Thibadeau, R., Just, M.A. and Carpenter, P.A. (1982). A model of the time course and content of human reading. Cognitive Science, 6, 101–155.
Underwood, G., Bloomfield, R. and Clews, S. (1988). Information influences the pattern of eye fixations during sentence comprehension. Perception, 17, 267–278.
Underwood, G., Clews, S. and Everatt, J. (1990). How do readers know where to look next? Local information distributions influence eye fixations. Quarterly Journal of Experimental Psychology, 42A, 39–65.
Van Orden, G.C. (1987). A rows is a rose: Spelling, sound, and reading. Memory and Cognition, 15, 181–198.
Vitu, F. (1991). The influence of parafoveal processing and linguistic context on the optimal landing position effect. Perception and Psychophysics, 50, 58–75.
Vitu, F. and O'Regan, J.F. (1995). A challenge to current theories of eye movements in reading. In: J.M. Findlay, R. Walker and R.W. Kentridge (Eds.), Eye Movement Research: Mechanisms, Processes, and Applications. Amsterdam: North Holland, pp. 381–393.
Vitu, F., O'Regan, J.K., Inhoff, A.W. and Topolski, R. (1995). Mindless reading: Eye movement characteristics are similar in scanning letter strings and reading text. Perception and Psychophysics, 57, 352–364.
Vitu, F., O'Regan, J.K. and Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. Perception and Psychophysics, 47, 583–600.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 269-293). Oxford: Elsevier.

CHAPTER 12

Eye Movements During Scene Viewing: An Overview

John M. Henderson and Andrew Hollingworth
Michigan State University

Abstract

How do the semantic and visual characteristics of local scene regions influence the placement and duration of eye fixations during scene viewing? First, we review research on eye movement behaviour during scene viewing, focusing particularly on the influence of semantic information on eye movement behaviour. Second, we identify a number of factors that may influence eye movement behaviour in scenes, and suggest directions for future research. Finally, we propose a descriptive model of eye movement control in complex scenes.

Eye Guidance in Reading and Scene Perception/G. Underwood (Editor)

Overview

In this chapter our goal is to provide an overview of eye movement patterns during scene viewing. There are at least three important reasons to understand eye movements in scene viewing. First, eye movements are critical for the efficient and timely acquisition of visual information during complex visual-cognitive tasks, and the manner in which eye movements are controlled to service information acquisition is a critical question. More generally, the interaction between vision, cognition, and eye movement control can be seen as a scientifically tractable testing ground for theories of the interaction between input, central, and output systems (Henderson, 1996). The vast majority of our current knowledge of eye movement control in complex visual-cognitive tasks derives from studies of reading, but a complete theory will require generalization to other ecologically valid tasks like scene viewing. Second, how we acquire, represent, and store information about the visual environment is a critical question in the study of perception and cognition. The tradition in the study of scene perception (and in perception and visual cognition generally) has been to study performance in tasks that use static, briefly presented images as stimuli. However, vision is a dynamic process in which representations are built up over time from multiple eye fixations. The study of eye movement patterns during scene viewing contributes to an understanding of how information in the visual environment is dynamically acquired and represented. Finally, eye movement data provide an unobtrusive, online measure of visual and cognitive information processing. In order to capitalize on this measure, it will be necessary to develop a more complete understanding of the manner in which visual-cognitive processing is reflected by eye movement behaviour.

This chapter is divided into three sections. First, we briefly review the literature on eye movement behaviour during scene viewing, with particular emphasis on where the eyes tend to fixate in a scene, and how long they tend to stay at a particular location. Our focus here is on static scenes. Reports of recent investigations of eye movements during the viewing of dynamic scenes can be found in Chapters 17–19. Second, we identify some largely unexplored factors that may affect the placement and duration of eye fixations in a scene. Finally, we offer a tentative descriptive model of eye movement control during scene viewing.

Review of eye movements during scene viewing

Eye movement behaviour during scene viewing can be divided into two relatively discrete temporal phases, *fixations*, or periods of time when the point of regard is relatively (though not perfectly) still, and *saccades*, or periods of time when the eyes are rotating at a relatively rapid rate to reorient the point of regard from one spatial

position to another. Useful pattern information is acquired during the fixations, with little useful pattern information taken in during the saccades due to a combination of visual masking and central suppression (Matin, 1974).

During fixations, the quality of the information acquired falls off rapidly and continuously from the center of the point of regard (fixation position) due to the optical properties of the eyes and the neural structure of the retina and visual cortex, with the highest quality visual information acquired from the spatial area immediately surrounding that point. Two important issues for understanding eye movement control during scene viewing are *where* the fixation position tends to be centered during scene viewing, and *how long* the fixation position tends to remain centered at a particular location in a scene. We will address these issues of fixation position and fixation duration next.

Where do viewers look in a scene?

Effects of general region informativeness on fixation position

The first systematic exploration of fixation positions in scenes was reported by Buswell (1935), who asked 200 participants to look at 55 pictures of different types of artwork under a variety of viewing instructions. An important result was that fixation positions were found to be 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 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 the information in the scene, and by extension, to perceptual and cognitive processing of the scene. Buswell concluded that "Eye movements are unconscious adjustments to the demands of attention during a visual experience. The underlying assumption in this study is that in a visual experience the center of fixation of the eyes is the center of attention at a given time." (Buswell, 1935, pp. 9–10).

Buswell's finding that informative scene regions tend to receive more fixations has been replicated many times. In the first study to explore this relationship analytically, Mackworth and Morandi (1967) divided each of two colour photographs into 64 square regions, and a group of participants then rated the informativeness of each region based on how easy it would be to recognize on another occasion. A new group of viewers then examined the pictures to decide which one of the two they preferred. Fixation density (the number of discrete fixations) in each of the 64 regions in each scene was found to be related to the informativeness rating of the region, with regions rated more informative receiving more fixations. Regions that received low informativeness ratings were often not fixated 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. Mackworth

and Morandi (1967) also found that viewers were as likely to place their fixations on informative regions in the first two seconds of scene viewing as in other two-second intervals, providing evidence for relatively early, peripherally-based scene analysis.

The two pictures used by Mackworth and Morandi (1967) were visually simple: One depicted a pair of eyes within a hooded mask, and the other was a coastal map. Using images of more complex scenes taken predominantly from the Thematic Apperception Test, Antes (1974) provided additional evidence that region informativeness affects fixation position. Like Mackworth and Morandi (1967), Antes (1974) asked one group of viewers to rate each scene region according to the degree to which it contributed to the total amount of information conveyed by the whole picture. A different group of viewers then examined the scenes while their eye movements were recorded. Their task was to decide which scene they preferred. There were two main results relevant to fixation position. First, the density of fixations in a scene region was highly correlated with that region's informativeness, with regions rated more informative receiving more fixations, replicating Mackworth and Morandi (1967). Second, the first fixation position selected by a viewer (following the experimenter-induced initial fixation position at the center of the scene) tended to be within an informative region of a scene, suggesting rapid control of fixation position by scene characteristics.

