target object was consistent with the scene and 12 in which it was not. Across participants, each scene appeared in each context condition an equal number of times. The order of scene presentation (and hence condition presentation) was determined randomly for each participant. The entire experiment lasted approximately 45 min.

Apparatus

The stimuli were displayed at a resolution of 800 × 600 pixels on NEC Multisync XE 15 in monitor driven by a Hercules Dynamite Pro super video graphics adapter card. The screen refresh rate was 100 Hz. The contours of the scenes appeared black (pixels off) against a gray (pixels on) background (the gray was created by setting the red, green, and blue channels to an intensity value of 16, where white is an intensity value of 64 on each channel). Display luminance was generally low and was individually adjusted to a comfortable level for each participant. The room was otherwise illuminated by a low-intensity, indirect light source.

Eye movements were monitored using a Generation 5.5 Stanford Research Institute dual Purkinje image eyetracker (Crane, 1994; Crane & Steele, 1985), which has a resolution of 1 min arc and a linear output over the range of the visual display used. A bite bar and forehead rest were used to maintain the participant's viewing position and distance. The position of the right eye was tracked, although viewing was binocular. Signals were sampled from the eyetracker using the polling mode of the Data Translations DT2802 analog-to-digital converter, producing a sampling rate of better than 1,000 Hz.

The eyetracker and display monitor were interfaced with a microcomputer running a 90-MHz Pentium processor. The computer controlled the experiment and maintained a complete record of time and eye position values over the course of each trial.

Results

Eye Movement Data Analysis

Raw 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 arc) in 15 ms or less. Manual inspection of the raw data files confirmed that this criterion effectively eliminated saccades while preserving slow drifts. Once saccades had been identified, fixation positions and durations were computed over the remaining data. Fixation positions and durations were initially computed independently of the positions of the target objects. The duration of a fixation was the elapsed time between two consecutive saccades. During a fixation, the eyes often drift. The scored position for a given fixation was taken to be the mean of the position samples (in pixel values) taken during that fixation weighted by the durations of each of those position samples (see Henderson, McClure, Pierce, & Schrock, 1997).

Scoring regions for each target object were defined by constructing a rectangular box around the target object that was large enough to encompass both the consistent and inconsistent object for a given scene. The pixel coordinates of the box were then taken as the position of the target. The same box was used for both the consistent and inconsistent context conditions for a given object, so that the size of the scoring regions was equated across context conditions. Finally, each fixation in the scene was determined to be within or outside of the box based on its position value. Individual fixations less than 90 ms, or greater than 1,000 ms, were eliminated as outliers. All data reduction and analysis were conducted using automated analysis software.

Typical viewing patterns are shown in Figures 2A and 2B for 2 participants looking at the barroom scene containing a consistent and an inconsistent target object, respectively. The straight lines represent saccades. The fixations, represented as dots, are numbered in order. Consistent with the results of previous studies, viewers generally distributed their fixations across a large part of each scene in this experiment, with the majority of fixations landing on or near

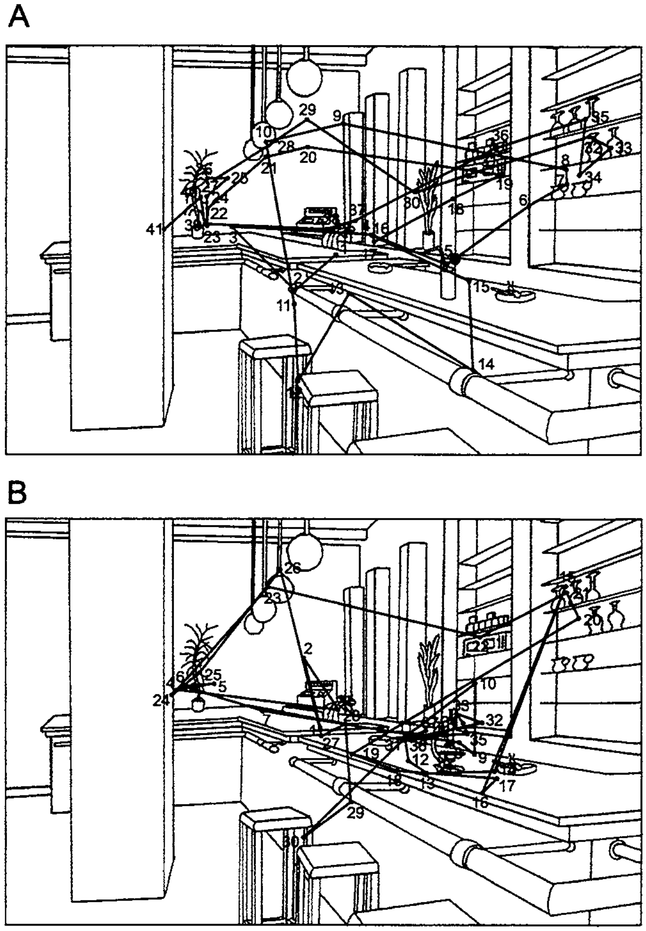

Figure 2. Sample viewing patterns for one participant on a barroom scene containing the consistent target object (A) and the inconsistent target object (B) in Experiment 1. The straight lines represent saccades, and the fixations, represented as dots, are numbered in order.

objects (Antes, 1974; Buswell, 1935; Mackworth & Morandi, 1967; Yarbus, 1967; see Henderson & Hollingworth, 1998).

Several measures were calculated to quantify viewers' eye movement patterns on the target objects as a function of contextual constraint. These measures are reported below. Many of these measures have been shown to reflect visual and cognitive processes in visual object perception (Henderson et al., 1997) and in reading (e.g., Rayner & Pollatsek, 1989). For each of these measures, two ANOVAs were conducted. One analysis treated participant as a random effect and contextual constraint as a within-subjects factor (reported as $F1$), whereas the other treated scene as a random effect and contextual constraint as a within-items factor (reported as $F2$). One item was removed from all analyses because the target object was inadvertently placed at the initial fixation point. In addition, trials on which the target was fixated on the first (i.e., experimenter-induced) fixation were excluded from the analysis, although this occurred only twice across all trials, accounting for 0.5% of the data.

Extrafoveal Semantic Analysis During Scene Perception

The primary issue we examined was the degree to which eye movements during scene viewing would be initially controlled by the semantic characteristics of peripheral scene regions. As discussed in the introduction, Loftus and Mackworth (1978) provided evidence that viewers were more likely to move their eyes to semantically inconsistent than consistent objects early during scene viewing, with a higher probability that they would move their eyes immediately to a semantically inconsistent object. In contrast, De Graef et al. (1990) failed to obtain this effect. In addition, Loftus and Mackworth found that viewers executed saccades that were on average 7.40° to informative objects and 7.25° to noninformative objects. These values are surprising in comparison to the majority of the picture viewing literature, in which mean saccadic amplitudes of 3°–4° are more typical and saccadic amplitudes of 7° are rare (Henderson & Hollingworth, 1998).

In the present experiment, we examined three measures of eye movement behavior to address the degree to which semantic informativeness could control early fixation patterns in complex, natural scenes. These measures were the probability of immediate target fixation, the number of fixations to the target, and the amplitude of the initial saccade to the target. Each of these measures was conditionalized on fixation of the target object, so nonfixations did not contribute to the computed means. The data from these measures are summarized in Table 1.

Probability of immediate target fixation. The probability of immediate target fixation was defined as the proportion of trials in which the target object was fixated after the first saccade in the scene. Overall, the mean probability of

Table 1
Measures of Extrafoveal Semantic Analysis (Means and Standard Errors by Participants) in Experiment 1

	Scene context condition			
Measure	Consistent	Inconsistent	*SEM*	Difference
Probability of immediate target fixation	.12	.09	.028	.03
No. of fixations to target	9.7	10.7	0.416	−1.0
Amplitude of initial saccade to target (deg)	3.21	2.86	0.117	0.35*

*$p < .05$ by participants.

