Chapter 7

Natural scenes and activities

7.1 Introduction

In the preceding three chapters we have examined three situations in which active vision has been studied intensively. Orienting, reading and visual search are very different activities but progress has been possible in all three cases because of the constraints on the visual activity required. In the case of orienting, both visual material and task are extremely simple; in the case of reading, the text is ordered so that there is no choice to be made of saccade direction; in the case of visual search, the task is highly specified. In this chapter we consider active vision in less constrained situations.

In Chapter 1, we listed a set of investigative questions for active vision suggesting, for example, that a full understanding of active vision would entail an ability to predict when and where the eyes moved for each gaze shift. We are far from a detailed understanding at this level of eye scanning and indeed it may be an unrealistic aim for the free viewing situation. However the gaze shifts that occur during picture and scene viewing have long proved a fascinating study area. In this chapter we review the work that has been carried out, showing that the principles emerging from studies of simpler areas can still operate. We have selected studies that have either historical or theoretical importance. We have had to exclude for reasons of space any detailed discussion of work in several important applications areas, e.g. radiology (Kundel *et al.* 1978; Krupinski, 1996); driving (Mourant and Rockwell, 1972; Land and Lee, 1994; Chapman and Underwood, 1998); sports (Williams and Davids, 1998; Williams *et al.* 1999).

In view of the wide prevalence of the moving image in contemporary life, it is remarkable that very little progress has been made in investigations of how people scan such images. An early finding by Guba *et al.* (1964) found that viewers had a strong tendency to look at the face of a narrator, even when a strong animated distractor was also present. More recently Tosi *et al.* (1997) have shown that rapid motion sequences in moving pictures or objects generate a strong tendency to fixate the screen centre. Yet, in general, this topic is ripe for future investigation.

7.1.1 Early studies of picture scanning

Two classic studies of eye scanning, one American and one Russian, devoted much time to analysing records of individuals looking at pictorial material. Guy Buswell used a photographic technique to record eye movements both of reading (Buswell, 1922, 1937) and picture viewing (Buswell, 1935). In the study of pictures, he used artistic material from a variety of sources, such as the classical print ‘The Wave’ by Hokusai (Fig. 7.1). He noted wide individual differences in all his measures but also showed that eye fixations in general concentrated on particular areas of the picture. So, for example, in pictures containing human figures, these figures were disproportionately likely to receive fixations. He offered relatively little detailed interpretation of his results, suggesting indeed that such an issue might be of more interest to the art expert than the psychologist.

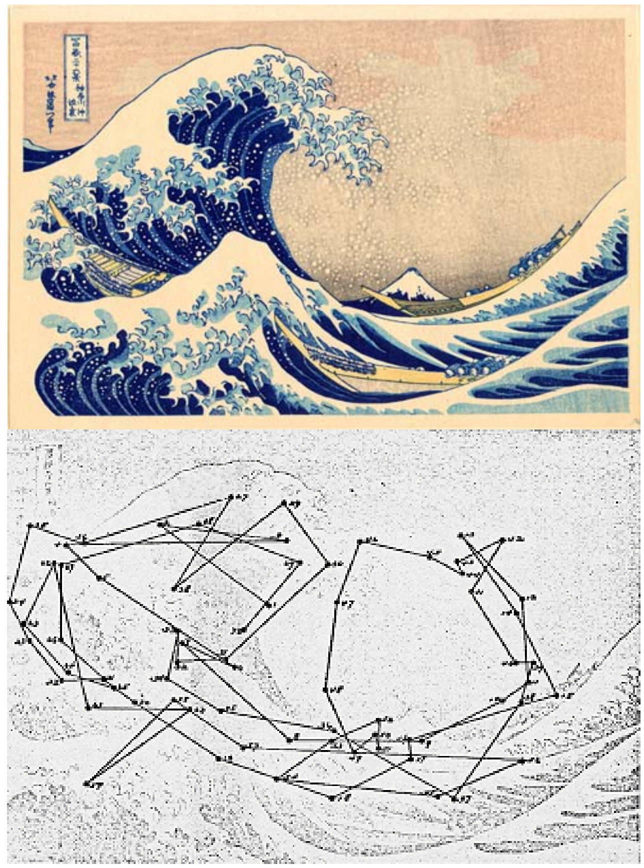


Figure 7.1 'The Wave' by Hokusai, and an eye scan of the same picture reported by Buswell in his early studies of eye scanning.

Nevertheless, some interesting generalisations emerge. Buswell made detailed measurements of fixation durations during picture scanning and noted a tendency for fixation duration to increase with increased viewing time. He suggested that a general survey, in which the viewer takes in the main areas of a picture, is followed by a more detailed study of localised regions. Buswell compared scans over coloured versus black/white versions of the same picture and reported that colour had surprisingly little effect on eye scanning. Finally he compared art students with others and found that art training resulted in a tendency to make shorter fixations.

Alfred Yarbus pursued a solitary line of investigation in Moscow at a time when the political vicissitudes of the Cold War prevented much East-West scientific interchange. His work was consequently little known in the West until an admirable translation was made by Lorrin Riggs, an American oculomotor specialist (Yarbus, 1967). Yarbus used a 'suction cap', which fitted onto the eyeball, and had a small mirror attached. A light beam reflected from a mirror attached to the cap was deviated as the eye rotated. This enabled a photographic record of the eye movements to be obtained (a similar technique was in use at the same time in the UK and the US for studying miniature eye movements (§ 2.3.2).

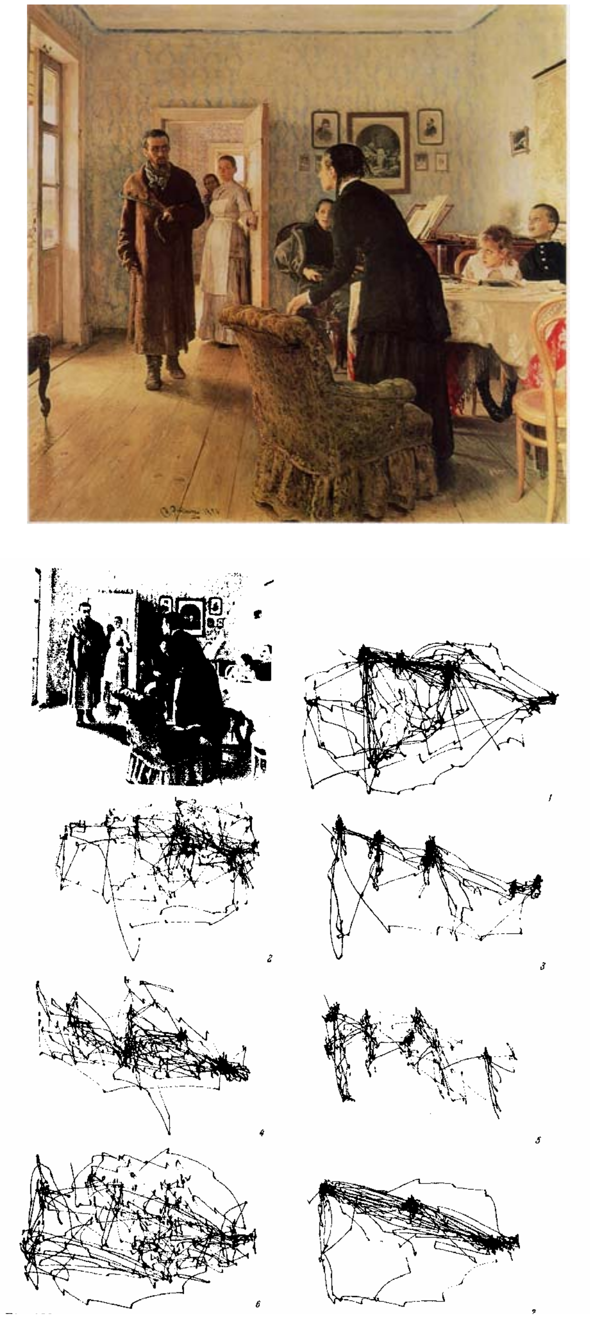


Figure 7.2 Examples of eye scanning records obtained by Yarbus (1967). Observers were given different instructions while viewing the picture 'They did not expect him' by Ilya Repin. Each of traces shows a three minute record of eye scanning with the following instructions 1) Free examination, 2) following request to estimate the material circumstances of the family, 3) following request to give the ages of the people, 4) following request to surmise what the family were doing before the visitor's arrival, 5) remember the clothes worn by the people in the picture, 6) remember the position of the people and objects in the room, 7) estimate how long the "unexpected visitor" had been away.