In summary, the studies reviewed in this section suggest that the positions of individual fixations in a scene, including the position of the fixation after the first saccade, are determined in part by the informativeness of scene regions, with more fixations being directed to more informative regions. However, because region informativeness was determined by experimenter intuition (Buswell, 1935; Yarbus, 1967) or by viewer ratings (Antes, 1974; Mackworth and Morandi, 1967), visual and semantic informativeness were probably correlated in these studies. Therefore, it is not possible to determine whetner there is an independent effect of semantic informativeness (i.e., the meaning of a region) beyond visual informativeness (i.e., the presence of discontinuity in texture, colour, luminance, and depth) on the positions of fixations in a scene. This issue is important because it is related to the question of whether fixation positions reflect cognitive operations as well as perceptual processes during scene viewing. If so, then semantically informative regions should be more likely to receive fixations during scene viewing, holding visual informativeness constant. We turn to this issue next.

Effects of semantic informativeness on initial fixation positions

In perhaps the first study to investigate the influence of semantic informativeness on fixation location, Loftus and Mackworth (1978) presented viewers with line drawings of scenes in which a manipulated target object was either high or low in semantic informativeness. Semantic informativeness was defined as the degree to

which an object was predictable within the scene, with the logic that an object unlikely to be found in a scene is more informative than an object likely to be found there. Importantly, visual informativeness was controlled by exchanging objects across scenes. For example, a farm scene could contain either a tractor (low informativeness) or an octopus (high informativeness). An underwater scene contained the same two objects, so that the semantic informativeness of the target objects was reversed. The two target objects occupied the same position in each scene. Participants viewed the scenes for four seconds each in preparation for a later recognition test. There were two main findings with respect to fixation location. First, viewers tended to fixate the inconsistent objects earlier during the course of scene viewing. Second, and more interestingly, viewers were more likely to fixate the semantically informative objects immediately following the first saccade within the scene. Because the distance of the saccade to the target objects averaged 6.5–8° of visual angle, these data suggest that viewer's could determine in a single fixation the semantic informativeness of an object based on peripheral information, and that semantic informativeness could then exert an immediate effect on eye movement control.

De Graef, Christiaens and d'Ydewalle (1990) investigated the influence of semantic informativeness on eye movement patterns during scene viewing using a visual search task: Viewers searched line drawings of scenes for object-like figures that were not associated with any identifiable real-world object ("non-objects"). Using the same manipulation as had Loftus and Mackworth (1978), pre-specified target objects were placed in the scenes, and these objects were either semantically consistent or inconsistent (referred to by De Graef et al. as probability violations) with the scene. (Other types of violations were used as well, but we will focus on the semantic consistency here.) In contrast to Loftus and Mackworth (1978), De Graef et al. (1990) found no evidence that semantically inconsistent objects were fixated earlier than consistent objects. In fact, when De Graef et al. (1990) plotted the cumulative proportion of targets fixated as a function of informativeness, they found that viewers were no more likely to fixate the inconsistent than the consistent objects for the first 8 fixations. Our examination of this cumulative probability distribution (De Graef et al., 1990, Fig. 2) suggests to us that after the first 8 fixations in a scene, there was even some tendency for viewers to fixate the consistent objects sooner than the inconsistent objects. Clearly, these data do not support the view that the eyes are immediately drawn to semantically informative objects.

We recently conducted two new experiments to provide additional evidence concerning the role of semantic informativeness on eye movement patterns during scene perception (Henderson, Weeks and Hollingworth, 1999). In the first experiment, we attempted to replicate and extend Loftus and Mackworth (1978). We constructed 24 line drawings of real-world scenes generated from photographs (De

Graef et al., 1990). Semantically uninformative (consistent) target objects were drawn independently for each scene. Pairs of objects were then inserted into two yoked scenes to create two scenes in which the objects were informative and two in which the objects were uninformative, as shown in Fig. 1. The two target objects in a pair were always placed in the same location in a given scene so that the distance from the initial fixation point and lateral masking from surrounding contours would be controlled. During the experiment, we asked viewers to look at the scenes in preparation for a later memory test (which was, in fact, never given). The viewers were shown each of the 24 scenes once, half containing the informative target object for that scene and half containing the uninformative target object. Whether a given scene contained the informative or uninformative object was counterbalanced across viewers.

In contrast to Loftus and Mackworth (1978) but similar to De Graef et al. (1990), we found that viewers were no more likely to fixate the more informative target object than the less informative object early during scene viewing. First, viewers were no more likely to fixate the informative than the uninformative target after the first saccade in the scene, fixating the target immediately on about 10% of the trials in both conditions. Viewers were also no more likely to fixate the informative target after two saccades, fixating both types of objects after the first or second saccade in about 20% of the trials. Second, viewers initially landed on a target object after an average of about 11 fixations in the scene regardless of the semantic informativeness of the object. Third, the magnitude of the initial saccade to the target object was about 3°, and there was no evidence that these saccades were longer to the informative targets. These data suggest that the eyes are not initially driven by peripheral semantic analysis of individual objects.

In a second experiment, we introduced a visual search task to provide additional evidence concerning the relationship between semantic informativeness and initial fixation placement. During each trial, viewers were provided the name of a target object and then shown a line drawing of a scene. The viewer's task was to determine as quickly as possible whether the target object was present in the scene. Because of the instructions, viewers should have been highly motivated to find the targets as quickly as possible. If semantically informative objects can draw the eyes from peripheral regions of the scene, informative objects should be found more quickly than uninformative objects. As in the first experiment, however, viewers were no more likely to fixate the informative than the uninformative target after the first saccade in the scene. Instead, uninformative targets were fixated sooner (after about 3 fixations) than informative targets (after about 3.5 fixations). This finding presumably resulted from the fact that the positions of the uninformative objects were more constrained by the scenes, and so they were easier to find. For example, a blender in a kitchen is likely to appear on a counter-top rather than on the floor or elsewhere in the scene. A blender in a farmyard, by comparison, might appear just about