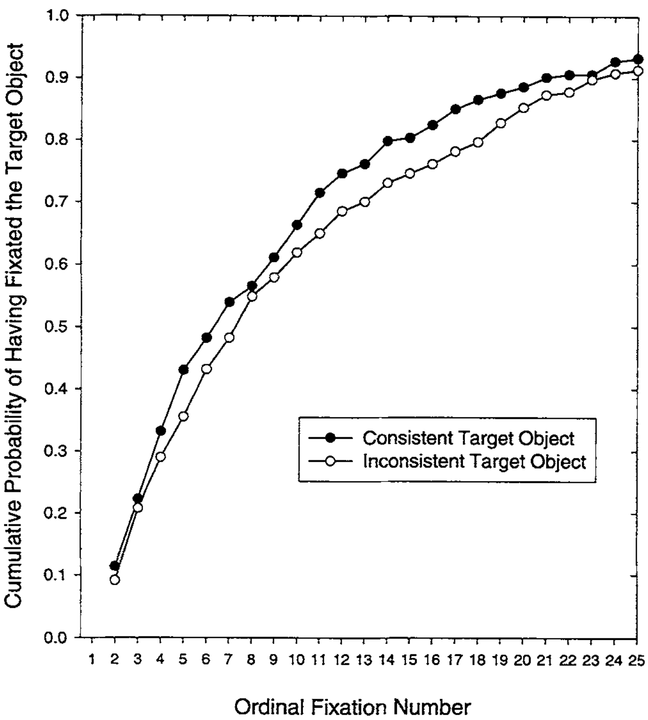

Figure 3. Cumulative probability of having fixated the target object as a function of the ordinal fixation number and semantic consistency in Experiment 1.

fixating the target after the first saccade was .11 and was unaffected by semantic consistency. The probability of immediately fixating the target object was .12 in the consistent condition and .09 in the inconsistent condition ($F1 < 1, MSE = .0141; F2 < 1, MSE = .0175$).

A similar pattern was observed for the cumulative probability of having fixated the target after the first two saccades in the scene. Overall, the mean probability of having fixated the target after two saccades was .22 and was unaffected by semantic consistency. The probability was .23 in the consistent condition and .21 in the inconsistent condition ($F1 < 1$, $MSE = .0179; F2 < 1, MSE = .0147$).

The cumulative probability of having fixated the target object as a function of ordinal fixation number and target object consistency is shown in Figure 3.[1] Note that fixations of the target occurring on Fixation 1 were excluded because that fixation was not under participant control. If any trend is present in this figure, it is that consistent target objects were fixated sooner than inconsistent target objects. De Graef et al. (1990, Figure 2) found a similar pattern in their nonobject search task, with a trend toward earlier fixation of consistent target objects (base condition) than inconsistent target objects (probability violation condition).

Number of fixations to the target. This measure was defined as the number of discrete fixations on the scene before the first fixation in the target object region. This value included both the first (i.e., experimenter-induced) fixation on the scene and the initial fixation in the target object region. Overall, the first fixation on the target was 10.2 fixations into scene viewing. There were 9.7 fixations to the target in the consistent condition and 10.7 fixations in the inconsistent condition, a difference that was not reliable, $F1(1, 17) = 3.070, MSE = 3.120, p = .09; F2 < 1, MSE = 12.66$. Given that each time the eyes landed in a new region, they made 1.7 fixations in that region on average (see the gaze fixation count measure for the consistent condition below), these data suggest that on average the eyes fixated about 4 ($10.2/1.7 - 2$) scene regions between the initial fixation on the scene (at the initial fixation point) and the first entry into the target object region.

Amplitude of initial saccade to the target. The amplitude of the initial saccade to the target was defined as the length of the saccade that first brought the eyes into the target object region. Overall, the amplitude of the initial saccade to the target was 3.03°. There was a trend toward a 0.35° effect of semantic consistency, which was reliable by subjects but not by items, $F1(1, 17) = 4.519, MSE = 647$, $p < .05; F2(1, 22) = 2.456, MSE = 1{,}712, p > .10$,[2] with an initial saccade of 3.21° to the consistent object and 2.86° to the inconsistent object. Interestingly, this marginal difference in saccadic amplitude across semantic consistency observed in the present study was in the direction opposite to that observed by Loftus and Mackworth (1978). Furthermore, although Loftus and Mackworth found that the average saccade length to a target object in their study was about 7.32°, the amplitude of an average saccade during scene viewing has typically been found to be about 3°–4° (Antes, 1974; Saida & Ikeda, 1979; Shiori & Ikeda, 1989; Van Diepen, Wampers, & d'Ydewalle, 1998; see Henderson & Hollingworth, 1998), values that are consistent with our observation of an average saccadic amplitude of 3.03°. Together, then, the bulk of the evidence suggests that during scene viewing, the semantic analysis of an object in a scene takes place within a relatively limited region within and near to the fovea and that the eyes are not drawn to semantically informative (inconsistent) peripheral scene regions.

Fixation Density

The second question we investigated was whether viewers would differentially distribute their fixations over objects in scenes as a function of semantic consistency. Friedman (1979) suggested that there were three important differences between her study and Loftus and Mackworth's (1978) study that might have accounted for the difference in results relating semantic consistency and fixation density. First, Friedman (1979) suggested that the scenes used by Loftus and Mackworth (1978) may have been visually simpler than the scenes she used and so the observed tendency to return to

[1] The cumulative probabilities in this figure were computed by weighting all individual data points equally. Thus, the values in this figure (and in Figure 5) may differ slightly from the means reported in the text, which are based on participant means.

[2] The ANOVAs were computed in pixel values, and hence the mean square error values are reported in pixel values. The means were converted to degrees of visual angle for ease of exposition.

semantically informative objects in the former study may have been due to the lack of other objects at which to direct attention. Second, Friedman suggested that the relatively short (4-s) viewing times used by Loftus and Mackworth may not have allowed viewers ample time to distribute their fixations uniformly over the objects in the scenes, whereas the relatively long (30 s) viewing times she used would have allowed them the opportunity to distribute their fixations more evenly over the scene. Third, Friedman suggested that the instructions she used, which emphasized a later memory test in which viewers would be required to note a change in the detail of any object, may have induced viewers to distribute their fixations more evenly across the objects, in contrast to the instructions used by Loftus and Mackworth, which specified preparation for a later recognition test that emphasized the picture as a whole.

In the present study we investigated the role of semantic consistency on fixation density using Loftus and Mackworth's (1978) manipulation of consistency. However, we did not use short viewing times, our scenes were relatively complex, and we gave participants the same viewing instructions as did Friedman (1979). If the difference in the fixation density result between Loftus and Mackworth and Friedman was caused by the viewing times, scene complexity, and instructions, then we should replicate Friedman's result of equivalent fixation densities for consistent and inconsistent objects.

To investigate the degree to which semantic consistency would affect fixation density, we examined two eye movement measures, the proportion of target objects fixated, and the number of discrete entries into the target object regions. The data from these measures are summarized in Table 2.

Proportion fixated. Proportion fixated was defined as the proportion of trials in which at least one fixation landed in the target object region. This measure reflected the degree to which the target objects were fixated and therefore the percentage of trials contributing to other measures of eye movement behavior on the target objects. Overall, the target objects were fixated on .95 of the trials. This proportion did not differ as a function of context, with consistent targets fixated on .94 of the trials and inconsistent targets fixated on .96 of the trials, $F1 < 1$, $MSE = .0059$; $F2(1, 22) = 1.355$, $MSE = .0032$, $p > .25$. Thus, the 15 s of viewing time provided ample opportunity for participants to fixate the majority of target objects. In addition, there was no evidence that viewers were more likely to fixate an inconsistent object than a consistent object.

Table 2
Measures of Fixation Density (Means and Standard Errors by Participants) in Experiment 1

	Scene context condition			
Measure	Consistent	Inconsistent	*SEM*	Difference
Proportion fixated	.94	.96	.018	−.02
No. of entries	2.45	2.90	0.096	−0.45*

*$p < .05$ by participants and items.

Number of entries. Number of entries was defined as the number of times that the eyes moved to the target object region after a saccade that originated from a launch position beyond that region. Number of entries serves as an index of the degree to which the eyes moved from elsewhere in the scene to the target object and is similar to previously used measures of fixation density (Friedman, 1979; Loftus & Mackworth, 1978). Overall, viewers moved their eyes to the target object 2.67 times on average. Fewer entries were made to the target in the semantically consistent condition (2.45) than in the inconsistent condition (2.90), $F1(1, 17) = 10.65$, $MSE = .1667$, $p < .005$; $F2(1, 22) = 23.21$, $MSE = .1246$, $p < .001$.

These data support those reported by Loftus and Mackworth (1978). It appears that fixation density in a scene region can be controlled by the semantic characteristics of that region; in the present case, the eyes tended to return to semantically inconsistent objects over the course of scene viewing. This effect does not require that viewing time be short, that the scenes be visually simple, or that the instructions focus on the scene as a whole rather than on the processing of object detail. Instead, it seems that fixations are distributed unevenly over the objects in a scene (Antes, 1974; Buswell, 1935; Mackworth & Morandi, 1967; Yarbus, 1967) and in particular that semantically interesting and informative objects tend to be refixated (Loftus & Mackworth, 1978).

Semantic Influences on Object Processing Time

To determine whether semantic consistency affected the amount of time that viewers spent processing the target objects, we examined eight measures of target processing time. These measures were first-pass gaze duration, first-pass gaze fixation count, average first-pass fixation duration, second-pass gaze duration, total fixation duration, total fixation count, average fixation duration, and the frequency distributions of fixation durations.[3] Each of these measures was conditionalized on fixation of the target object, so nonfixations did not contribute to the computed means. The data from these measures are summarized in Table 3.