As well as studying basic aspects of eye movement control, Yarbus recorded the scanning eye movements of observers looking at pictures. He used a set of pictures, such as Fig. 7.2, from the Russian realist school and other sources. His fascinating records have become deservedly well known through frequent reproductions. Yarbus confirmed many of Buswell's findings, for example the minimal influence of colour on eye scanning patterns (also reported by Tosi *et al.* 1997). He appreciated that scanning patterns were very sensitive to the specific mental task of the observer, showing this clearly by comparing scans where observers viewed pictures after having been given very specific instructional questions (Fig. 7.2). In one aspect, a clear difference from Buswell's account appears. Yarbus, although at times suggesting that a scan is first used to form a 'general impression', used his data to draw the conclusion that "additional time spent on perception is not used to examine the secondary elements, but to re-examine the most important elements" (Yarbus, 1967, p. 193). It is certainly the case that the records of Yarbus show this repetitive characteristic. This is quite surprising given the records were often made over an abnormally long viewing period (many seconds).

7.1.2 Average characteristics of eye movement patterns during picture viewing.

Buswell, Yarbus and a number of subsequent workers have accumulated much data on average parameters of the eye movements made during picture scanning. These measures have been reviewed in a number of places (Rayner and Pollatsek, 1992; Henderson and Hollingsworth, 1998). Figures in the range 2-4 deg are the most commonly reported estimates of average saccade amplitude during picture viewing. It should be noted that saccade size reports are generally given in degrees visual angle. In the study of saccades during reading (§ 5.2), absolute measures of saccade size vary with the size of the print and it is much more appropriate to use a measure in terms of number of characters of text. It is very likely that the same consideration should apply in the case of scenes or pictures, since it is evident that a pictorial representation can be scaled to a wide range of sizes.

Fixation durations when viewing pictures and scenes show a skewed distribution, with a mode at 230 ms, a mean of 330 ms and a range from less than 50 ms to more than 1000 ms (Henderson and Hollingsworth, 1999a). The finding that the mean fixation duration increases as viewing continues was noted by Buswell (1935) and has been replicated several times (Viviani, 1990). For example, Antes (1974) reported that the mean duration of early fixation was 215 ms and that this increased progressively to 310 ms after a few seconds inspection.

Just as in the case of reading (Chapter 5) a variety of low level and high level factors affect fixation durations during picture viewing. Loftus (1985) showed that presenting a line drawing at a low contrast resulted an increase in mean fixation duration in comparison with that found with a normal contrast presentation, from ca. 300 ms to ca. 400 ms with an even greater effect on the first fixation (200 ms to 500 ms). Low pass filtering has a similar effect (Mannan et al. 1995). Low pass filtering is an operation carried out on a visual image which removes high spatial frequency, fine detail, information. The effect is similar to blurring. More cognitive factors likewise affect fixation duration. Fixation durations on objects that are implausible in the scene context are generally longer (§ 7.2.5).

7.1.3 Scanpaths

An idea that excited considerable interest was the 'scanpath' (Noton and Stark, 1971). Noton and Stark claimed that when a particular visual pattern is viewed, a particular sequence of eye movements is executed and furthermore that this sequence is important in accessing the visual memory for the pattern. They obtained some experimental support for their first postulate by showing reproducible scanpaths, albeit in a situation with very large and very dim displays. However there seems almost no evidence for the second postulate and several findings that are incompatible with it. Walker-Smith *et al.* (1977) recorded scanning patterns when observers carried out tasks involving facial recognition (faces were one type of item used by Noton and Stark). When an individual test face was examined to determine a match to a memorised face, some reproducible scanning sequences were observed. However when observers were asked to make a direct comparison between two simultaneously presented faces, there was no suggestion that a scanpath occurred on one face at a time. Rather a repeated scanning between the two faces, resembling a feature by feature comparison, occurred. Parker (1978) carried out a task which involved memorising a scene containing about six objects, always in identical locations. Subjects generally fixated each object in the scene in a preferred viewing order although use of a different order was not disruptive. Other work (e.g. Mannan *et al.* 1997; Hodgson *et al.* 2000) has failed to find evidence for consistent replicable scanning sequences. In spite of occasional revivals of the idea (Stark and Ellis, 1981), it is now widely accepted that the scanpath suggestion in its original strong form was of little generality.

Arising in part from interest in scanpaths, a number of workers have developed techniques to capture statistical regularities in the pattern of eye scanning. One approach has considered statistical dependencies in parameters such as saccade direction or fixation location (Ellis and Stark, 1986; Ponsoda *et al.*1995). An example of their use has been discussed in connection with visual search in Chapter 6 (Gilchrist and Harvey, 2000). The simplest form of such a sequential dependence is the Markov process in which the properties of the immediately preceding saccade constrain the probabilities of the one currently programmed. Markov analyses have been used with human scanning data (Bozkov *et al.* 1977) and have been proposed as a way of controlling robot vision scanners (Rimey and Brown, 1991). Several papers have explored the relationship between scanning statistical properties and image statistics (Mannan *et al.* 1996, 1997; Krieger *et al.* 2000). Relationships emerge but the major component appears to relate to the simple generality described in the next section.

7.1.4. The gaze selects informative details

Buswell and Yarbus both reported that eye fixations tended to fall on the important details of pictures. Mackworth and Morandi (1967) elaborated this into the formal statement that forms the title of this section. Mackworth and Morandi generated an informativeness measure for different regions of a picture. This was achieved by cutting up the picture into 64 separate segments. A separate group of 20 observers were shown these segments individually and asked to rate how informative they appeared to be using a nine-point scale. The informativeness rating obtained for a region was highly predictive of the probability that the region would be fixated when an observer viewed the picture in a preference task.

Whilst useful as a general summary of observations, there are grave difficulties in using this statement in any more detailed way. It is tempting to believe that the idea of informativeness might be the key that allows cognitive activity to be revealed from patterns of eye scanning. However, as Viviani (1990), has argued with particular force, it is simply not possible to use eye scanning patterns in this way. Viviani makes a number of critical points. Scene information is generally available at a multiplicity of spatial scales and eye gaze measures cannot reveal whether a particular scale is selected. Thus "what meets the fovea is only a part of what meets the (mind's) eye" (Viviani, 1990, p 360). Indeed, the direction of eye gaze may not necessarily even correspond to the direction of attention (Chapter 3). Even more fundamental is the fact that informativeness depends on the cognitive task being undertaken (cf. Fig. 7.2). Although uniform or uniformly repetitive areas of a scene can never convey much information, in general there is little relationship between the purely visual properties of a scene and their information content. Finally, cognitive processes may often take place in parallel whereas eye scanning is necessarily a serial operation. Although this led Viviani to dismiss many attempts to use eye scan data, he expressed more optimism about situations in which eye scanning is properly matched to an articulated theory of cognitive activity for the task in question.