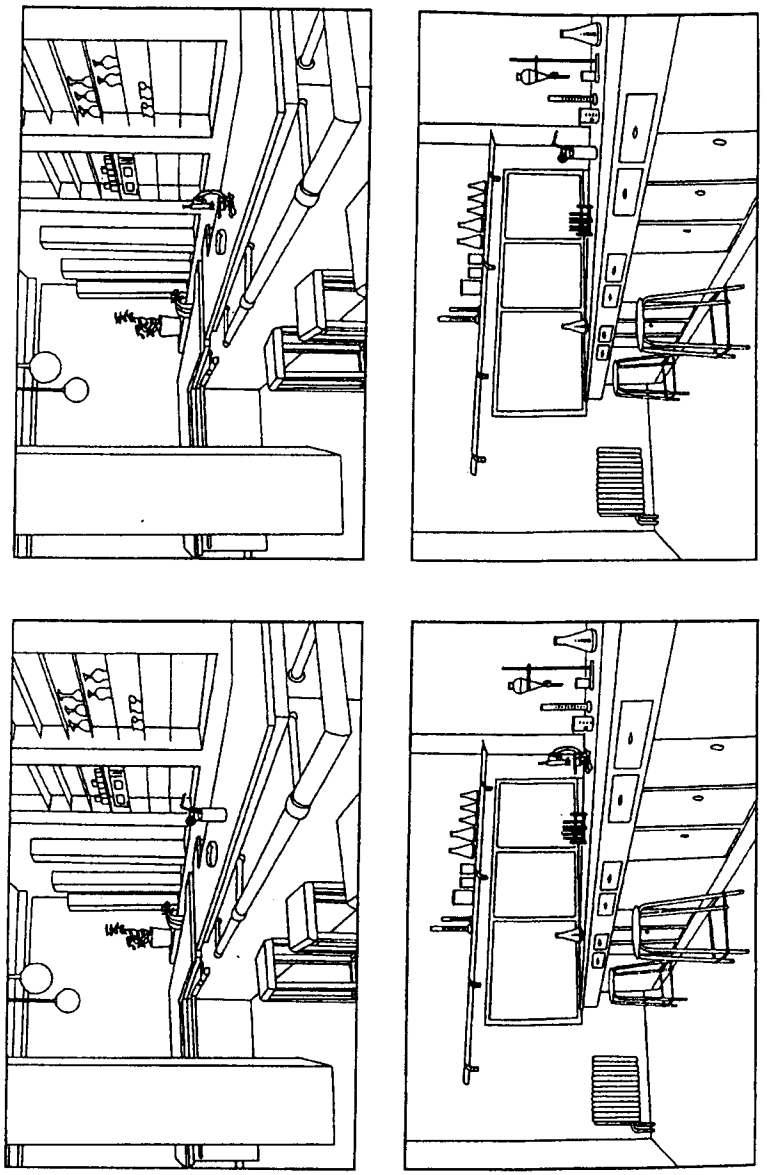


Fig. 1. Pairs of objects inserted into two yoked scenes (bar and laboratory) to create two scenes in which the objects were informative and two scenes in which the objects were uninformative.

anywhere, and would thus be more difficult to find. Finally, it is of interest that viewers moved their eyes to the targets more quickly in the second experiment (after about 3 fixations) than in the first (after about 11 fixations), suggesting that they could use peripheral visual information to guide their search. Even so, there was no evidence in either experiment that the eyes were drawn to semantically informative objects.

In summary, four experiments have examined the effects of semantic informativeness on initial fixation placement. Of these, one experiment has shown that the eyes are drawn to inconsistent object (Loftuś and Mackworth, 1978), while three have shown that they are not (one experiment reported by De Graef et al., 1990, and two experiments reported by Henderson et al., 1999). Why might Loftus and Mackworth (1978) have found that viewers' initial fixations were drawn to semantically informative objects? One possible explanation is simply that the Loftus and Mackworth (1978) result was due to statistical error. This explanation seems possible given the relatively low spatial and temporal resolution of the eyetracking equipment that was available at the time of that study.

If we assume that the Loftus and Mackworth result was not due to statistical error, there are at least three other potential explanations for the inconsistency across studies. First, semantic informativeness and visual informativeness may have been correlated in the Loftus and Mackworth experiment (De Graef et al., 1990; Rayner and Pollatsek, 1992). This problem might arise if, for example, the consistent target objects were initially drawn in the scenes, and then the target objects were swapped across scenes. If this were true, then the result of semantic informativeness on initial fixations may actually have been due to visual factors. While we cannot say for certain whether this was a problem in the Loftus and Mackworth experiment, we do know that it was not a problem in our study: All scenes were created in the same way, and target objects were drawn independently of the scene backgrounds. Second, it could be that the scenes used by Henderson et al. (1999) and by De Graef et al. (1990) were visually more complex than those used by Loftus and Mackworth (1978). For example, if there were fewer contours in the Loftus and Mackworth scenes, then there may have been less lateral masking of individual objects, and so it may have been easier for viewers to semantically analyse peripheral objects. A third and related possibility is that the difference in results across studies may have been due to a difference in the size of the scenes used across studies. Larger scenes might lead to greater peripheral semantic analysis because the objects in the scenes would potentially be larger. In our study, the scenes subtended 10×14.5° while the Loftus and Mackworth (1978) scenes subtended 20×30°. Contrary to this hypothesis, however, De Graef et al. (1990) used scenes that subtended 20×30°, but as discussed above, they observed no influence of peripheral object semantics on early fixation placement.

There is an additional point that leads us to believe that the Loftus and Mackworth (1978) result was anomalous. Loftus and Mackworth (1978) observed an average saccadic amplitude of over 7° in their study. This average is roughly twice as large as the average saccadic amplitude typically observed in scene viewing experiments. For example, viewers in both of our experiments moved their eyes to the target objects from about 3–4° away, and very few saccades were in the 6–8° range (Henderson et al., 1999). (We report further evidence concerning distributions of saccadic amplitudes during scene viewing below.) The smaller saccadic amplitudes observed in our study were not due to the size of our scenes. Antes (1974) presented scenes that subtended 20×20° and observed average saccadic amplitudes in the same range as we did. Saida and Ikeda (1979) had participants view 14.4×18.8° pictures in preparation for a later memory test. In their control condition in which the entire scene was visible throughout the trial, the modal saccade length was under 2° and very few saccades were greater than 4°. Shiori and Ikeda (1989) reported that the median saccade size in a non-degraded viewing condition of their study was about 3° in 15×15° pictures, with 75% of all saccades between about 1.5 and 5.5° (estimated from Shiori and Ikeda, 1989, Fig. 10). Van Diepen, De Graef and d'Ydewalle (1995) found average saccadic amplitudes of about 3.4° when viewers searched for "non-objects" in line drawings of scenes that subtended 16×12°. (We ignore here conditions in the Saida and Ikeda (1979), Shiori and Ikeda (1989), and van Diepen et al. (1995) studies in which the amount of the scene that was visible during each fixation was manipulated using a window or mask that moved contingent on eye position; see Chapter 15 for information on these manipulations.) Overall, then, the saccadic amplitudes observed by Loftus and Mackworth (1978) appear to be anomalous given the remainder of the picture viewing literature.

Effects of semantic informativeness on fixation density

Fixation density can be defined as the number of discrete fixations within a given region. As reviewed above, viewers tend to cluster their fixations within informative regions of a scene (Antes, 1974; Buswell, 1935; Mackworth and Morandi, 1967; Yarbus, 1967). An examination of the figures presented by Buswell (1935) and Yarbus (1967) suggest that these clusters are not entirely determined by visual factors, but instead that viewers tend to concentrate their fixations on regions that are semantically interesting. Other evidence for an influence of scene semantics on fixation density comes from the manipulation of viewing instructions by Yarbus (1967). Yarbus found that when looking at a picture of I.E. Repin's *An Unexpected Visitor*, viewers tended to concentrate their fixations on the people in the picture and particularly on their faces when they were attempting to determine the ages of the people, but tended to distribute their fixations more widely over the scene when they were attempting to estimate the material circumstances of the family.