First-pass gaze duration. First-pass gaze duration was defined as the sum of the durations of all fixations between first entry and first exit in a target object region. This measure was thus the sum of the time spent fixating the target object from when the eyes initially landed on the object until the eyes first left that object. First-pass gaze duration has been used as a measure of object encoding during an object identification task (Henderson et al., 1997) and has been shown to reflect the influence of semantic consistency on object encoding during scene perception (De

[3] Another commonly used measure of object processing time is first fixation duration (e.g., De Graef et al., 1990; Henderson, Pollatsek, & Rayner, 1987). The proper interpretation of first fixation duration is currently controversial (e.g., Henderson & Ferreira, 1990; O'Regan, 1992; Rayner & Fischer, 1996; Vitu et al., 1995). Evidence from the present research bearing on this issue will be presented in a separate report.

Table 3
Measures of Processing Time (Means and Standard Errors by Participants) in Experiment 1

Measure	Scene context condition			
	Consistent	Inconsistent	*SEM*	Difference
First-pass gaze duration (ms)	500	669	33.22	−169*
First-pass gaze fixation count	1.68	2.10	0.089	−0.42*
Average first-pass fixation durations (ms)	298	319		−21
Second-pass gaze duration (ms)	590	710	45.60	−120**
Total fixation duration (ms)	1,350	2,028	94.17	−678*
Total fixation count	4.5	6.5	0.290	−2.0*
Average fixation durations (ms)	301	311		−10

$^{*}p < .05$ by participants and items. $^{**}p < .10$ by participants; $p < .05$ by items.

Graef et al., 1990; Friedman, 1979; Loftus & Mackworth, 1978).[4]

In the present study, overall mean first-pass gaze duration in the target object regions was 584 ms. Gaze durations were reliably shorter by 169 ms in the consistent context condition (500 ms) than in the inconsistent context condition (669 ms), $F1(1, 17) = 12.96$, $MSE = 19{,}867$, $p < .005$; $F2(1, 22) = 12.01$, $MSE = 25{,}112$, $p < .005$. Taking gaze duration as a reflection of the initial encoding time for an object, these data suggest that scene context influences the encoding of an object when that object is first encountered in a scene (De Graef et al., 1990; Friedman, 1979; Loftus & Mackworth, 1978).

The difference in first-pass gaze duration across semantic consistency was 169 ms in the present study, 342 ms for the highest versus lowest likelihood conditions in Friedman (1979), 34 ms in Loftus and Mackworth (1978), and 31 ms in De Graef et al. (1990). The fact that these values differ widely across studies suggests that the nature of the stimulus materials and/or the viewer's task can have a great influence on the degree to which semantic consistency influences fixation time.

We also investigated whether semantic consistency affected the fixation time on an object when that object happened to be fixated during an early fixation in the scene. De Graef et al. (1990) provided evidence in their nonobject search paradigm that semantic consistency did not influence fixation durations on a target object when that object happened to be fixated early during scene viewing (on one of the seven initial fixations). However, it could be that the viewers in that study were oriented away from processing the meaning of the scene by the search task (De Graef et al., 1990; Rayner & Pollatsek, 1992). To investigate this question in the present study, we conducted a post hoc analysis to examine first-pass gaze duration on the target objects given that the target was fixated early during scene viewing. There were not enough data to examine gaze duration given that the object was fixated immediately (after the first saccade in the scene), but we were able to compute first-pass gaze duration for 16 participants given that the target was fixated after one or two saccades and for 17 participants given that the target was fixated after one, two, or three saccades in the scene. We were also only able to conduct analyses over participants because there were too many empty cells in the items analysis.

The mean first-pass gaze duration on a target object given that it was fixated after one or two saccades was 513 ms. There was a nonreliable 117-ms difference between semantic consistency conditions, with an initial first-pass gaze duration of 455 ms for the consistent object and 572 ms for the inconsistent object, $F1(1, 15) = 2.189$, $MSE = 49{,}656$, $p > .10$. The first-pass gaze duration on the target given that it was fixated after one, two, or three saccades in the scene was 526 ms. There was a reliable 147-ms consistency effect, with an initial first-pass gaze duration of 452 ms for the consistent object and 599 ms for the inconsistent object, $F(1, 16) = 5.64$, $MSE = 32{,}566$, $p < .05$. These data thus suggest that fixation time on an object is affected by the consistency of the object with the scene, even when the object happens to be fixated relatively early during scene viewing.

First-pass gaze fixation count. First-pass gaze fixation count was defined as the number of individual fixations between first entry and first exit of the target object region. This count is the number of fixations whose durations were summed to produce the first-pass gaze duration. Overall, the mean first-pass gaze fixation count in the target object regions was 1.89. First-pass gaze fixation count was reliably affected by semantic consistency, $F(1, 17) = 11.11$, $MSE = .1414$, $p < .005$; $F2(1, 22) = 11.20$, $MSE = .1624$, $p < .005$, with .42 fewer first pass gaze fixations in the consistent condition (1.68) than the inconsistent condition (2.10).

Average first-pass fixation duration. Average first-pass fixation duration was defined as the first-pass gaze duration divided by first-pass gaze fixation count. The mean average first-pass fixation duration was 309 ms and was 21 ms shorter in the consistent condition (298 ms) than in the inconsistent condition (319 ms). No inferential statistics were computed on these data because they were not independent of the gaze duration and gaze fixation count data.

[4] Both Friedman (1979) and Loftus and Mackworth (1978) used a fixation time measure that they referred to as "duration of the first fixation on a target object." This measure is equivalent to our first-pass gaze duration.

Second-pass gaze duration. Second-pass gaze duration was defined as the sum of all fixations between second entry into and second exit from a target object region. Second-pass gaze duration is a measure of the processing time on an object the second time the eyes move to it from another region of the scene. The overall mean second-pass gaze duration in the target object regions was 650 ms. Second-pass gaze durations were shorter by 120 ms when the object was consistent with the scene context (590 ms) than when it was inconsistent (710 ms). This difference was reliable by items and approached reliability by subjects, $F1(1, 17) = 3.423$, $MSE = 37{,}421$, $p = .08$; $F2(1, 21) = 5.504$, $MSE = 53{,}890$, $p < .05$. These data suggest that semantic consistency continues to influence object processing when the object is fixated a second time.

Friedman (1979) reported that fixation times on low-likelihood objects decreased with additional fixations on those objects. As can be seen in Table 3, in the present study, gaze durations for both semantically consistent and inconsistent target objects increased from the first to second pass. There was also a reduction in the size of the consistency effect, from 169 ms on the first encounter to 120 ms on the second encounter. To determine whether these differences were reliable, we conducted an additional ANOVA treating semantic consistency (consistent vs. inconsistent) and gaze duration moment (first pass vs. second pass) as within-subjects and within-items factors. The results indicate that the main effect of semantic consistency was reliable, $F1(1, 17) = 13.06$, $MSE = 28{,}673$, $p < .005$; $F2(1, 22) = 14.69$, $MSE = 36{,}312$, $p < .005$. The main effect of fixation moment approached reliability by participants but not items, $F1(1, 17) = 3.044$, $MSE = 25{,}549$, $p = .10$; $F2 < 1$, $MSE = 32{,}919$. Most importantly, there was no interaction between fixation moment and semantic consistency ($F1 < 1$, $MSE = 28{,}614$; $F2 < 1$, $MSE = 44{,}085$). Thus, we have no evidence that either average gaze durations or the size of the consistency effect changed from the first to the second encounter with the object, although this analysis was low in statistical power and should be treated cautiously.

Total fixation duration. Total fixation duration was defined as the sum of the durations of all fixations within the target object region from scene onset to scene offset. Averaged over scenes and conditions, the mean total fixation duration on the target objects was 1,689 ms. Total fixation duration was reliably affected by semantic consistency, $F1(1, 17) = 25.91$, $MSE = 159{,}617$, $p < .001$; $F2(1, 22) = 22.08$, $MSE = 249{,}996$, $p < .001$, with total durations 678 ms longer in the inconsistent condition (2,028 ms) than in the consistent condition (1,350 ms).

Total fixation count. Total fixation count was defined as the total number of discrete fixations within a target object region from scene onset to scene offset. This count was the number of fixations whose durations were summed to produce the total fixation duration measure. Overall, the mean total fixation count on the target object regions was 5.5 fixations. These data showed a similar pattern to the total fixation time data, with a reliable 2.0 more fixations in the inconsistent condition (6.5 fixations) than in the consistent condition (4.5 fixations), $F1(1, 17) = 25.12$, $MSE = 1.5086$, $p < .001$; $F2(1, 22) = 26.27$, $MSE = 1.945$, $p < .001$.