7.2 Analytic studies of scene and object perception

7.2.1 Scenes and objects

Historically, the mid-1970s represented something of a watershed in studies of active vision. At this point, the gaze-contingent methodology was introduced into studies of reading (Chapter 5) and analyses such as those of the amplitude transition function were developed in visual orienting (Chapter 4). Furthermore, considerable theoretical developments occurred in the areas of object perception and scene perception. In this section we summarise these developments.

The visual world presents a *scene,* which in general contains a number of well defined and often well located *objects*. It is usually accepted that scene perception involves more global properties of the visual world than simply the collection of objects. Thus a countryside scene might consist of objects such as a farmhouse, animals, hedges, but it is clear also that configurational coherence is an essential characteristic of a scene. Object perception can be studied in isolation from scene perception but it is less clear that the converse is true. It is not in general possible to envisage a scene devoid of objects, which creates a practical problem in disentangling scene perception from object perception. The scene provides context for the objects and one question that has received extensive investigation is whether objects are more readily perceived in an appropriate scene context (§ 7.2.5).

7.2.2 Theories of object perception and scene perception

Individual objects viewed foveally can be recognised rapidly and generally within a single fixation (§ 7.2.3). The process by which this occurs has been, and is still, a matter of active and fierce debate but all contenders accept that foveal recognition of single objects is an area where the passive vision approach, i.e. massive parallel processing of retinal information, is legitimate. Only a very brief sketch of the different theories of object perception can be given here. One early approach was to consider an object as made up of a distinctive collection of *features* and to postulate that object recognition was achieved through detection of features and feature combinations. The inadequacy of this view was argued, amongst others, by Marr (1982), whose view of vision was that a set of computational processes transformed the retinal image into a *three dimensional representation*, which formed the basis for recognition. Biederman (1987) developed this idea into a more fully elaborated theory of object recognition. In his theory, objects are represented as an organised collection of *geon* primitives. Geons, or geometric icons, are elementary three-dimensional shapes that can be recognised on the basis of properties such as parallel sides that produce a viewpoint invariant two-dimensional projection.

An alternative to Biederman's theory suggests that three-dimensional representations are not primarily involved in object recognition (although they may be needed to support visual actions – § 2.2.2). The alternative suggestion is that objects are recognised on the basis of their 2D properties with a small number of stored two-dimensional 'views' of an object operating in conjunction with interpolation processes (Poggio and Edelman, 1990; Tarr and Bülthoff, 1995). Current thinking suggests that the key biological importance of recognition may have resulted in a multiplicity of visual recognition systems (Logothetis and Sheinberg, 1996). An additional complication is that, for any particular experimental task, objects may be partially recognised on the basis of a few simple features. Such partial recognition may be adequate, for example, to allow an experimental subject to show object discrimination, particularly where the set of possible objects is small and is shown repeatedly.

Scene perception has been less investigated and no theoretical position has emerged of comparable strength to those in object perception. Biederman (1981; Biederman *et al.* 1982) attempted to analyse what might be involved in the 'configurational coherence' concept which, as discussed above, differentiates a scene from a random collection of objects. He suggested, using an analogy with analyses of language material, that scenes could be regarded as *schemas*, providing a frame in which objects are viewed. He proposed several properties in which an object present in a real world scene would normally satisfy. The object would have *support*; in general objects are not free floating and thus part of the scene will be underneath the object. The object would occlude the scene background; *interposition* of background and object occurs only rarely with transparent materials. Support and interposition can be likened to the syntactic constraints in linguistic analyses. Corresponding to the semantics of language, the object would have a high *probability* of occurring within the scene frame. Cookers are unlikely to occur in a street scene. Objects would also be the appropriate *size* and have the appropriate *position* within the scene. Biederman and his group have carried out a number of experiments showing that these relationships affect the speed of perception when scenes are viewed. If scenes are presented in which the relationships of support, interposition etc are violated, then recognition is slowed.

7.2.3 Are eye movements necessary for scene and object perception?

Studies on picture memory have demonstrated that picture memory is good, even with very brief exposures. For example, Intraub (1980), using a set of 250 pictures cut from magazines, compared recognition memory after a subset had been shown. With 6 seconds viewing time, recognition rates were 94%. With tachistoscopic exposure durations, the rate fell to about 80% (and the false positive rate rose from 8% to 11%). This confirms earlier findings (Potter and Levy, 1969; Loftus, 1972) that the opportunity to inspect a picture with eye movements is beneficial, but nevertheless the memory ability for material shown in a single glimpse is still impressive. However, it is probable that the majority of the pictures used contained prominent objects viewed foveally.

Object recognition deteriorates quite rapidly in the parafovea and periphery. Nelson and Loftus (1980) allowed subjects a limited amount of time to scan a picture and they then performed a recognition test, viewing a version of the picture in which one object had been changed. Their task was to detect the changed object. Detection rate for small objects (1 deg) was above 80% for objects that had been directly fixated and higher if they received two fixations. However it fell to 70% for objects where the closest original fixation had been at a distance from the object in the range 0.5deg - 2 deg. and to a chance level (or in one case slightly above chance) for objects viewed more peripherally. This suggests that the gaze needs to be directed to within 2 deg of an object in a scene for it to be reliably encoded (see also Irwin and Zelinsky, 2002). As discussed in Chapters 2 and 4, the absolute eccentricity of a stimulus is likely to be less important than the eccentricity in relation to the object size and viewing distance. However, with rare exceptions (Saida and Ikeda, 1979) this relationship does not appear to have been investigated formally in studies of object perception. Work discussed in the next section suggests that information from objects seen in isolation can be obtained more peripherally but only when the object is a target for a forthcoming saccade. Other estimates of objects perceived in scenes (Rayner and Pollatsek, 1992) have emerged with broadly similar figures. One exceptional result (Loftus and Mackworth, 1978) suggesting that objects could be identified in more extreme peripheral locations is discussed in detail later (§ 7.2.5).

Some significant scene information that is not dependent on object identification can be acquired in a single glimpse. Schyns and Oliva (1994) presented subjects with 50 ms views of scenes that had been either high-pass or low-pass spatial frequency filtered. The low-pass filtering prevented recognition of objects in the scene but nevertheless the scene could be 'recognised' although 'recognition' in this experiment implied merely discrimination from a small number of possibilities. Schyns and Oliva argue that scene recognition normally depends on the low spatial frequency configurational information, but the system is also flexible enough to allow recognition from high-pass filtered versions (Schyns and Oliva, 1997). Recent work has examined the possibility that the rapid recognition stage may be based on statistical properties, for example related to the colour distribution (Oliva and Schyns, 2000)

Biederman (1981; Biederman *et al.* 1982) has also argued that scene information is captured within a single glimpse. He based this argument on experiments in which subjects are given a brief presentation (150 ms) of a scene and have to respond whether a prespecified object is present in the scene at a location specified by a flashed post-exposure marker. The scenes were line drawings containing 8-12 objects. Detection accuracy was about 70% (with a 20% false alarm rate) in scenes with normal structural relationships. However accuracy showed a substantial decline when the structural setting of the cued object violated the structural relationships specified in § 7.2.2. No effect was found if the violations occurred in relation to an object other than the target object (an 'innocent bystander'). His conclusion, that scene properties are processed and can affect object recognition, seems justified. However, as in the case of visual search (Chapter 6), only a limited number of objects can be processed in parallel. Further discussion of the relationship between scene perception and object perception occurs subsequently (§ 7.2.5).