Fixation densities in a scene region can be influenced both by the number of fixations made within that region each time it is examined (including the first time), and by the number of times viewers look back to that region. The figures presented by Buswell and Yarbus provide some qualitative evidence that both the number of initial fixations and the number of looks back to a scene region are affected by the informativeness of the region. There is also quantitative evidence supporting these conclusions. First, we have shown that the number of fixations viewers make in a region when that region is first fixated is affected by scene semantics (Henderson et al., 1999). In addition, there are two studies that provide quantitative evidence that viewers tend to return their gaze to semantically informative regions over the course of scene viewing (Loftus and Mackworth, 1978; Henderson et al., 1999). In our study, we found that viewers looked to informative objects about 3.3 times and to uninformative objects about 2.6 times on average over the course of 15 seconds of scene viewing.

In contrast to the results reported by Loftus and Mackworth (1978) and Henderson et al. (1999), Friedman (1979) found no effect of informativeness (likelihood) on the number of discrete looks to an object from a position beyond that object (Friedman and Liebelt, 1981). In that study, Friedman (1979) used a correlational approach to investigate the relationship between semantic consistency and eye movement patterns. Participants viewed line drawings of real-world scenes in preparation for a memory test in which "they would have to later be able to distinguish between the original pictures and new pictures in which, for example, only a small detail on one object would be different." Each scene contained objects that had been rated for their likelihood within the scene by a separate group of participants. A likely explanation for the lack of effect of semantic informativeness in the Friedman (1979) study is that the overall manipulation of informativeness was relatively weak; objects ranged continuously from very likely to somewhat likely in the scenes, with no truly unlikely objects. In our study (Henderson et al., 1999) as well as that of Loftus and Mackworth (1978), when a scene contained a semantically inconsistent object, that object was highly anomalous in the scene. Thus, the effect of semantic informativeness on fixation density was probably easier to detect in these latter studies.

Summary

In summary, the results of the past scene viewing studies indicate that the positions of fixations within a scene are non-random, with fixations clustering on informative scene regions (Antes, 1974; Buswell, 1935; Henderson et al., 1999; Mackworth and Morandi, 1967; Yarbus, 1967). However, the specific effect of semantic informativeness beyond that of visual informativeness on fixation position is less clear. Loftus and Mackworth (1978) observed that viewers tended immediately to fixate semantically informative objects, but neither De Graef et al. (1990) nor Henderson et al. (1999) were able to replicate this effect. At the same time, both Loftus and Mackworth (1978) and Henderson et al. (1999) observed that viewers tended to look back more often to semantically informative than to uninformative scene regions, while Friedman (1979) did not observe this effect.

How long do viewers look at different scene regions?

While initial studies of eye movement patterns during scene viewing did not report viewing time measures (Antes, 1974; Buswell, 1935; Mackworth and Morandi, 1967; Yarbus, 1967), later research provides good evidence that the amount of time viewers fixate a scene region is dependent on the informativeness of that region. At a macro level of analysis, the *total time* that a region is fixated in the course of scene viewing (the sum of the durations of all fixations in that region) is correlated with the number of fixations in that region. Because, as discussed in the preceding section, fixation density is higher for visually and semantically informative scene regions, total viewing time spent on those regions also tends to be longer.

At a micro level of analysis, one can ask whether the durations of individual fixations and temporally contiguous clusters of fixations in a region (rather than the sum of all fixations) are also affected by region informativeness. Several commonly used micro-level measures of fixation time include *first fixation duration* (the duration of the initial fixation in a region), *first pass gaze duration* (the sum of all fixations from first entry to first exit in a region), and *second pass gaze duration* (the sum of all fixations from second entry to second exit in a region). In a recent series of experiments, van Diepen and colleagues have manipulated the quality of the visual information available during each fixation using a moving mask paradigm (see Chapter 15). Viewers searched for non-objects in real-world scenes, and the image at fixation was normal or was degraded. Image degradation was manipulated by reducing the contrast or overlaying a noise mask on the fixated region. When the image was degraded beginning at the onset of fixation, first fixation durations were longer than in a control condition, suggesting that the duration of the initial fixation is controlled, at least in part, by the acquisition of visual information from the fixated region. The van Diepen et al. study (Chapter 15) is the only direct exploration of the influence of visual factors on fixation duration during scene viewing that we are aware of, and there is currently no direct data concerning whether first fixation durations or gaze durations in a scene are affected by other correlates of visual informativeness such as contour density or contrast.

The effects of semantic informativeness on micro measures of fixation time during scene viewing have been studied more extensively. Loftus and Mackworth (1978) found that first pass gaze durations were longer for semantically informative (i.e., inconsistent) objects. Friedman (1979) similarly showed that first pass gaze duration on an object was correlated with the rated likelihood of that object in the

scene, with longer gaze durations on objects that were less likely to be found in a particular scene.[1] Using the non-object counting task, De Graef et al. (1990) also found that first pass gaze durations were longer for semantically inconsistent objects, though this difference appeared only in the later stages of scene viewing. Finally, Henderson et al. (1999) found that first pass gaze duration and second pass gaze duration, as well as total fixation duration, were longer for semantically informative than uninformative objects.

The influence of semantic informativeness on the duration of the very first fixation on an object is less clear. De Graef et al. (1990) found that overall, first fixation durations on an object did not differ as a function of semantic informativeness. However, when first fixation duration was examined as a function of fixation moment (whether an object was fixated during the first or second half of all the fixations on the scene within which it appeared), first fixation durations on objects that were first encountered relatively late during scene exploration (following the median number of total fixations) were shorter on semantically uninformative (consistent) objects. We have recently analysed the first fixation duration data from our study (Henderson et al., 1999). Overall, we did not observe an effect of semantic informativeness, with mean first fixation durations of 317 ms in the informative condition and 314 ms in the uninformative condition, $F < 1$. In a subsequent analysis, we used a median split to divide the data into first fixations that occurred during the first versus second half of scene exploration, and again found no effect of semantic informativeness for either fixation moment. It appears that if they exist, effects of region semantics on first fixation durations during scene viewing are fragile.

Factors that may influence eye movement patterns during scene viewing

While there has been reasonable consistency in the eye movement patterns that have been observed across scene viewing studies, there are also some notable differences, as discussed above. It is often difficult to determine the cause of these differences because there are a number of potentially important factors that vary from study to study. These factors include image size, viewing task, viewing time per scene, image content, and image type. Table 1 summarizes the values of these factors used in the studies reviewed above. These factors could each produce main effects and could also interact with each other in complex ways to influence dependent measures of eye movement behaviour such as saccadic amplitudes, fixation positions, and fixation durations.