Average fixation duration. Average fixation duration was defined as total fixation time divided by total fixation count. The mean average fixation duration in the target object region was 307 ms and was 10 ms shorter in the consistent context condition (301 ms) than in the inconsistent context condition (311 ms). No inferential statistics were computed on these data because they were not independent of the total time and total fixation count data.

Frequency Distributions of Fixation Durations Over the Entire Scene

Finally, we computed the frequency distributions of the durations of all fixations over the scenes as a function of semantic consistency. All individual fixations were segregated into 20-ms bins. Frequencies for each bin were then divided by the total number of fixations in that condition. The resulting distributions are shown in Figure 4. As can be seen, the distributions of individual fixations were similar over the scenes in the consistent and inconsistent target conditions.

Discussion

The purpose of Experiment 1 was to examine eye movement patterns during complex real-world scene viewing. The main issue was the extent to which the semantic analysis of peripheral scene regions can control the placement of initial fixations in a scene. We found that the initial saccade in a scene was no more likely to be to a semantically inconsistent than a semantically consistent object. In fact, to the extent that there was any influence at all on initial saccades to the target objects, that influence tended to favor the semantically consistent objects, with some tendency for viewers to move their eyes more quickly and farther to the semantically consistent objects. Importantly, though, the amplitude of the average saccade was about 3°, a value that is typical of the scene viewing literature (Henderson & Hollingworth, 1998). These data strongly suggest that objects are semantically interpreted within a relatively limited region within and near to the fovea during viewing of complex scenes.

One explanation for the finding that the eyes were not drawn to inconsistent objects in the present study is that the meaning of the scene was not processed early enough during scene viewing. This hypothesis does not appear to have much a priori plausibility, given the strong evidence that scene semantics can be determined quickly enough to influence object processing even when the scene is shown for a duration that is shorter than a typical fixation (e.g., Biederman et al., 1982; Boyce & Pollatsek, 1992; Boyce, Pollatsek, & Rayner, 1989). Importantly, we have similarly found clear effects of scene semantics on object processing during a single fixation with the present scenes and target objects (Henderson, Hollingworth, & Weeks, 1996). For example, participants are strongly biased toward saying that an object that is semantically consistent versus inconsistent

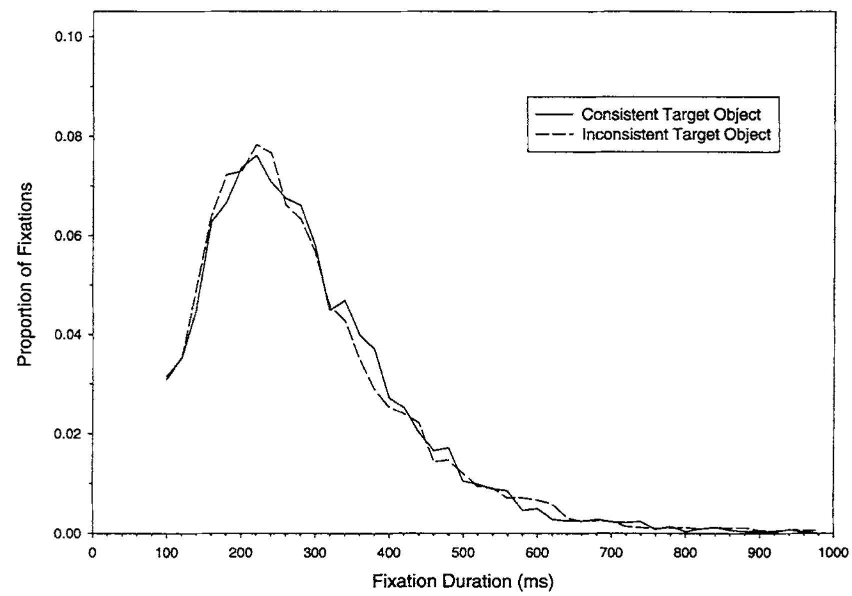

Figure 4. Frequency distributions of fixation durations over the entire scene, for all participants and all scenes, as a function of target object semantic consistency in Experiment 1. The bin width was 20 ms.

with the scene was present in the scene following a 200-ms presentation of the scene (Hollingworth & Henderson, 1998). This bias must be based on a rapid determination of the meaning of the scene. Furthermore, the results of Experiment 2 demonstrate that participants can quickly determine the meaning of the scene during a visual search task.

In addition to the main question of interest in this study, we also examined two subsidiary issues. First, we investigated whether viewers would differentially distribute their fixations over the objects in a scene as a function of semantic consistency. Loftus and Mackworth (1978) found that viewers returned their gaze to semantically inconsistent objects more often than to semantically consistent objects over the course of scene viewing. These results are similar to those found in studies that have examined only visual informativeness or that have not distinguished visual and semantic informativeness (Antes, 1974; Buswell, 1935; Mackworth & Morandi, 1967; Yarbus, 1967).

Given the evidence that viewers tend to return their gaze to semantically inconsistent objects, why did Friedman (1979) fail to observe this effect? One possibility may have to do with Friedman's manipulation of semantic consistency. In that study, objects in scenes ranged continuously from highly consistent to somewhat inconsistent (e.g., hockey sticks in a kitchen scene). If the manipulation was weaker in Friedman's study because the objects were less clearly inconsistent, and if viewers tended to distribute their fixations in a graded fashion across the multiple, somewhat inconsistent objects that were present, the overall effect of inconsistency on fixation density would be smaller and therefore more difficult to detect. In the present experiment, in contrast, when a scene contained an inconsistent object, that object was highly anomalous and was the only such object in the scene. Thus, a relatively larger proportion of the fixations would be expected to return to that object and therefore the effect of semantic inconsistency on fixation density would be easier to detect.

The final question we investigated in Experiment 1 was the extent to which the semantic consistency between an object and its scene would be reflected in different measures of eye movement behavior. The primary finding was that all of the measures of object processing time that we examined showed robust effects of semantic consistency. Specifically, viewers produced shorter first-pass gaze durations, second-pass gaze durations, and total viewing durations on the semantically consistent objects than on the semantically inconsistent objects. This pattern was also observed for measures of the number of discrete fixations made on the objects, with fewer first-pass gaze fixations and total fixations on the consistent target objects than on the inconsistent objects. Similarly, viewers looked back to the consistent objects fewer times, with fewer entries into the target object region when the object was consistent with the scene. Finally, the increases in processing time seemed to be caused both by an increase in the number of fixations and the durations of those fixations, as shown by the average fixation durations.

Experiment 2

In Experiment 1, viewers' eyes were not initially drawn to semantically inconsistent objects in the visual periphery. One interpretation of this result is that viewers found it difficult to evaluate the semantic consistency between an object and its scene until they had fixated relatively close to that object. An alternative hypothesis, however, is that viewers had little motivation to move their eyes to the inconsistent target objects. For example, because the participants were attempting to learn the scenes for an expected memory test, they may not have felt the need to fixate inconsistent objects any earlier than they fixated consistent objects. It could be that participants were more motivated to fixate the inconsistent objects quickly in Loftus and Mackworth's (1978) study because overall viewing time was relatively short.

The purpose of Experiment 2 was to determine whether participants would be able to move their eyes more quickly to semantically inconsistent target objects if they were strongly motivated to find the target objects as quickly as possible. To investigate this possibility, we used a visual search task in which participants were asked to determine as quickly as possible whether a prespecified target object was present in a given scene. In half the target-present trials, the target was semantically consistent with the scene, whereas in the other half, the target was semantically inconsistent with the scene. If the semantic inconsistency of a peripheral scene region can draw the eyes, as suggested by Loftus and Mackworth (1978), then participants should be able to make an earlier and longer saccade to the target object in the inconsistent condition than the consistent condition and so should find the inconsistent objects faster than the consistent objects. In the extreme case, the first saccade in the scene should more often be directed to the inconsistent than the consistent target object (Loftus & Mackworth, 1978).

Method

Participants

Twenty-six Michigan State University undergraduates took part in this experiment. All participants had normal or corrected-to-normal vision, were naive about the purposes of the research, and received partial credit for their introductory psychology courses or remuneration for their participation. None of these participants had taken part in Experiment 1.

Stimuli

The stimuli were 20 line drawings of scenes, as described in Experiment 1. The scenes were modified from Experiment 1 in the following manner. First, the 10 scenes from van Diepen and De Graef's (1994) set that were least familiar to North American viewers were eliminated to increase the contextual constraint for individual target objects across the set of scenes. These were replaced by six new scenes created by electronically tracing the principal contours of photographs of common environments from around the East Lansing, Michigan, area. Old and new scenes were stylistically indistinguishable. Second, a number of the scene pairings were changed to accommodate the new scenes, and several new target objects were drawn to maintain a high degree of contextual constraint across newly paired scenes. Third, all scene contours were presented as darker anti-aliased, gray-scale lines (16 levels of gray) against a lighter gray background to make them appear more smooth and sharp.