Another paradigm that has proved productive in examining the amount that can be taken in from a brief presentation is that of conceptual identification. Intraub (1981) presented pictures of objects at a rapid sequential rate (113 ms/picture). The task was to identify a picture that belonged to a specified category (e.g. food). Performance was high (60%) and good performance was also obtained in the converse task with negative instructions (identify a non-food item in a sequence of food items). More recently, a similar task has been employed with more complex material (Thorpe et al. 1996; Fabre-Thorpe et al. 1998; Delorme et al. 2000). Participants in these experiments are required, for example, to identify whether or not a briefly exposed scene contained an animal. Good identification is found with exposure durations as brief as 20 ms. Moreover, by recording brain (evoked response) activity in the paradigm, it has been shown that significant differences are found as early as 150 ms in the responses evoked on target trials in comparison with those on non-target trials. A detailed analysis in relation to viewing position has not been made in these experiments, but in the example scenes shown, the target objects are in central or near central view.

The conclusions from this section are that objects viewed foveally or in the close parafovea can be identified and categorised very rapidly. However the evidence suggests that objects not fixated closer than 2 or 3 degs are not recognised and thus eye movements are in general necessary for the identification of objects within scenes. A certain amount of scene information (the *gist*) can be extracted from a single glimpse of a scene and this allows the scene schema to be evoked with very little delay. It is likely that eye movements over a scene will provide additional information, as well as information about specific objects. However at present no metric exists for scene measurement and so this proposition has not been tested.

7.2.4 Object perception in peripheral vision.

.

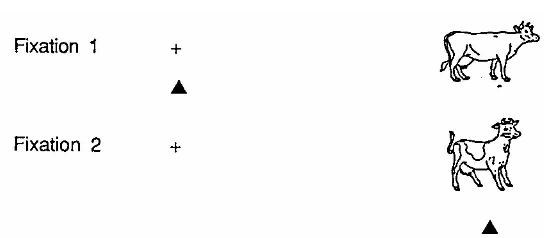


Figure 7.3 Procedure used by Pollatsek et al. (1984) to demonstrate preview benefit in picture naming. The black triangle symbolises the eye location. The picture is altered during the saccadic movement to the object.

A series of systematic studies has investigated the extraction of information about objects seen in peripheral vision using a similar *boundary technique* to that employed in the study of reading (§ 5.3.1). Recollect the logic of the technique is to display material to be identified in peripheral vision that may change as the subject's eyes make a saccadic movement to view it. The change occurs as the eye crosses an imaginary boundary and the aim of the approach is to investigate the extent to which peripheral preview assists the recognition process.

A series of studies at the University of Massachusetts, Amherst, has used the boundary technique (Fig. 7.3) to explore the *preview benefit* of pre-saccadic material in peripheral vision. Pollatsek et al. (1984) measured the speed of object naming when various forms of preview occurred. They found that full peripheral preview produced a substantial benefit (120 ms). Benefits, albeit of a reduced magnitude, also occurred if the previewed picture came from the same category as the target picture. A small benefit, also occurred when the previewed picture had the same overall visual shape as the picture (e.g. carrot > baseball bat). This latter effect occurred only with tests at 10 deg eccentricity, and not at 5 deg.

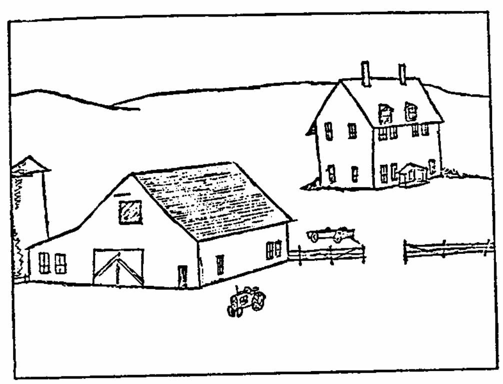
Henderson et al. (1989) ran another gaze-contingent experiment to examine a further question. Would preview information be taken in automatically from any location in the periphery, or would the information only be extracted from the location to which the saccade was directed. Displays were shown of four objects, located at the corners of an imaginary square, and the subjects were asked to scan these in a fixed order. Processing was speeded if the object could be seen in the peripheral location about to be fixated, but not if it was located elsewhere. Peripheral information can be usefully extracted from pictorial material viewed in isolation although only at the destination point of a planned saccade. Studies in visual search (Chapter 6) have also demonstrated, under some circumstances, that processing of pictorial material is possible in peripheral vision but once again with displays consisting of discrete elements.

The work discussed up to now has shown that peripheral preview benefits later foveal viewing. Henderson et al. (1997) demonstrated that objects can be recognised when only ever seen in peripheral vision. They used the foveal mask gaze-contingent technique to create a situation of an artificial foveal scotoma (§ 5.3.1). Subjects viewed a linear array of three pictorial line drawings of objects chosen from the Snodgrass and Vanderwart (1980) collection and were required to indicate whether a subsequently named item was part of the set. The objects were of overall size 1.5 – 2.0 deg and were separated by 2.4 deg. In the scotoma condition, the object to which the gaze was directed was obscured. Performance in the scotoma condition was high (85% - 93%) and only slightly below that in the normal viewing condition (95%). Henderson et al contrast this result with the disastrous decline with a foveal scotoma when reading performance is assessed (§ 4.3.2).

Attempts have been made to describe the useful field of view for picture perception using a gaze-contingent window technique in a similar manner to that used in studies of reading. Saida and Ikeda (1979) created a viewing situation with an electronic masking technique in which subjects saw only a central square region of an everyday picture. Viewing time and recognition scores were impaired unless the window was large enough for about half the display to be visible. This fraction applied both to 14 deg x 18 deg displays and to 10 deg by 10 deg ones. A follow up study by Shiori and Ikeda (1989) looked at the situation where the picture was seen normally inside the window but in reduced detail outside. Even very low-resolution detail (very well below the acuity limit at the peripheral location) aided performance considerably. Shiori and Ikeda used a pixel shifting technique to reduce resolution. Van Diepen et al. (1998) have developed a technique whereby a moving window can be combined with Fourier filtering. Somewhat surprisingly, performance is better when the degraded peripheral material is low-pass filtered than when it is high-pass filtered.

7.2.5 Scene context and object perception.

A question that has provoked intense interest is whether scene context facilitates object recognition. An early piece of work on scene perception (Biederman, 1972) found this to be the case. Biederman carried out an experiment comparing recognition speed for detecting an object when it was present in a coherent scene with that when the same scene was jumbled by cutting it into pieces that were randomly rearranged. The object's contours were preserved during the jumbling operation but the time taken to detect its presence was significantly longer. Biederman interpreted this to indicate that scene context was important in facilitating object recognition although Henderson (1992) has pointed out an alternative possible interpretation of the results which is that the additional contours introduced by the jumbling operation added noise to the situation, thus slowing detection.