1 We note that both Loftus and Mackworth and Friedman called their measures *duration of the first fixation*, though the measures are equivalent to what has commonly been called gaze duration. Their eyetracking equipment did not have the spatial resolution to allow these investigators to examine the true first fixation duration.

Table 1

Summary of methods for eye movement scene studies

Study	Image size	Viewing task	Viewing time per scene	Image type/content
Buswell (1935)	varied	generally choose which images are pleasing	self-paced	colour paintings and images of other works of art
Mackworth and Morandi (1967)	16×16°	decide which image preferred	10 s	colour photographs of a mask and a coastline
Yarbus (1967)	varied	varied	varied; up to 30 min	colour paintings and images of other works of art
Antes (1974)	no more than 20°	decide which image preferred	20 s	monochrome shaded drawings (mostly from TAT test)
Loftus and Mackworth (1978)	20×30°	prepare for a later recognition memory test	4 s	black and white line drawings of real-world environments
Friedman (1979)	20×30°	prepare for memory test in which a small detail of one object may have changed	30 s	black and white line drawings (with some shading) of real-world environments
Friedman and Liebelt (1981)	20×30°	prepare for memory test in which a small detail of one object may have changed	30 s	black and white line drawings (with some shading) of real- world environments
De Graef, Christiaens and d'Ydewalle (1990)	20×30°	count non-objects	8 s	black and white line drawings of real-world environments
Henderson, Weeks and Hollingworth (1999) Exp. 1	10×14.5°	prepare for memory test in which a small detail of one object may have changed	15 s	black and white line drawings of real-world environments

(*continued*)

Table 1 (*continuation*)

Study	Image size	Viewing task	Viewing time per scene	Image type/content
Henderson, Weeks and Hollingworth (1999) Exp. 2	10×14.5°	search for a pre-specified target object	until response	black and white line drawings of real-world environments
Henderson and Hollingworth (1997)	10×14.5°	prepare for memory test in which a small detail of one object may have changed	15 s	black and white line drawings of real-world environments

Image size

One example of how variation in one factor can make interpretation across studies difficult arises in the case of the effect of the semantic informativeness of a scene region on the amplitude of a saccade to that region. As discussed above, it is possible that the amount of the visual field subtended by a depicted scene affects saccadic amplitudes, and that the influence of semantics on amplitude is mediated by this factor. While our review above led us to conclude that mean saccadic amplitudes range between about 2 and 4° despite scene size when the scene subtends between 10 and 20°, there are no published studies that were designed to directly examine saccadic amplitudes as a function of scene size. It is possible that cross-experiment comparisons are misleading because other factors have not been held constant, and that saccadic amplitudes do scale with scene size. Only studies designed to directly test these possibilities will be able to answer this question.

Viewing task

Another important variable in scene viewing is the task given to the viewer. Buswell (1935) and Yarbus (1967) both presented evidence that viewers place their fixations in a scene differently depending on the viewing task. However, these studies were descriptive in that the conclusions were based on a qualitative analysis of the viewing behaviour of particular individuals on specific scenes under differing viewing instructions. In the study described above, we compared eye movement patterns in a memory preparation task and a visual search task (Henderson et al., 1999). This is, to our knowledge, the only study that has held the nature of the stimulus image constant and quantitatively examined the influence of viewing task on eye movement patterns. As discussed above, our results showed that participants

made fewer fixation in a scene prior to fixation on a particular object when they were searching for that object than when they were trying to memorize the scene. We must point out, however, that in this study the comparison of eye movement patterns across tasks was accompanied by variations in other factors. For example, different groups of viewers took part in each task, and the amount of time viewers were allowed to look at the scenes differed in the two viewing tasks, with 15 seconds of viewing time in the memory task and self-termination of the scenes in the visual search task.

Viewing time per scene

Scene viewing time is another potentially important factor in determining viewing patterns over scenes. In the studies reviewed above, scene viewing time has ranged from a minimum of 4 seconds per scene (Loftus and Mackworth, 1978) to a maximum of 30 minutes per scene (Yarbus, 1967), though one has to wonder at the patience of the viewer in the latter case. Other studies have used scene presentation durations that fall between these extremes, as shown in Table 1. In Buswell's study, scene viewing time was determined by the viewer (Buswell, 1935), and there were large individual differences in the length of time viewers wanted to look at the scenes. Thus, it is not clear what the appropriate duration for scene presentation should be. In addition, there has been some evidence that viewing patterns change over time. For example, Friedman (1979) found that the difference in gaze durations on high and low probability objects decreased from 342 ms on first entry (first pass gaze duration) to 78 ms on the third and higher re-entries. This change appeared to be due to a decrease in gaze durations on each entry for low probability objects, but not for high probability objects. Given that viewing patterns and eye movement measures may change over the course of scene viewing, these measures may also change depending on how long viewers are allowed to look at a picture. There are currently no data available on this topic.

Image content

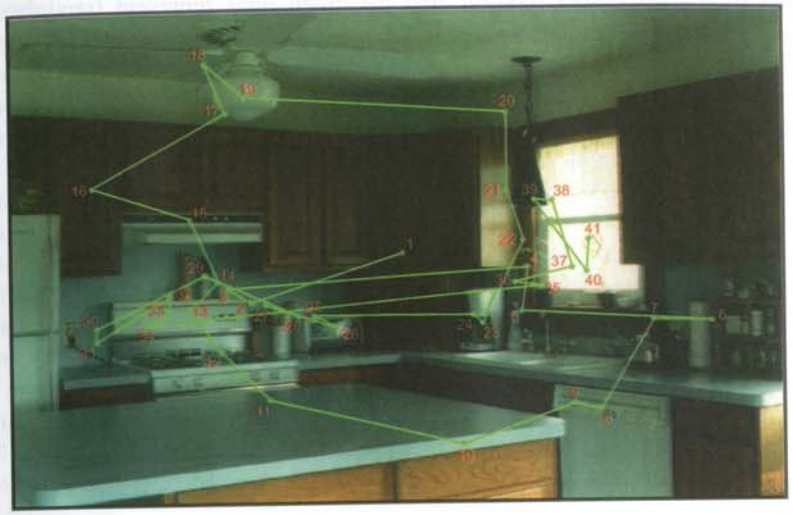
The content of the images presented to viewers in eye movement studies has varied markedly, as can be seen in Table 1. At one extreme, Buswell (1935) and Yarbus (1967) obviated the problem of image content by using a wide variety of types of scenes, while at the other extreme, Mackworth and Morandi (1967) presented only two images, one of a hooded face, and the other of an aerial view of a coastline. Both of these latter images contained large areas of uniform background. It is not clear what effect the use of such a restricted set of images has on viewers' eye movements. It may be that scenes with different content (e.g, outdoor versus indoor, large-scale spaces versus small-scale spaces) produce systematic effects on eye movement patterns. Currently, this is another unexplored issue.