Procedure

The procedure was the same as in Experiment 1, with the following exceptions: Participants were instructed that their task was to locate objects in scenes. They were informed that before the presentation of each scene, a target object would be specified by a label and that their task was to indicate as quickly as possible via a buttonpress whether that target object was present in the scene. After the instructions were reviewed, the eye-tracking apparatus was calibrated as described for Experiment 1. Once calibrated, the participant was given 5 practice trials, 2 target-present trials in each of the context conditions and 1 target-absent trial, using scenes that were not included in the experimental stimuli. After the practice trials, any questions the participant had were answered. The eyetracker was then recalibrated and the participant was shown 26 experimental trials, 10 trials containing target objects in each of the two context conditions and 6 target-absent filler trials.

A trial consisted of the following events. First, the calibration screen was shown and calibration was checked. The eyetracker was recalibrated whenever calibration was deemed inaccurate. After the calibration check, the participant fixated the center marker to indicate that he or she was ready for the trial to begin. The experimenter then started the trial: The fixation display was replaced by the target label, which was presented for 3,000 ms. The scene was then presented and remained visible until the participant pushed one of two buttons to indicate whether the target object was present.

A given participant saw all 20 test scenes, 10 in which the target object was consistent with the scene and 10 in which it was not. Each participant also saw six filler scenes in which the target object was not present; three of these scenes were preceded by a consistent target object label and three with an inconsistent target object label. Across participants, each test scene appeared in each target-present context condition an equal number of times. The order of scene presentation (and hence condition presentation) was determined randomly for each participant. The entire experiment lasted approximately 45 min.

Apparatus

The apparatus was the same as in Experiment 1, with the exception that buttonpresses were collected using a button panel connected to a dedicated input–output (I/O) card that was interfaced with the computer.

Results

In the following analyses, one item was removed because the target object was inadvertently placed at the initial fixation point. In addition, trials using the "chicken" target label were eliminated, as some participants did not accept that the label referred to the target picture (a rooster). The elimination of these latter trials resulted in two empty cells in the items analyses, which were filled by the mean of the other items in that condition. Finally, trials on which the target was fixated on the first (i.e., experimenter-induced) fixation were excluded from the analysis, although this occurred only twice across all trials, accounting for 0.4% of the data.

Percent Correct

Overall, participants found the target objects in the target-present condition on 96.8% of the trials. This value did not differ as a function of semantic consistency, with 97.9% correct in when the target was consistent with the scene and 95.7% correct when the target was inconsistent with the scene, $F1(1, 25) = 1.499$, $MSE = .0040$, $p > .2$. The false-alarm rate when the target was absent was 6.4% overall and tended to be higher in the consistent condition (11.5%) than in the inconsistent condition (1.3%), $F1(1, 25) = 5.332$, $MSE = .0256$, $p < .05$; $F2(1, 18) = 4.000$, $MSE = .0079$, $p > .10$. This pattern of hit and false-alarm rates suggested that participants were biased to respond that the target was present when it was semantically consistent with the scene.

Extrafoveal Semantic Analysis

Because we were interested in the degree to which participants could find target objects as a function of the semantic consistency of those targets with the scenes in which they appeared, the primary dependent measures reflected the time to find the target object. These measures were the total search time, the probability of immediate target fixation, the number of fixations to the target, and the amplitude of the initial saccade to the target. Each of these measures was conditionalized on a correct "present" response. The eye movement measures were also conditionalized on fixation of the target object, so nonfixations did not contribute to the computed means. The data from these measures are summarized in Table 4.

Total search time. Total search time was defined as the total amount of time that the scene was presented from initial onset until the decision button was pressed to trigger scene offset. If the eyes are attracted to semantically inconsistent objects, total search time should be shorter in the inconsistent condition than in the consistent condition. Overall, the mean total search time was 1,241 ms. However, semantically consistent objects tended to be found faster than inconsistent objects, with overall search times of 1,174 and 1,309 ms in the consistent and inconsistent conditions, respectively, $F1(1, 25) = 8.031$, $MSE = 29{,}478$, $p < .01$; $F2(1, 18) = 3.269$, $MSE = 63{,}298$, $p < .10$.

Probability of immediate target fixation. Overall, the mean probability of fixating the target after a single saccade was .23 and was unaffected by semantic consistency. The probability of immediately fixating the target object was .26 in the consistent condition and .20 in the inconsistent condition, $F1(1, 25) = 1.248$, $MSE = .0414$, $p > .25$; $F2 < 1$, $MSE = .0320$. This value is about twice that observed in Experiment 1, suggesting that the search instructions effectively induced participants to seek out the targets.

The cumulative probability of having fixated the target after two saccades was .67. This probability was reliably higher in the consistent condition (.74) than in the inconsistent condition (.59), $F1(1, 25) = 9.609$, $MSE = .0295$, $p < .01$; $F2(1, 18) = 4.205$, $MSE = .0406$, $p = .05$. Thus, early in the time course of scene viewing, participants were more likely to have detected a target object in the consistent than in the inconsistent condition.

The cumulative probability of having fixated the target as a function of semantic consistency and ordinal fixation number is shown in Figure 5 (see Footnote 2). In general, targets were fixated much earlier than in Experiment 1, as would be expected under search instructions. In addition, there is a clear pattern indicating earlier fixation of consistent target objects than inconsistent target objects.

Number of fixations to the target. Overall, the first fixation on the target was 3.29 fixations into scene viewing. There was a reliable effect of semantic consistency, with 3.11 fixations in the consistent condition and 3.46 fixations in the inconsistent condition, $F1(1, 25) = 10.409$, $MSE = .1569$, $p < .01$; $F2(1, 18) = 4.503$, $MSE = .2996$, $p < .05$.

Amplitude of initial saccade to target. Overall, the amplitude of the initial saccade to the target was 3.68°. The initial saccade was 3.86° to the consistent target object and 3.49° to the inconsistent target. This difference was reliable by items but not by participants, $F1(1, 25) = 1.724$, $MSE = 2{,}827$, $p > .15$; $F2(1, 18) = 4.503$, $MSE = 1{,}097$, $p < .05$ (see Footnote 1). These values are roughly half the size of the 7.32° average saccade length observed by Loftus and Mackworth (1978).

Fixation Density

In Experiment 2, fixation density was examined using the proportion fixated measure. Number of entries was not analyzed because the fixation on the target was typically terminated by a manual buttonpress indicating the presence of the target.

Overall, the target objects were fixated on .81 of the trials. The consistent targets were less likely to be fixated (.79 of the trials) than the inconsistent targets (.83 of the trials), although this difference was not reliable, $F1(1, 25) = 1.742$, $MSE = .0153$, $p > .15$; $F2 < 1$, $MSE = .0242$.

Table 4
Measures of Extrafoveal Semantic Analysis (Means and Standard Errors by Participants) in Experiment 2

Measure	Scene context condition: Consistent	Inconsistent	*SEM*	Difference
Total search time (ms)	1,174	1,309	33.67	−135**
Probability of immediate target fixation	.26	.20	.040	.06
No. of fixations to the target	3.11	3.46	0.078	−0.35*
Amplitude of initial saccade to the target (deg)	3.86	3.49	0.203	0.37

*$p < .05$ by participants and items. **$p < .05$ by participants; $p < .10$ by items.

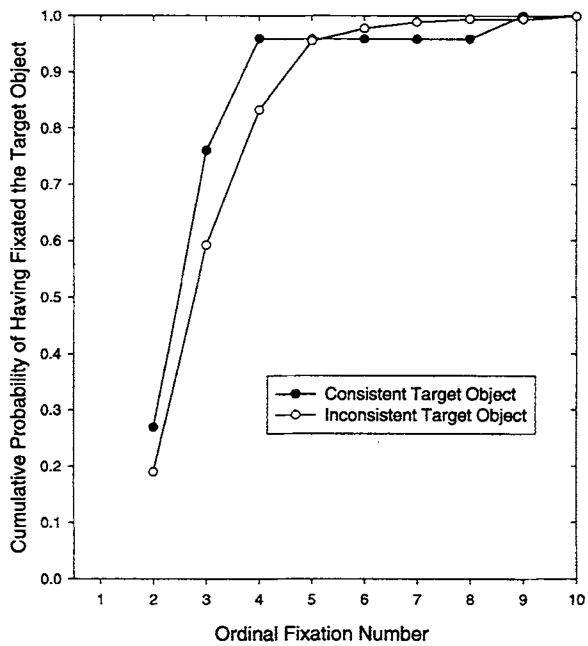

Figure 5. Cumulative probability of having fixated the target object as a function of the ordinal fixation number and semantic consistency in Experiment 2.