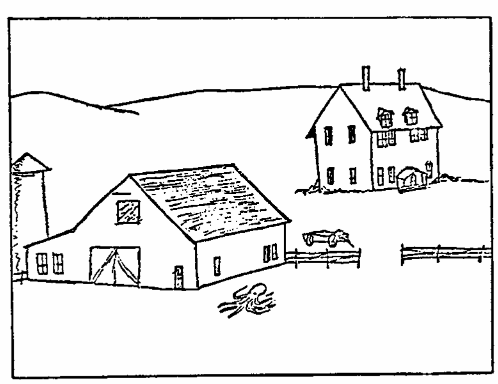


Figure 7.4 Example of pictures used in the study of Loftus and Mackworth (1978). Different individuals viewed each version of the picture with instructions to memorise. The measures of interest concerned the time at which the inconsistent object (in this case the octopus) affected the eye scanning pattern

An important subsequent study that supported this position was reported in 1978 by Loftus and Mackworth. We shall describe this study in detail because of its significance for the general theme of the book and because it has considerably shaped subsequent thinking and experimentation. Loftus and Mackworth's result implied that objects were assessed in relation to the scene context and furthermore that this assessment was carried out over a substantial part of the visual field. It thus provided support for the passive vision approach which, as we outlined in Chapter 1, may be regarded as an implicit statement about scene perception. Passive vision claims that vision achieves a representation in which three-dimensional information, adequate to support object recognition, is present over the entire visual scene.

Loftus and Mackworth aimed to determine the relative significance of high-level and low-level properties in controlling fixations on a scene (§ 7.1.4). They carried out the following experiment. Line drawings of scenes were constructed in which one object in the scene might be carefully replaced with a drawing of a different object, the replacement having broad visual similarity, but a very different contextual probability of appearing in the scene. Thus, in a farmyard picture, a tractor was replaced by an octopus of similar size and shape (Fig. 7.4), and in a stadium scene, a footballer was replaced by a ballet dancer. Control cases occurred with the tractor in the farmyard and the octopus in an underwater scene. Observers were instructed to expect a recognition test and their eye movements were recorded as they viewed the scenes. Loftus and Mackworth were interested in how quickly the incongruous object was detected, and how far in the periphery the object was at the time of detection. The estimate of the latter measure came from the size of the saccade that immediately preceded the fixation on the incongruous object. Loftus and Mackworth found that incongruous objects were looked at more often and for longer durations, that fixations subsequent to the very first fixation on the scene were more likely to fall on the incongruous object than the control context appropriate object, and that the average saccade size prior to looking at the critical object was about 7 deg.

The last two of these results imply that the scene is analysed to a very high level using peripheral vision. In some way, it appears that objects in peripheral vision were identified as being improbable in the particular scene context. This suggests that the object is implicitly recognised and furthermore related to the scene properties. The peripheral recognition abilities found in the Loftus and Mackworth study are much greater than those found in visual search experiments discussed in Chapter 6. The visual search experiments used isolated objects, rather than contextualised objects. It is thus conceivable that superior perceptual abilities occur in scenes. However, as discussed below and in detail by Henderson and Hollingworth (1998,1999a), attempts to replicate the results of Loftus and Mackworth have found much less evidence for such perceptual discriminations and have led to increasing concern over the result

Friedman and Liebelt (1981) analysed data from subjects viewing scenes in which incongruous objects might be present. Incongruity affected the duration of fixations once the objects were fixated but had no effect on the probability that an object would be fixated. This pattern of results has now been found in a number of other studies. De Graef *et al.* (1990) used a task in which subjects viewed drawings of scenes containing objects and also containing one or more constructed 'non-objects'; geometrically plausible sculpted constructions of similar size to real objects in the scene. The task was to report on the non-objects but the interest of the study was to compare fixations in a task where object-decisions (corresponding to the more familiar lexical decisions of § 5.4) were important. De Graef et al. used the same manipulation as Loftus and Mackworth to include objects that were both plausible and implausible in the scene context. De Graef et al. found *no* evidence that implausible objects were more likely than plausible ones to be fixated any stage in the viewing period. They did in part replicate the first finding of Loftus and Mackworth. Implausible objects, once fixated, held the gaze for longer. However these longer fixation durations did not appear in the early fixations (at least the first eight) of the viewing sequence.

Other work does suggest that an immediate effect of scene context can be found when an object is first fixated. Boyce and Pollatsek (1992) took advantage of the very powerful potential of visual transients to attract the eyes (Chapter 4). Subjects were presented with a scene and shortly (75 ms) after, a target object was 'wiggled' (given a brief small displacement before returning to its original position). The task was to name the object. The wiggle ensured that the subjects' eyes are directed towards the object. Boyce and Pollatsek found that the naming latency of objects presented in context was substantially shorter than that of those violating the scene context.

Henderson and Hollingworth (1998), in a review of the area, report two further experiments of their own attempting, and failing, to replicate the Loftus and Mackworth finding that contextually implausible objects attract fixations. They conclude that the weight of evidence is now against a position that suggests detailed object analyses can be carried out over a wide region of peripheral vision during scene viewing. The alternative view, that our subjective impression of an immediate pictorial reality is illusory, has received increasing support from recent studies, particularly of he phenomenon discussed in the next section.

7.2.6 Change blindness

Grimes (1996), working in the laboratory of George McConkie, reported some results where a saccade-contingent paradigm (§ 5.3.1) was employed to study picture perception. In the studies reported by Grimes, an observer views a picture and at some point whilst a saccade is in progress, aspects of the picture were changed. Participants viewed full colour pictures of scenes and were told to expect a memory test. They were also forewarned that something in the scene might occasionally change and were asked to report any such changes. Surprisingly large changes were often undetected. Thus, in a picture of two men wearing hats, swapping the hats was never detected. Movement of a child in playground scene, involving an image size change of 30% showed a detection rate of under 20%. Changing the colour of a parrot, occupying 25% of the picture space, from brilliant green to brilliant red, was occasionally (18% of occasions) not detected. Further important work with this paradigm (McConkie and Currie, 1996; Currie *et al.* 2000) will be discussed in Chapter 9.

One implication of the failure to detect changes in Grimes' experiment is that, contrary to expectation, the relevant information for detecting changes is just not available in any visual representation of the scene. If this were so, making the change simultaneous with a saccade would not be important. Of course if changes are made to a scene during a fixation, visual transients are introduced and these transients have the effect of directing attention to the area of the change (Chapters 3 and 4). Several slightly different ways have now been used to get around the problem that this introduces. One effective approach has been to mask a scene change by making several simultaneous conspicuous changes at different locations in the scene. This has been termed the 'mudsplash' technique because of the similarity to a familiar event on car windscreens. Investigations using this technique (Rensink et al. 1997; O'Regan *et al.* 1999) have shown that surprisingly large changes can fail to be detected.

The results have shown that, providing the attention capture introduced by visual transients is in some way masked, viewers frequently fail to detect substantial scene modifications. This striking phenomenon has been termed *change blindness*. Most people are somewhat amazed when first encountering the change blindness phenomenon. This was shown formally by Levin *et al*. (2000) who carried out a study in which individuals were asked to indicate how likely they would be to notice particular types of scene change in a test. There was a considerable discrepancy between the expected changes and the reported changes when the individuals were given the test. The participants had no formal training in vision. Change blindness is clearly incompatible with the 'picture in the head' account which we have termed passive vision. Levin's finding suggests that the passive vision assumptions are found widely in the population and do not depend on specific experience with passive vision theories.