Image type

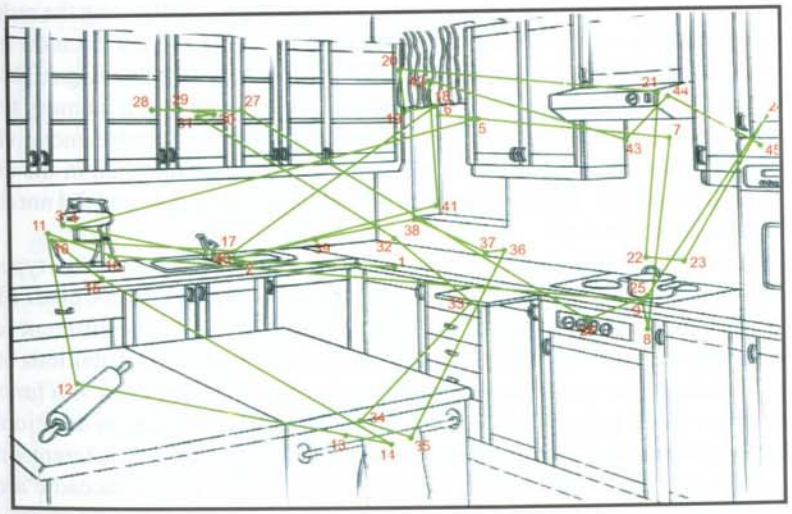
The final potentially important factor that we will discuss here is the manner in which a scene is depicted. As shown in Table 1, scene depiction has varied from line drawings (e.g., Friedman, 1979; Henderson et al., 1999; Loftus and Mackworth, 1978) to monochrome shaded drawings (Antes, 1974) to colour paintings (Buswell, 1935) and colour photographs (Mackworth and Morandi, 1967). All of the work that has so far been conducted to examine the influence of semantic informativeness on eye movement patterns has used line drawings as stimuli (De Graef et al., 1990; Friedman, 1979; Henderson et al., 1999; Loftus and Mackworth, 1978). It is not yet clear to what extent the results generated from one type of image type will generalize to other image types. Furthermore, it will ultimately be important to determine whether the results that are derived from images that depict real-world scenes generalize to the visual world itself, that is, to the situation in which the viewer is looking at the actual visual environment. The introduction of viable head-mounted eyetracking equipment in the last few years should help to encourage the exploration of this latter issue.

In order to begin to get a feel for the influence of image type on eye movement patterns, we recently conducted a study in which we contrasted viewing behaviour on line drawings, colour photographs, and computer-rendered 3-D colour images of real-world scenes (Henderson and Hollingworth, 1997). Eight viewers examined each of 30 scenes for 15 seconds each. The viewing instructions were the same as those used by Henderson et al. (1999): Viewers were told that after they had viewed all of the scenes, they would be given a memory test in which they would have to discriminate the test scenes from new scenes in which only a small detail of a single object might be changed. Ten exemplars of each of three image types (colour photographs, line drawings, and computer-rendered 2-D images from 3-D models) were presented. All images depicted common real-world scenes. The colour photographs and line drawings depicted the same 10 categories of scenes, while the rendered images depicted rooms in a house. All of the images were viewed by the same set of 8 participants. The images were presented in a random order determined individually for each participant, so that participants would be less likely to develop different viewing strategies for different image types. The three image types were presented on the same SVGA display system at the same resolution (800×600 pixels) and visual angle (10×14.5°). Eye movement data were collected using a Fourward Technologies Generation 5.5 dual-Purkinje image eyetracker. Further details of our general method can be found in Henderson et al. (1999).

The eye movement data were analysed using analysis software developed in our laboratory (see Henderson et al., 1999). Colour Plate 1a shows the eye movement pattern of one viewer on a colour photograph of a kitchen, and Colour Plate 1b shows the pattern for that same viewer on a line drawing depicting a similar scene. In the figure, the green dots represent fixations, the red numbers indicate the ordinal



Colour Plate 1a. Viewing pattern for one participant viewing a photograph of a kitchen. Green dots represent discrete fixations, ordinally numbered in red. Green lines represent saccadic vectors.



Colour Plate 1b. Viewing pattern for the same participant viewing a line drawing of a kitchen. Green dots represent discrete fixations, ordinally numbered in red. Green lines represent saccadic vectors.

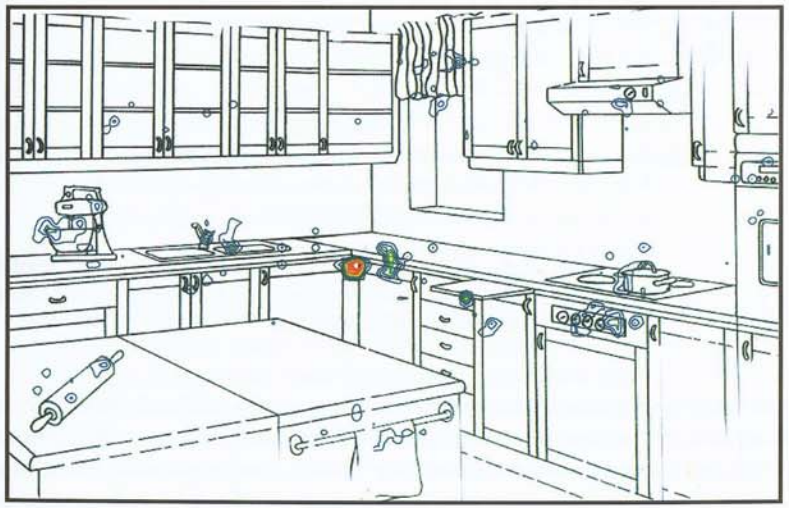
number of the associated fixation, and the green lines represent (straightened) saccade vectors. As can be seen in the figure, this viewer tended to distribute her fixations over a relatively large area of the scene, with more fixations concentrated on the more distant counter top where there were many objects than on the closer but empty counter top. Colour Plate 2 presents a contour plot of total fixation time summed across the eight viewers for the kitchen photograph (Plate 2a) and line drawing (Plate 2b). In this figure, cooler colours represent less total fixation time, and hotter colours represent more fixation time, with the colours ordered from dark blue through bright red. This figure illustrates that the viewers as a group spent the majority of their time fixating the informative regions of the scenes. A particularly striking example of this tendency can be seen by comparing the total fixation times (Colour Plate 2) on the closer counter top for the photograph and line drawing. In the colour photograph where the close counter was empty, very little fixation time was spent in that region of the scene. In contrast, in the line drawing the close counter contained a rolling pin, and fixation time clearly was devoted to that object. This contrast points out very nicely how fixation time is directed to informative scene regions.