Frequency Distributions of Fixation Durations Over the Entire Scene

The frequency distributions of all fixation durations over the scenes as a function of semantic consistency were generated in the same way as in Experiment 1 and are shown in Figure 6. In general, the distributions of individual fixation durations were highly similar across the semantic consistency manipulation. It is interesting to compare this distribution with that from Experiment 1. Although the modes of both distributions were similarly placed (at approximately 220 ms), the distributions from Experiment 1 contained more durations in the 300- to 600-ms range. To investigate this potential difference in fixation duration more directly, we calculated the mean fixation duration across all fixations on the scenes for each experiment. In Experiment 1 the mean fixation duration was 285 ms, whereas in Experiment 2 it was 247 ms, a difference of 38 ms. The memory task in Experiment 1 may have biased participants to increase the durations of individual fixations to facilitate committing those objects to memory, whereas the search task in Experiment 2 required only that the target object be discriminated from other objects in the scenes.

Discussion

The purpose of Experiment 2 was to provide an additional test of the hypothesis that the eyes are attracted by peripheral semantic inconsistency during real-world scene viewing. Participants were asked to determine whether prespecified target objects were present in line drawings of complex, natural scenes. Each target object was either semantically consistent or inconsistent with the scene within which it appeared. If semantic inconsistency in the visual periphery can draw the eyes, then targets should be fixated (and therefore found) more quickly when they are inconsistent with the scene. In contrast to this prediction, the main result in Experiment 2 was that semantically consistent targets were found more quickly, were fixated earlier, and tended to

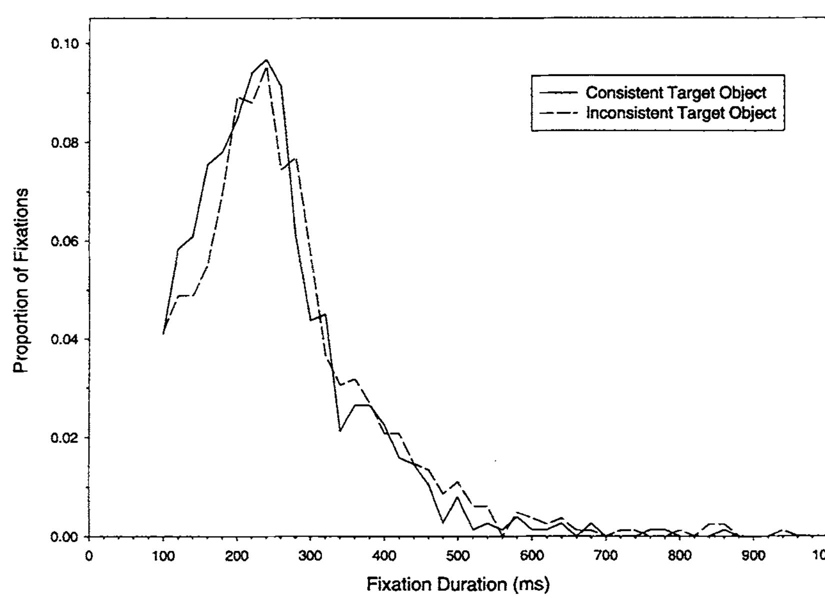

Figure 6. Frequency distributions of fixation durations over the entire scene, for all participants and all scenes, as a function of target object semantic consistency in Experiment 2. The bin width was 20 ms.

be fixated after a saccade of larger magnitude than inconsistent targets. Furthermore, although mean saccade lengths to the targets were larger in Experiment 2 (3.68°) than in Experiment 1 (3.05°), they were still well below the saccadic amplitudes observed by Loftus and Mackworth (1978). In fact, they were much more similar to the saccadic amplitudes reported in other scene scanning experiments (Antes, 1974; Saida & Ikeda, 1979; Shiori & Ikeda, 1989; van Diepen et al., 1998; see Henderson & Hollingworth, 1998). The results, then, provide converging evidence with those of Experiment 1 for the conclusion that the eyes are not immediately drawn to semantically inconsistent objects in scenes.

Why were consistent objects found sooner than inconsistent objects? A likely explanation is that participants used visual features of the target objects, in conjunction with knowledge about the likely positions of consistent targets, to drive their eye movements. For example, when looking for a mixer in a kitchen, the eyes could be driven by a combination of visual factors (parafoveal blobs that have mixerlike features) and semantic knowledge of the scene (this is a kitchen, and mixers are generally found on flat surfaces above the floor). In contrast, when looking for a live chicken in the kitchen, the latter source of information is no longer available or might even be misleading. Thus, search based on appropriate nearby visual features plus general knowledge about the scene layout based on a global analysis (and perhaps local analyses of fixated regions) can explain the benefit for the consistent targets. Importantly, there was no evidence at all that the eyes were immediately drawn to semantically inconsistent targets in the visual periphery despite the fact that the participants had strong motivation to find and fixate those targets if they could.

General Discussion

The present research was designed to investigate eye movement behavior during complex real-world scene viewing. Our primary focus was the extent to which semantic analysis of a scene region in the visual periphery can control the placement of the initial fixations in the scene. In addition, we also examined the extent to which the semantic relationship between an object and its scene influences the distribution of fixations in the scene over the course of scene viewing and the extent to which the semantic relationship between an object and its scene affects on-line measures of eye movement behavior. These issues are addressed individually.

Control of Initial Fixation Placement

To what extent does the semantic analysis of scene regions in the visual periphery control the placement of eye fixations in a scene? In an initial study bearing on this question, Loftus and Mackworth (1978) provided evidence that viewers tend to move their eyes a greater distance to a semantically inconsistent than a semantically consistent object that occupies the same position in a scene. However, De Graef et al. (1990), using a similar manipulation of semantic consistency, were unable to replicate this effect. In Experiment 1 of this research, we used Loftus and Mackworth's (1978) viewing task but used complex, realistic scene stimuli like those used by De Graef et al. (1990). In contrast to Loftus and Mackworth but similar to De Graef et al., we found that participants did not fixate inconsistent objects earlier or from farther away than they fixated consistent objects. Furthermore, the average initial saccade to the target objects in Experiment 1 was about 3°, suggesting that initial saccades to objects were typically controlled by information that was present in the fovea and parafovea during the current fixation. In Experiment 2, we introduced a visual search task to provide converging evidence concerning peripheral semantic analysis and fixation placement. Participants were first provided with a target object label and were then shown a scene for free viewing. The task was to indicate as quickly as possible whether the target object appeared in the scene. In this task, participants were motivated to move their eyes immediately to semantically inconsistent objects if they were able to do so. As in Experiment 1, there was no evidence that the eyes were drawn to semantically inconsistent objects. Instead, consistent target objects were detected faster and fixated sooner than inconsistent objects. As in Experiment 1, the average saccade to the target object in both conditions was less than 4°. Our conclusion is that the eyes are not initially controlled by peripheral semantic analysis of local scene regions.

To date, we know of four experiments that have examined the effects of semantic consistency on eye movement patterns. Of these, one has shown that the eyes are drawn to inconsistent objects (Loftus & Mackworth, 1978), and three have shown that they are not (De Graef et al., 1990, and the two experiments reported here). Why might Loftus and Mackworth have found that viewers' initial fixations were drawn to semantically inconsistent objects in the periphery? There seem to be at least three possible explanations for the differences across studies. First, the scenes used in the present research and by De Graef et al. were visually more complex than those used by Loftus and Mackworth. Loftus and Mackworth's stimuli contained a small number of objects and background contours and a large amount of empty space. In contrast, the line drawings derived from photographs that we (and De Graef et al., 1990) used contained many objects and contours. Pollatsek, Rayner, and Collins (1984) demonstrated that viewers can identify line drawings of objects at 10° from fixation before an eye movement when the objects are presented in an otherwise empty visual field. More complete semantic analysis of the peripheral objects in Loftus and Mackworth's study could have resulted in peripheral semantic control of initial saccades. The relatively long saccades observed by Loftus and Mackworth can also be explained by the visual simplicity of the scenes used in that study: Given that there were relatively few objects that were separated by a relatively large amount of blank space, given that viewers are likely to saccade to objects rather than empty space and given that the average saccadic amplitude should reflect the distance between objects, one would expect large average saccade amplitudes under these conditions.