7.3 Dynamic scenes and situations

The various situations discussed up to this point have had in common the fact the observer was simply taking in information from the visual environment. We have argued that eye scanning forms an active way of obtaining and selecting information in these situations. Although such observing occurs frequently in everyday life, a situation that is probably even more common is one in which the viewer is also engaged in carrying out some action. Under these circumstances, the pattern of eye scanning must be integrated into an overall action sequence. We turn now to examine some recent studies where this integration has been explored.

7.3.1 Deictic vision.

Ballard *et al.* (1992, 1997, see also Hodgson *et al.* 2000) devised an artificial manipulative task carried out by mouse-controlled manipulations of a computer screen display. The set up allowed a detailed record to be kept of both the manipulative actions and the eye scanning of the individual carrying out the action. The block assembly task that they used is shown in Fig. 7.5. Ballard et al used this situation to illustrate the following theoretical account of active vision.

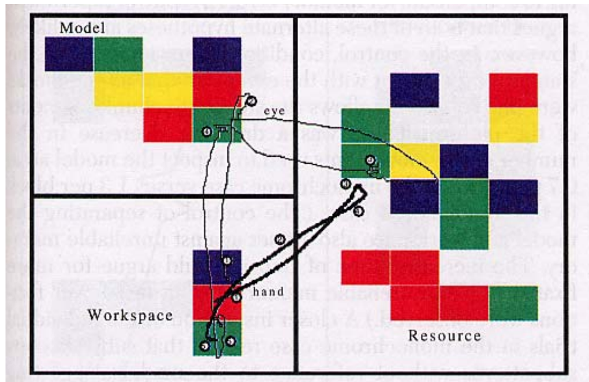


Figure 7.5 The 'blocks' task studied by Ballard et al. The subject operates on a computer display and has the task of assembling a copy of the Model (top left) in the Workspace (bottom left). To do this it is necessary to operate with the mouse, obtaining blocks from the Workspace (right) using a click and drag procedure. The traces show an episode of visual and motor activity. The heavy trace shows the mouse movement while the light trace shows the corresponding eye scan record.

The crux of the theory is that the eyes are used *deictically*. The term *deictic* refers to the intrinsic ability of a certain action to serve as a pointer to information. From general considerations of computational theory, Ballard et al. argue that the use of pointers is essential for a cognitive system to operate. Eye pointing is one of several ways in which the cognitive system operates using pointers. Eye pointing becomes particularly important when vision interfaces with cognitively controlled action (as discussed in Chapter 2, visual control of non-cognitive aspects of action does not necessarily need eye pointing).

Fixation on an object allows "the brain's internal representations to be implicitly referred to an external point" (Ballard et al. 1997, p. 72). The deictic strategy employs this pointer as part of a general 'do it where I'm looking' strategy to select objects for action. This is the crucial reason for directing the eyes to an object and of course is supplemented by the enhanced visual resolution that occurs through foveal vision. The enhanced resolution and ready mobility of the eyes result in gaze shifts being the process of choice for this deictic activity although covert attention may provide an alternative additional pointer (see Chapters 3 and 6). Further additional pointers relate to purely memory-based activity. However, a consistent finding that emerges in part from the experimental investigations presented in the paper is that use of memory pointers is avoided if an alternative is available.

The data collected from the block assembly task supported this characterisation of the underlying cognitive operations. Blocks were invariably fixated before they were operated on. Furthermore there was clear evidence that the preferred strategy involved making minimal demands on any internalised memory. In the block assembly task, as has been found in other tasks, many more saccades were made than appear necessary on a logical basis. The most common sequence observed in the block assembly task was eye-to-model, eye-to-resource, pick-from-resource, eye-to model, eye-to-construction, drop-at-construction. The first eye-to-model shift would appear to be to acquire the colour information of the next block to be assembled, a suitable block is then found in the resource space. The second look at the model establishes, or at least confirms, the location of this block in the pattern, which is then added to the construction. This second look could be avoided if the location information was also stored on the first look. On occasions, the second look was omitted indicating that such use of memory was an option. However these sequences were much less common than the sequences in which a return was made to the model.

Ballard et al conclude that memory minimisation is a significant feature of activity in the situation. Cognitive representations (in this case the position of the block in the model) are computed as late as possible before the necessary action. This *'just-in-time'* strategy, it is argued, minimises both memory and computational loads. This is in complete contrast with the traditional passive vision account involving intensive computation of an internalised representation. Detailed representations are computed only as they are needed. What is held in memory is the location in the display where the visual information can be obtained. Efficient use of serial procedures involving eye fixations in this way minimises both the computational load and the memory load.

7.3.2 Vision supporting everyday actions.

An important series of ground-breaking studies has been carried out by a group led by Michael Land. Land has an established reputation as a biologist with an interest in invertebrate vision (e.g. Land, 1995) and his interest in dynamic patterns of human vision emerged from that background. He devised and built a light head mounted video based eye tracking system (Land and Lee, 1994), which enabled a record to be built up of the fixation positions adopted by an observer during a variety of active tasks (see Fig. 7.6). Tasks studied have ranged through driving (Land and Lee, 1994), table tennis (Land and Furneaux, 1997), piano playing (Land and Furneaux, 1997) and tea-making (Land et al. 1999).

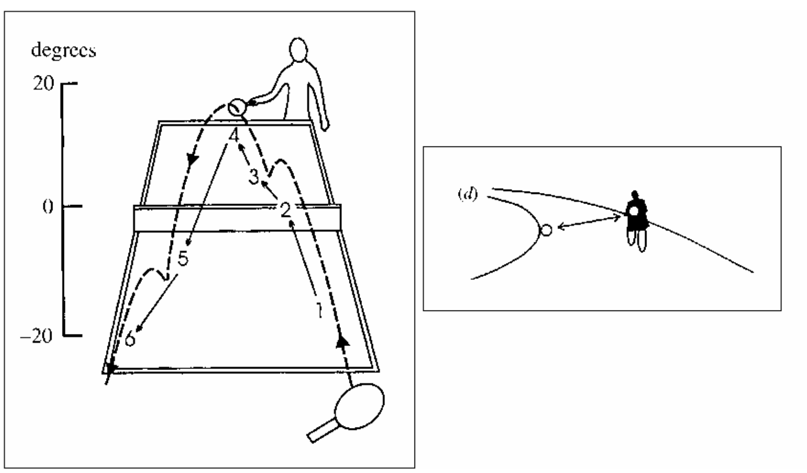


Figure 7.6 Two tasks studied by Land and Furneaux (1997). The left diagram shows a typical eye scan as an individual plays table tennis. The right hand diagram shows part of an eye scan of a driver negotiating a road curve in the presence of a cyclist.

The results, obtained during active vision supporting action, all demonstrate the strength of the principle the gaze is directed to the points of the scene where information is to be extracted. Exactly which regions of the visual scene are important for the task is often easy to define, such as when objects are being manipulated. However, in certain cases, the eye records have revealed, in an unexpected way, how vision is used. While driving around corners on a curving road, the eyes fixate the inside edge of the corner. While playing table tennis, the eyes are very active. Their activity takes roughly the same path as the ball. Contrary to popular belief, they do not *follow* the ball. The eye works in an anticipatory way. As the opponent makes the return shot, the player fixates the top of the net, using the clearance at this point to judge the ball's trajectory.