In a quantitative analysis of these data, we found small but reliable differences in eye movement parameters as a function of image type. First, viewers made an average of 36.5 fixations in each scene. They tended to fixate the photographs reliably fewer times (34.8) than the line drawings (36.8) or rendered scenes (37.9). Second, the duration of each fixation was on average 327 ms. Offsetting the reduced number of fixations on the photographs, the duration of the average fixation in the scene was reliably longer for photographs (336 ms) than for line drawings (324 ms) or rendered scenes (321 ms). Given the instructions to prepare for a memory test, it is possible that fixations were longer on the photographs because more visual information was available to commit to memory in photographs than in the other, more schematic stimuli. Finally, the mean saccade length was 2.4° and did not differ as a function of image type.

Despite the small differences in eye movement parameters across image type, the consistency of the viewing patterns is quite striking, as can be seen in Colour Plates 1 and 2. Also, the specific nature of the fixations and saccades in the different scene types was very similar. For example, Fig. 2 shows the frequency distributions of the fixation durations (top panel) and saccadic amplitudes (bottom panel) as a function of image type for all participants. As can be seen in the figure, fixation duration and saccadic amplitude distributions were remarkably similar for the different image types, with modal fixations durations of about 220 ms and modal saccadic amplitudes of about 0.5°.

For the purposes of comparison, we have also plotted the data from a reading study in Fig. 2. In this study, conducted by Fernanda Ferreira and Melissa Johnson, 36 participants each read 20 paragraphs of text for comprehension. The text was

Colour Plate 2a. Contour plot of fixation times summed across eight viewers for a photograph of a kitchen. The scene was originally presented in colour. On this image, cooler colours represent less total fixation time, and hotter colours represent more fixation time, with the colours



Colour Plate 2b. Contour plot of fixation times summed across eight viewers for a line drawing of a kitchen. Cooler colours represent less total fixation time, and hotter colours represent more fixation time, with the colours ordered: dark blue, light blue, dark green, light green, yellow, red.

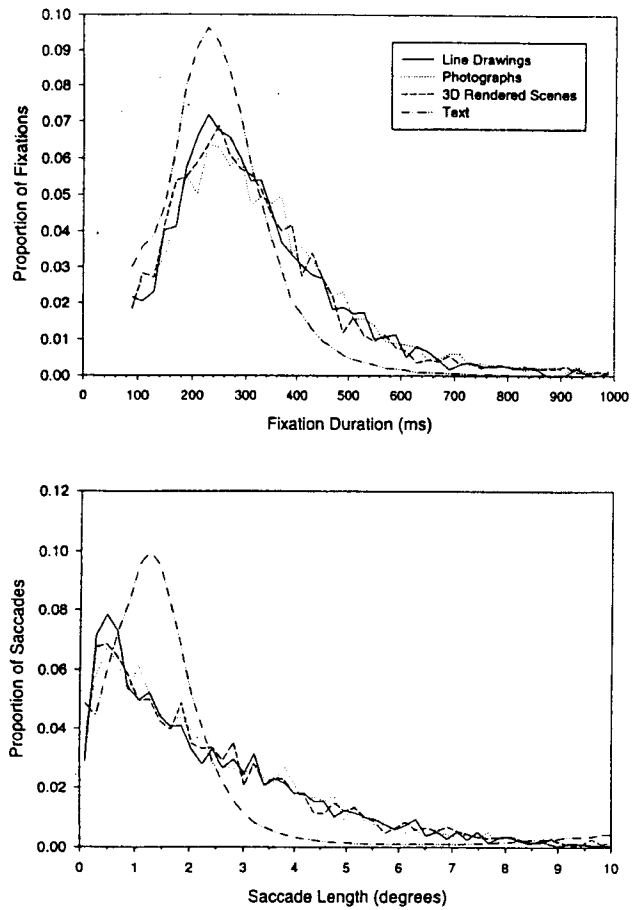


Fig. 2. Frequency distributions of the fixation durations (top panel) and saccadic amplitudes (bottom panel) for participants viewing scenes (line drawings, colour photographs, and colour 3D renderings), and reading text.

presented in graphics mode on the same display system, and subtended the same visual angle as the scenes we used as stimuli. Importantly, the eye movement data were collected using the same eyetracking system, and were analysed using the same software parameters for determining the onset of a fixation and saccade as we used in the scene study. To our knowledge, this is the first report of a direct comparison of eye movement behaviour in reading and image viewing. As can be seen in the top panel of Fig. 2, the modal fixation duration in reading and scene

viewing was the same. However, fixation duration was considerably less variable for reading, with fewer fixations lasting longer than 340 ms. The greater number of longer fixations in scene viewing appears to account for the common finding that mean fixation duration is longer in scene viewing than in reading. The bottom panel of Fig. 2 shows that the modal saccadic amplitude is longer but less variable in reading than in scene viewing. A generalization that can be extracted from Fig. 2 is that both fixations and saccades are more variable in scene viewing than in reading. Of course, these comparisons must be viewed with some caution; a task that emphasizes memory (as used in the scenes viewing study) may differ in important ways from a task that emphasizes comprehension (as used in the reading study).

A saliency map framework for eye movement control in scene viewing

In this final section we want to outline a model of eye movement control in scene viewing (Henderson, 1992; Henderson et al., 1999). This framework is, at this point, descriptive rather than quantitative, but we believe that it is specific enough to generate new predictions. The framework is also couched in such a way that it could be computationally modelled, and several of the proposed components have been modelled as independent modules for other purposes (e.g., Mahoney and Ullman, 1988). The framework is also in the spirit of the computational model proposed by Reichle et al. (1997; see also Chapter 11). The saliency map framework expands on ideas originally discussed by Henderson (1992), which in turn extended the model of eye movement control in reading proposed by Morrison (1984) and elaborated upon by Henderson and Ferreira (1990; Henderson and Ferreira, 1993) and Rayner and Pollatsek (1989; see Chapter 11). The framework is meant to account for fixation placement and fixation duration for those fixations that are directed in the service of visual analysis and cognitive processing. The framework is not meant to account for more fine-grained eye movement behaviour such as micro-saccades and ocular drift, and also ignores other oculomotor phenomena like the global effect (Findlay, 1982) and the optimal viewing position effect (O'Regan, 1992; Vitu et al., 1995), though of course we do not deny their existence.

In the saliency map framework, a representation of potential saccade targets is generated from an early parse of the scene into regions of potential interest and a background that is relatively undifferentiated. This initial parse is derived from a fast early analysis of the low frequency information available during an initial fixation in the scene. The positions of the regions of potential interest are coded in a representation of visual space and are assigned a saliency weight. The combination of spatial position and saliency weight is the saliency map (Mahoney and Ullman, 1988). Initially, region salience is determined by visual factors such as luminance, contrast, texture, colour, contour density, and so on, because this is the only information that is available about each region. Salience may also initially be

modified by top-down factors such as the viewer's task, but only if the task can be based on these visual factors. For example, the salience of scene regions that are the same shape as a search target (e.g., rectangular) would be increased, leading to relatively efficient search, as found by Henderson et al. (1999). Further, a global semantic analysis of the scene could contribute to search by constraining the likely position of semantically consistent targets.