Note that the scene stimuli used in the present research and in De Graef et al. (1990) are themselves a significant visual simplification of natural environments. Although our stimuli were derived from photographs of real-world scenes, these line drawings did not preserve all the contours of the photographic images. There is evidence, however, that the scene stimuli we used produce patterns of eye movements that are highly similar to those produced when full-color photographs of natural scenes are viewed. In a recent study (summarized by Henderson & Hollingworth, 1998), eye movements were monitored while participants viewed either line drawings, full-color photographs, or 3-D computer-rendered images of natural scenes under instructions to prepare for a subsequent memory test. Unlike the current research, the scenes contained only semantically consistent objects. Mean saccade lengths and mean fixation durations were the principal measures examined. Although fixation durations were reliably longer on photographs than line drawings, this difference was small (12 ms). Importantly, there was no difference in the spatial extent of saccades across stimulus type, with a mean saccade length of approximately 2.4°. Given that the line drawings used in the present research and in De Graef et al. (1990) led to eye movement patterns like those found in photographs of scenes, it seems likely that the level of visual complexity present in these images is representative of that found in natural, complex scenes.

A second potential explanation for the difference in results between Loftus and Mackworth (1978) on the one hand and De Graef et al. (1990) and the present research on the other hand is that semantic consistency and visual consistency may have been correlated in Loftus and Mackworth's experiment, as suggested by several theorists (De Graef et al., 1990; Rayner & Pollatsek, 1992). The result of such a correlation would be that the effect of semantic consistency on initial saccades would actually be attributable to visual factors. This problem could arise if, for example, the consistent target objects were initially drawn in the scenes and then the target objects were swapped across scenes. The problem with this method is that different scenes might be drawn in slightly different styles, so that objects drawn in one scene would be visually more consistent with the rest of the scene than would objects exchanged across scenes. Although we do not know for certain whether this was a problem in Loftus and Mackworth's experiment, we do know that it was not a problem in the present research: Target objects were drawn independently of both the consistent and inconsistent scene background.

A third possible explanation for the difference in results between the present research and those of Loftus and Mackworth (1978) is the difference in the visual angle subtended by the scenes in the two studies: Our scenes subtended 10° × 14.5°, whereas Loftus and Mackworth's scenes subtended 20° × 30°. Larger scenes might lead to greater peripheral semantic analysis and therefore to longer saccades. Contrary to this hypothesis, however, the scenes used by De Graef et al. (1990) subtended 20° × 30°, but, like us, they observed no early influence of peripheral object semantics on fixation placement. Furthermore, the average saccadic amplitudes that we observed in Experiments 1 and 2 were similar to the averages reported in the majority of scene viewing studies reported in the literature across a range of scene sizes (Henderson & Hollingworth, 1998). Importantly, evidence from several studies that have explicitly examined peripheral object processing during scene viewing suggests that object information is not typically acquired from beyond about 4.5° from fixation (Nelson & Loftus, 1980; Shiori & Ikeda, 1989; van Diepen et al., 1998).

A final possible explanation for the discrepancy in results is simply statistical error in Loftus and Mackworth's (1978) experiment. This explanation seems possible given the relatively low spatial (>1°) and temporal (60 Hz) resolution of the eye-tracking equipment that was available at the time of that study.

Fixation Density During Scene Viewing

The second question we investigated in the present research was whether semantic consistency would influence the density of eye fixations on the objects in a scene. Loftus and Mackworth (1978) found that viewers returned their gaze to semantically inconsistent objects more often than to semantically consistent objects over the course of scene viewing. On the other hand, Friedman (1979) found that the number of times that the eyes moved to fixate a particular object was equivalent across levels of semantic consistency. The results of Experiment 1 replicate the Loftus and Mackworth results concerning fixation density: Viewers moved their eyes more often to semantically inconsistent than to semantically consistent objects. These data show that the eyes tend to return to semantically informative scene regions over the course of scene viewing.

The most likely explanation for the lack of a fixation density difference as a function of semantic consistency in Friedman's (1979) study would seem to rest on the nature of the semantic consistency manipulation used in that study. In Friedman's experiment, the objects in each scene ranged from completely consistent to somewhat inconsistent. In contrast, in both Loftus and Mackworth's (1978) study and our research, the target objects were semantically anomalous in the inconsistent condition. Thus, the difference in results is likely attributable to the nature and strength of the semantic consistency manipulation across studies.

Semantic Processing and Measures of Eye Movement Patterns

The final issue we investigated was the extent to which semantic consistency would be reflected in different measures of eye movement behavior. From a theoretical perspective, this issue is important because, as we have shown, different components of eye movement behavior may be controlled by different aspects of a viewed scene. A complete model of eye movement control will have to account for all aspects of the eye movement record. In addition, from an empirical perspective, it is important to determine the extent to which eye movement behavior in reading generalizes to eye movement behavior in scene viewing because many of the current theories of eye movement control derive from studies of reading (e.g., Henderson & Ferreira, 1990; Morrison, 1984; Rayner & Pollatsek, 1989). The primary finding from Experiment 1 was that all of the measures of object processing time that we examined showed robust effects of semantic consistency: Viewers produced shorter first-pass gaze durations, second-pass gaze durations, and total viewing durations on the semantically consistent objects than on the semantically inconsistent objects. This pattern was also observed for measures of the number of discrete fixations made on the objects, with fewer first-pass gaze fixations and total fixations on the consistent target objects than on the inconsistent objects. Similarly, viewers looked back to the consistent objects fewer times, with fewer entries into the target object region when the object was consistent with the scene. Finally, the increases in processing time seemed to be caused both by an increase in the number of individual fixations and the durations of those fixations.

Eye Movement Control During Scene Viewing: A Saliency Map Framework

In this section we outline a framework for understanding eye movement control during scene viewing. The eye movement control framework that we propose is a modification of a framework originally proposed by Henderson (1992b), which in turn was an extension of the model proposed by Morrison (1984; see also Henderson & Ferreira, 1990; Rayner & Pollatsek, 1992) to account for the cognitive control of eye movements in reading. This framework is meant to account for macrolevel eye movement behavior such as the positioning of the eyes in a scene and the maintenance of fixations at those positions in the service of visual and cognitive processing. The framework is not meant to account for microlevel eye movement behavior such as convergence, nystagmus, and ocular drift and ignores other oculomotor factors such as the global effect (Findlay, 1982) and the optimal viewing position effect (O'Regan, 1992; Vitu, O'Regan, Inhoff, & Topolski, 1995). Although these phenomena are important, they are beyond the scope of our discussion here.

In our *saliency map framework,* a map of potential saccade targets is generated from an early parse of the scene into "blobs," or visually differentiated regions of potential interest, and a background that is relatively undifferentiated. This initial parse may be based on a fast initial analysis of the low-frequency information available in the scene and may also lead to an initial generation of the scene category (Schyns & Oliva, 1994). After the initial parse, each region of potential interest is assigned a weight within a saliency map (Mahoney & Ullman, 1988). The saliency map is a representation of space that codes the positions and relative salience of scene regions.

Initially, salience is determined by low-level stimulus factors such as luminance, contrast, texture, color, contour density, and so on, with a bias toward higher salience for closer regions (Rayner & Pollatsek, 1992). Salience may also be modified by the viewer's task as long as the task modification can be based on low-level stimulus factors. During any given fixation, visuospatial attention is allocated to the region that is assigned the highest weight within the saliency map at that time (Koch & Ullman, 1985) and the eyes are programmed to saccade to the attended region (Henderson, 1992b; Henderson, Pollatsek, & Rayner, 1989). Thus, initial fixation placement will be determined by visual rather than semantic factors. Once the eyes land, the amount of time they remain in a region will be determined primarily by the latency of successful completion of processing at that region. Completion of processing is taken to include both perceptual and cognitive analysis. For example, to the extent that semantic interpretation and memory encoding are required for the viewer's task, these factors will be considered in the determination of successful processing. After successful completion of processing, the saliency weight for that region will be reduced and attention will be released. Attention then shifts to the new region with the highest saliency weight and the eyes are programmed to move to that region. Finally, if processing of the current region is unlikely to be successful (or will take too long) given the present fixation position (e.g., because information is accumulating too slowly), then attention will shift within the current region and a refixation within that region will be programmed in an attempt to increase the likelihood of successful completion (for a similar idea in reading, see McConkie, 1979). Thus, regions that are difficult to analyze will be more likely to receive refixations. Refixation rate on an object may also be increase if the eyes are not optimally positioned within that object (Henderson, 1993; see also O'Regan, 1992; Vitu et al., 1995).