Land's analysis of tea making (Land et al. 1999) introduces the concept of *object-related action* units (ORAs) which he regards as the basic units of an action sequence (in the tea-making case, 'find the kettle' and 'transport to sink' are examples of such units). These units, with very rare exceptions, are carried out sequentially and involve engagement of all sensorimotor activity on the relevant object or objects. The eyes move to the object, or the point at which the object's activity is directed, *before* the manipulation starts. In general the eyes anticipate the action by about 0.6 sec. There may be action that precedes the eye movement. In cases where it is necessary to reposition the body, the movement of the trunk occurs first, in turn a further 0.6 sec before the eyes start to move. During a single ORA, the eye saccades around the object but when shifting between one ORA and another, very large saccades might occur. Although most saccades had amplitudes between 2.5 and 10 deg, as a result of these large saccades, the mean saccade size over the whole action was thus 18- 20 deg. This is appreciably larger than previous estimates (§ 2.4).
.
The eye movements could, with only occasional exceptions, be placed into one of the following categories: *locate* - an object to be used; *direct* – fixate an object which is to be manipulated; *guide* – bring two objects (e.g. kettle and lid) into contact; *check* the state of an object (e.g. water level). The 'locate' part of the sequence frequently required a search process. Land et al suggest that search patterns show that memory for object locations is not precisely coded. During the 'guide' process, the eyes frequently made a sequence of fixations from one object to the other. During the 'locate' and 'check' stages, quite long fixations occurred, as though the eyes were waiting for events to catch up. In a similar way to the case of saccade amplitudes, this led to a long tail on the distribution of fixation durations and a mean fixation duration of 0.42 sec, significantly longer than that found in picture scanning (§ 7.1.2).

Although Land is less concerned with the computational aspect of the brain activities involved in active vision, there are evident similarities between his analyses and those of Ballard's group presented in the previous section. In both cases, visual actions are based around objects in a deictic way. A highly important feature of these analyses is the potential for a long overdue integration of work in eye scanning with emerging theories of sequential action (Norman and Shallice, 1986; Schwartz et al. 1995).

Recent studies of simple pointing and prehension movements have thrown further light on how these tasks are controlled through active vision. Neggers and Bekkering (2000, 2001) have demonstrated a powerful *gaze anchoring* phenomenon. They show that during a manual pointing task the gaze is directed at the target movement and until the hand reaches this target, it is immune to distraction from a peripheral stimulus which would normally be considered a powerful stimulus for orienting (§ 6.5). Johannson *et al.* (2001) studied a task where subjects reached to grasp an object and use it to operate a control target switch. They report that 'landmarks where contact events took place are obligatory gaze targets'. Gaze is also likely to be directed at obstacles in the path of the action, but this is less obligatory.

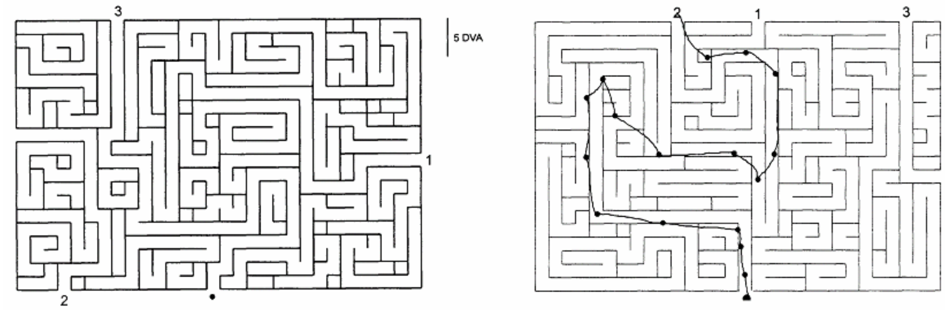


Figure 7.7. Maze task used by Crowe *et al*. (2000). Participants were presented with a maze (example in left diagram) asked to find which exit (1,2 or 3) corresponded with a path through the maze. An example of an eye scan while performing the task is shown on the right.

Another recent piece of work has also considered how eye movements might be involved in visual behaviour where the emphasis, rather than bringing a target object into foveal vision, concerns tasks of spatial routefinding. Crowe *et al.* (2000) looked at the microstructure of behaviour as participants worked out a route through a geometric maze (see Fig. 7.7). As participants worked out a route through a simple maze, their eye fixations fell at successive points on the route. However instances were observed where a large re-entrant section of the maze was not fixated. Crowe et al suggested that the omitted part of the route was in fact scanned with covert attention. Such an appeal to covert attention might obtain support from work on a task of following a continuous line in the presence of a second, intersecting, line.

7.4 Summary

This chapter has given a brief review of studies aimed at developing understanding of how eye scanning operates in less constrained visual environments. Research in this area is very active, and the wide variety of approaches reflect the excitement felt by individuals who converge on a common problem from a variety of start points.

CHAPTER 7 REFERENCES

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

Ballard, D. H., Hayhoe, M. M., Li, F, and Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. Philosophical Transactions of the Royal Society, Series B, **337B**, 331-339.

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

Biederman, I. (1972). Perceiving real world scenes. Science, **177**, 77-80.

Biederman, I. (1981). On the semantics of a glance at a scene. In Perceptual organization. (ed. M. Kubovy and J. R. Pomerantz), pp. 213-253, Lawrence Erlbaum, Hillsdale, NJ.

Biederman, I. (1987). Recognition by components : a theory of human image understanding. Psychological Review, **94**, 115-145.

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

Boyce, S. J. and 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.

Bozkov, V., Bohdanecky, Z. and Radil-Weiss, T. (1977). Target point selection during scanning eye movements. Biological Cybernetics, **27**, 215-220.

Buswell, G. T. (1922, 1937). Fundamental reading habits; a study of their development. Supplementary Educational Mongraphs, #21. How adults read. Supplementary Educational Monographs, #50. cited in Woodworth, R. S. and Schlosberg, H. (1954). Experimental Psychology, Methuen, London.

Buswell, G. T. (1935). How people look at pictures. University of Chicago Press, Chicago.

Chapman, P. R. and Underwood, G. (1998). Visual search of dynamic scenes: event types and the role of experience in driving situations. In Eye guidance in reading and scene perception, (ed. G. Underwood), pp. 369-393, Elsevier, Amsterdam.

Crowe, D. A., Averbeck, B. B., Chafee, M. V., Anderson, J. H. and Georgopoulos, A. P. (2000). Mental maze solving. Journal of Cognitive Neuroscience, **12**, 813-827

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

Delorme, A., Richard, G. and Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. Vision Research, **40**, 2187-2200.

Ellis, S. R. and Stark, L. (1986). Statistical dependency in visual scanning. Human Factors, **28**, 421-438.

Fabre-Thorpe, M., Richard, G. and Thorpe, S. J. (1998). On the speed of natural scene categoriation in human and non-human primates. Cahiers de Psychologie Cognitive, **17**, 791-805

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

Gilchrist, I. D. and Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. Current Biology, **10**, 1209-1212.

Grimes J (1996). On the failure to detect changes in scenes across saccades. In Perception, Volume 5 of the Vancouver Studies in Cognitive Science, (ed. K. Akins), pp. 89-110, Oxford University Press, New York.

Guba, E., Wolf, W., De Groot, S., Knemeyer, M., Van Atta, R. and Light, L. (1964). Eye movements and TV viewing in children. Audio-Visual Communication Review, **12**, 386-401

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

Henderson, J. M. and Hollingsworth, A. (1998). Eye movements during scene viewing: an overview. In Eye guidance in reading and scene perception. (ed. G. Underwood,), pp. 269-293, Elsevier, Amsterdam.

Henderson, J. M. and Hollingworth, A. (1999a). High-level scene perception. Annual Review of Psychology, **50**, 243-271.