According to the saliency map framework, the visual information acquisition system follows two simple rules: (1) Allocate visual-spatial attention to the scene region with the highest saliency weight (Koch and Ullman, 1985), and (2) Try to keep the eyes fixated on the attended scene region (Henderson, 1992; Henderson and Ferreira, 1990; Henderson and Ferreira, 1993). Because initially saliency weights are determined only by visual factors, initial attention allocation and initial fixation placement will be determined by visual rather than semantic characteristics of the scene. When the eyes are in fixation, the amount of time they remain stationary will primarily be determined by the amount of time needed to complete perceptual and cognitive analysis of that region. Once processing is complete, the saliency weight for that region will be reduced and attention will be released. Attention is then reallocated to the region that now has the highest saliency weight, and the eyes are programmed to move to that region (Henderson, 1992; Henderson, Pollatsek and Rayner, 1989). If perceptual and/or cognitive analysis of the currently fixated region is taking too long given the present fixation position (i.e., the rate of information acquisition is too low), then visual-spatial attention will be reallocated within the current region to optimize information acquisition, and a refixation will be programmed to the new locus of attention (Henderson, 1993; see also McConkie, 1979, for a similar explanation of refixations in reading). Selecting a sub-region within a region is assumed to be based on constructing a saliency map at a finer scale of resolution. The reallocation of attention within a scene region accounts for the finding that scene regions that are difficult to analyse are more likely to receive refixations (Henderson et al., 1999). Refixations may also be programmed based on oculo-motor factors alone (Henderson, 1993; O'Regan, 1992).

In the saliency map framework, initial movements of the eyes during scene viewing should be controlled by stimulus features rather than by cognitive features. However, as individual scene regions are fixated and cognitively analysed, saliency weights will be modified to reflect the relative cognitive interest of those regions. In other words, we assume that the source of the saliency weight for a given scene region will change from primarily visual to primarily cognitive interest as regions are fixated and understood. As scene viewing and understanding progresses, region salience will become heavily determined by factors such as semantic informativeness. The eyes will then be more likely to be sent to regions of cognitive salience rather than drawn by regions of visual salience, leading to greater fixation density and total fixation time on semantically interesting objects and scene regions.

In contrast to initial fixation placement, the amount of time the eyes initially remain fixated in a region, and the number of initial refixations in that region, should be affected by semantic aspects of the region right from the first time the region is fixated, because these aspects of eye movement behaviour are determined by the amount of time required to complete cognitive analysis of that region. In other words, the length of time the eyes remain in a region is controlled primarily by the needs of perceptual and cognitive analysis of the region. In addition, to the extent that additional looks back to a region are needed for additional cognitive analysis of that region, fixation times during these additional looks should also be influenced by the same factors that influence initial fixation times.

Conclusion

There had been something of a hiatus in the exploration of eye movements during scene viewing following the studies that were conducted in the 1960s and '70s. Now, after 20 years of relative inactivity, there has been a resurgence of interest in this topic, as exemplified by many of the chapters in this volume. We see this renewed interest as positive and necessary: while a great deal has been learned about eye movement behaviour during scene viewing, there are still a large number of unresolved questions. Ultimately, answers to these questions will provide a more complete understanding of the interface between perception and action, will contribute to our knowledge of scene perception, and will allow eye movement monitoring to fulfill its promise as a noninvasive, on-line measure of visual-cognitive processing.

Acknowledgements

We would like to thank Fernanda Ferreira for her lively discussions of the issues raised here, and several anonymous reviewers for their comments. The work described in this chapter was supported by grants from the U.S. Army Research Office and the National Science Foundation to John M. Henderson, 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.

References

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

Buswell, G.T. (1935). How people look at pictures. Chicago: University of Chicago Press.
De Graef, P., Christiaens, D. and 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. and Liebelt, L.S. (1981). On the time course of viewing pictures with a view towards remembering. In: D.F. Fisher, R.A. Monty and J.W. Senders (Eds.), Eye movements: Cognition and Visual Perception. Hillsdale, NJ: Erlbaum.
Henderson J. (1992). Visual attention and eye movement control during reading and picture viewing. In: K. Rayner (Ed.), Eye movements and Visual Cognition. 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. (1996). Visual attention and the attention-action interface. In: K. Aikens (Ed.), Perception: Vancouver Studies in Cognitive Science (Vol V). Oxford: Oxford University Press.
Henderson, J.M. and 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 and Ferreira, F. (1993). Eye movement control in reading: Fixation measures reflect foveal but not parafoveal processing difficulty. Canadian Journal of Experimental Psychology, 47 (Special Issue), 201–221.
Henderson, J. M. and Hollingworth, A. (1997). Eye movements during viewing of line drawings, color photographs, and computer renderings of natural scenes. Unpublished data.
Henderson, J.M., Pollatsek, A. and Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. Perception and Psychophysics, 45, 196–208.
Henderson, J.M., Weeks, P.A., Jr. and Hollingworth, A. (1999). Eye movements during scene viewing: Effects of semantic consistency. Journal of Experimental Psychology: Human Perception and Performance, in press.
Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology, 4, 219–227.
Loftus, G.R. and 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. and Morandi, A.J. (1967). The gaze selects informative details within pictures. Perception and Psychophysics, 2, 547–552.
Mahoney, J.V. and Ullman, S. (1988). Image chunking defining spatial building blocks for scene analysis. In: Z. Pylyshyn (Ed.), Computational Processes in Human Vision: An Interdisciplinary Perspective. Norwood, NJ: Ablex.
Matin, E. (1974). Saccadic suppression: A review and an analysis. Psychological Bulletin, 81, 899–917.

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.
McConkie, G.W. (1979). On the role and control of eye movements in reading. In: P.A. Kolers, M.E. Wrolstad and H. Bouma (Eds.), Processing of Visible Language, Vol. 1. New York: Plenum Press, pp. 37–48.
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. New York: Springer-Verlag, pp. 333–354.
Rayner, K. and Pollatsek, A. (1989). The Psychology of Reading. Englewood Cliffs, NJ: Prentice-Hall.
Rayner, K. and Pollatsek, A. (1992). Eye movements and scene perception. Canadian Journal of Psychology, 46 (Special Issue), 342–376.
Saida, S. and Ikeda, M. (1979). Useful visual field size for pattern perception. Perception and Psychophysics, 25, 119–125.
Shiori, S. and 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. and d'Ydewalle, G. (1995). Chronometry of foveal information extraction during scene perception. In: J.M. Findlay, R. Walker and R.W. Kentridge (Eds.), Eye Movement Research: Mechanisms, Processes, and Applications. Amsterdam: Elsevier.
Vitu, F., O'Regan, J.K., Inhoff, A.W. and Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning strings and reading texts. Perception and Psychophysics, 57, 352–364.
Yarbus, A.L. (1967). Eye Movements and Vision. New York: Plenum Press.