In this view of eye movement control during scene viewing, initial movement of the eyes should be controlled by stimulus features rather than by semantic features. There is good evidence that the eyes are attracted by salient physical characteristics in a scene. For example, the eyes are attracted to a small back-and-forth movement of an object in a scene (Boyce & Pollatsek, 1992), and this effect is found even when viewers are not aware of the movement (De Graef, 1998). There is also strong evidence that the eyes are drawn to static regions that are visually informative (Antes, 1974; Buswell, 1935; Mackworth & Morandi, 1967; Yarbus, 1967). In our framework, as individual scene regions are fixated and converted from meaningless blobs to meaningful elements over the course of scene scanning, saliency weights will be modified to reflect the relative cognitive interest of those regions. That is, after initial fixation at a region, the basis of the saliency weight for that region will change from primarily visual to primarily cognitive. Ultimately, as scene analysis progresses, salience will become heavily determined by factors such as semantic informativeness, the need for additional visual detail to aid with semantic interpretation, memory concerns, and so on. The eyes are then likely to be sent to regions of cognitive saliency rather than drawn by regions of visual saliency, leading to greater fixation density and total fixation time on semantically interesting objects and scene regions. However, until a fixation has landed relatively near (within about 3°–4°) to a particular region at least once, its saliency weight will be determined by visual rather than semantic factors. An interesting and unexplored question, in this view, is how semantic and visual factors trade off in the assignment of saliency weights to different scene regions over the course of scene viewing.

In contrast to initial fixation placement, initial fixation durations and refixation probabilities on a region should immediately be affected by semantic consistency. According to the proposed framework, the amount of time the eyes initially remain in a given region is determined by the amount of time required to complete perceptual and cognitive analysis of that region. Under the assumption that semantically inconsistent objects require more time to perceptually analyze, integrate with the scene representation, or commit to memory, first-pass gaze durations on these objects should be longer. In addition, the more difficult processing is for an object, the higher the likelihood an additional fixation within that object will be useful. Finally, to the extent that additional looks back to an object are used for additional cognitive analysis (e.g., additional memory encoding), gaze durations during these second-pass fixations should also be influenced by the same factors that influence first-pass fixations.

In the saliency map framework as outlined thus far, there is no role for the global semantics of the scene on initial fixation placement. However, there is good evidence that a global semantic analysis of a scene can take place within the first fixation on the scene (e.g., Biederman et al., 1982; Boyce et al., 1989; Henderson et al., 1996; Hollingworth & Henderson, 1998). This analysis appears to be based on low spatial frequency information in the scene (Schyns & Oliva, 1994), perhaps leading to the generation of a structural description of the global geometric properties of the scene (Biederman, 1994). Does this global analysis influence the saliency weights assigned to peripheral scene regions before the initial fixation of those regions? The results of our experiments, as well as those of De Graef et al. (1990), suggest that it does not. At the same time, however, it does appear that eye movements can be controlled to some extent by information about scene layout that has been derived from both global scene analysis and local analysis of fixated scene regions. For example, the results of the visual search task used in Experiment 2 demonstrate that the eyes tended to move more quickly to semantically consistent than semantically inconsistent objects during visual search.

Another striking difference between the results of Experiments 1 and 2 was the amount of scene viewing time that elapsed before the initial fixation on the target object. In Experiment 1, there were an average of 10.2 fixations before the initial fixation on the target, whereas in Experiment 2, there were only 3.3 fixations before initial fixation. Furthermore, in a qualitative analysis of the viewing paths taken by the eyes in the two experiments, we observed that the eyes tended to move to the targets in a much more direct line in the search task than in the memory task. To quantify this observation, we computed the ratio of the observed viewing path to the most direct (euclidian) path between the initial fixation position and the first fixation in the target object region for each participant in each scene. In Experiment 1, this ratio was 6.5 (i.e., the path taken by the viewer was 6.5 times longer than a direct path). In Experiment 2, this path was 1.9, or less than twice as long as a direct path. Thus, when participants were searching for a particular target object, they were able to find that object relatively efficiently.

These two findings—more rapid fixation of a consistent than an inconsistent target during visual search and more rapid fixation of a target during visual search than during scanning for memorization—could be taken to undermine our assertion that initial fixation placement is driven by visual rather than semantic aspects of local scene regions. However, these results can also be accommodated by our eye movement control framework in the following manner. When participants were given a target label before scene presentation in Experiment 2, they would be able to generate an expectancy about the visual characteristics of the target. These expectancies could then be used to increase the saliency weights for relatively nearby regions with visual similarity to the target. In this way, search could be guided by visual characteristics of the image, leading to a reduction in the number of regions that would need to be searched. Qualitative analysis of viewing paths supports this hypothesis: Participants often fixated objects that were visually similar to the target on their way to finding the target. For example, participants often fixated on a vase filled with flowers when they were searching for a lamp. In addition, as suggested earlier, a global analysis of the scene could also contribute to search by providing constraints on the likely positions of semantically consistent targets. Positional constraints could be used to modify the saliency weights of unidentified objects. Importantly, neither the finding of faster target fixation in Experiment 2 than in Experiment 1 nor the finding that targets were fixated faster in Experiment 2 when they were semantically consistent with the scene requires that any peripheral semantic analysis of target objects takes place.

Conclusions

The data from the present research, in conjunction with those from previous studies, converge on the following conclusions concerning eye movement behavior during real-world scene viewing: (a) Initial fixation placement in a complex, natural scene is not controlled by a peripheral semantic analysis of individual objects in the scene. (b) Once an object has been fixated, the eyes tend to remain fixated longer on that object if it is semantically informative (inconsistent) than uninformative (consistent) in the context of the scene. (c) The eyes tend to return to semantically inconsistent objects in a scene more often than to consistent objects. (d) Search paths to a specified object tend to be shorter to objects that are consistent with the scene than to objects that are inconsistent with the scene. Taken together, these results support a model of eye movement control during scene viewing in which the eyes are initially driven by visual factors and global scene semantics, with cognitive and semantic aspects of local scene regions playing an increasingly important role as scene exploration unfolds.

References

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103,* 62–70.

Biederman, I. (1994). Higher-level vision. In D. N. Osherson, S. M. Kosslyn, & J. M. Hollrbach (Eds.), *Visual cognition and action* (pp. 121–165). Cambridge, MA: MIT Press.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 531–543.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 556–566.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art.* Chicago: University of Chicago Press.

Crane, H. D. (1994). The Purkinje image eyetracker, image stabilization, and related forms of stimulus manipulation. In D. H. Kelley (Ed.), *Visual science and engineering: Models and applications* (pp. 15–89). New York: Marcel Dekker.

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

De Graef, P. (1998). Prefixational object perception in scenes: Objects popping out of schemas. In G. Underwood (Ed.), *Eye guidance while reading and while watching dynamic scenes* (pp. 313–336). Amsterdam: Elsevier Science.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317–329.

Findlay, J. M. (1982). Global processing for saccadic eye movements. *Vision Research, 22,* 1033–1045.

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

Friedman, A., & Liebelt, L. S. (1981). On the time course of viewing pictures with a view towards remembering. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 137–155). Hillsdale, NJ: Erlbaum.

Henderson, J. M. (1992a). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology, 46,* 319–341.

Henderson, J. M. (1992b). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 260–283). New York: Springer-Verlag.

Henderson, J. M. (1993). Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint. *Canadian Journal of Experimental Psychology, 47,* 79–98.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 417–429.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance while reading and while watching dynamic scenes* (pp. 269–293). Amsterdam: Elsevier Science.

Henderson, J. M., Hollingworth, A., & Weeks, P. A., Jr. (1996, November). *The influence of scene context on object perception.* Paper presented at the annual meeting of the Psychonomic Society, Chicago.

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. (1987). Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 449–463.

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

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127,* 398–415.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review, 104,* 524–553.

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 & Psychophysics, 2,* 547–552.

Mahoney, J. V., & Ullman, S. (1988). Image chunking defining spatial building blocks for scene analysis. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective* (pp. 169–209). Norwood, NJ: Ablex.

McConkie, G. W. (1979). On the role and control of eye movements in reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1, pp. 37–48). New York: Plenum.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 667–682.

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.

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 333–354). 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.

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics, 58,* 734–747.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology, 46,* 342–376.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language processes. *Language and Cognitive Processes, 4,* 21–50.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157.

Reilly, R. G., & O'Regan, J. K. (1998). Eye movement control during reading: A simulation of some word-targeting strategies. *Vision Research, 38,* 303–317.

Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics, 25,* 119–125.

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

Shiori, S., & Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception, 18,* 347–361.

van Diepen, P. M. J., & De Graef, P. (1994). *Line-drawing library and software toolbox* (Psychology Rep. No. 165). Leuven, Belgium: University of Leuven, Laboratory of Experimental Psychology.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). In G. Underwood (Ed.), *Eye guidance while reading and while watching dynamic scenes* (pp. 337–355). Amsterdam: Elsevier Science.

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning strings and reading texts. *Perception & Psychophysics, 57,* 352–364.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum.

Received August 20, 1997
Revision received January 14, 1998
Accepted January 14, 1998 ■