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

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., McClure, K. K., Pierce, S. and Schrock, G. (1997). Object identifiaction without foveal vision: evidence from an artificial scotoma paradigm. Perception and Psychophysics, **59**, 323-346.

Hodgson, T. L., Bajwa, A., Owen, A. M. and Kennard, C. (2000). The strategic control of gaze direction in the tower of London task. Journal of Cognitive Neuroscience, **12**, 894-907.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. Journal of Experimental Psychology, Human Learning and Memory, **6**, 1-12.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. Journal of Experimental Psychology, Human Perception and Performance, **7**, 604-610.

Irwin D and Zelinsky G J (2002). Eye movements and scene perception: memory for things observed. Perception and Psychophysics, **64**, 882-895.

Johansson, R. S., Westling, G., Bäckström, A. and Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. Journal of Neuroscience, **21**, 6917-6932.

Jonides. J., Irwin, D. E. and Yantis, S. (1982). Integrating information from successive fixations. Science, **215**, 192-194.

Jonides, J., Irwin, D. E. and Yantis, S. (1983). Failure to integrate information from successive fixations, Science, **222**, 188.

Kundel, H. L., Nodine, C. F. and Carmody, D. (1978). Visual scanning, pattern recognition and decision making in pulmonary nodule detection, Investigative Radiology, **13**, 175-181.

Krieger, G., Rentschler, I., Hauske, G., Schill, K. and Zetsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher order statistics. Spatial Vision, **13**, 201-214.

Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. Acta Radiologica, **3**, 137-144.

Land, M. F. (1995). The functions of eye movements in animals remote from man. In Eye movement research : mechanisms, processes and applications. (ed.. J. M. Findlay, R. Walker and R. W. Kentridge), pp. 63-76, Elsevier, Amsterdam.

Land, M. F. and Furneaux, S. (1997). The knowledge base of the oculomotor system. Philosophical Transaction of the Royal Society Series B, **352B**, 1231-1239.

Land, M. F. and Lee, D. (1994). Where we look when we steer. Nature, **369**, 742-743.

Land, M. F., Mennie, N. and Rusted, J. (1999). The roles of vision and eye movements in the control of activities of everyday living. Perception, **28**, 1311-1328.

Levin, D. T., Momen, N., Drivdahl, S. B. and Simons, D. J. (2000). Change blindness blindness: the metacognitive error of overestimating change-detection ability. Visual Cognition, **7**, 397-412.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. Cognitive Psychology, **3**, 525-551.

Loftus, G. R. (1985). Picture perception: effect of luminance level on available information and information extraction rate. Journal of Experimental Psychology, Human Perception and Performance, **114**, 342-356.

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.

Logothetis, N. K. and Sheinberg, D. L. (1996). Visual object recognition. Annual Review of Neuroscience, **19**, 577-621.

Mackworth, N. H. and Morandi, A. J. (1967). The gaze selects informative detail within pictures. Perception and Psychophysics, **2**, 547-552.

Mannan, S. K., Ruddock, K. H. and Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. Spatial Vision, **9**, 363-386.

Mannan, S. K., Ruddock, K. H. and Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. Spatial Vision, **10**, 165-188.

Mannan, S. K., Ruddock, K. H. and Wooding, D. S. (1997). Fixation sequences during visual examination of briefly presented 2D images. Spatial Vision, **11**, 157-178.

Mourant, R. R. and Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. Human Factors, **14**, 325-335.

Marr, D. (1982). Vision, W H Freeman, San Francisco.

Neggers, S. F. W. and Bekkering, H. (2000). Ocular gaze is anchored to the target of an ongoing pointing movement. Journal of Neuophysiology, **83**, 639-651.

Neggers, S. F. W. and Bekkering, H. (2001). Gaze anchoring to a pointing target is present during the entire pointing movement and is driven by a non-visual signal. Journal of Neurophysiology, **86**, 961-970.

Nelson, W. W. and Loftus, G. R. (1980). The functional visual field during picture viewing. Journal of. Experimental Psychology, Human Learning and Memory, **6**, 391-399.

Noton, D. and Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognising patterns. Vision Research, **11**, 929-942.

Norman, D. A. and Shallice, T. (1986). Attention to action: willed and automatic control of behaviour. In Consciousness and self-regulation: advances in research and theory. Volume 4. (ed. R. J. Davidson, G. E. Schwartz and D. Shapiro), pp 1-18, Plenum, New York.

Noton, D. and Stark, L. (1971) Scanpaths in eye movements during pattern perception. Science, **17**1, 308-311.

Oliva, A. and Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. Cognitive Psychology, **41**, 176-210.

O'Regan, J. K., Rensink, R. and Clark, J. J. (1999). Change-blindness as a result of mudsplashes. Nature, **398**, 34.

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

Poggio, T. and Edelman, S. (1990). A network that learns to recognise three dimensional objects. Nature, **343**, 263-266.

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

Pollatsek, A., Rayner, K. and Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. Journal of Experimental Psychology, Human Perception and Performance, **16**, 199-230.

Ponsoda, V., Scott, D. and Findlay, J. M. (1995). A probability vector and transition matrix analysis of eye movements during visual search. Acta Psychologica, **88**, 167-185.

Potter, M. C. and Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology, **81**, 10-15.

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

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

Rimey, R. D. and Brown, C. M. (1991). Controlling eye movements with hidden Markov models. International Journal of Computer Vision, **7**, 47-65.

Saida, S. and Ikeda, M. (1979). Useful visual field size for pattern perception. Perception and Psyhophysics, **25**, 119-125.

Schwartz, M. F., Montgomery, M. W., Fitzpatrick-DeSalme, E. J., Ochiopa, C., Coslett, H. B. and Mayer, N. H. (1995). Analysis of a disorder of everyday action. Cognitive Neuropsychology, **12**, 863-892.

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

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

Schyns, P.G. and Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. Perception, **26**, 1027-1038.

Snodgrass, J. G. and Vanderwart, M. (1980). A standardized set of 260 picutres: Norms for name agreement, image agreement, familiarity and visual complexity. Journal of Experimental Psychology, Human Learning and Memory, **6**, 174-215.

Stark, L. and Ellis, S. R. (1981). Scanpaths revisted: cognitive models direct active looking. In Eye movements: cognition and visual perception. (ed. D. F. Fisher, R. A. Monty and J. W. Senders), pp. 193-226, Lawrence Erlbaum Associates, New Jersey.

Tarr, M. J. and Bülthoff, H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Journal of Experimental Psychology, Human Perception and Performance, **21**, 1494-1505.

Thorpe, S. J., Fize, D. and Marlot, C. (1996). Speed of processing in the human visual system. Nature, **381**, 520-522.

Tosi, V., Mecacci, L. and Pasquali. E, (1997). Scanning eye movements made when viewing film: Preliminary observations. International Journal of Neuroscience, **92**, 47-52.

Van Diepen, P. M. J., Wampers, M. and d'Ydewalle, G. (1998). Functional division of the visual field: moving masks and moving windows. In Eye guidance in reading and scene perception, (ed. G. Underwood), pp. 337-355, Elsevier, Amsterdam.

Viviani, P. (1990). Eye movements in visual search. Cognitive, perceptual and motor control aspects. In Eye movements and their role in visual and cognitive processes. (ed. E. Kowler), pp. 353-393, Elsevier, Amsterdam.

Walker-Smith, G. J., Gale, A. G. and Findlay, J. M. (1977). Eye movement strategies involved in face perception. Perception, **6**, 313-326.

Williams, A. M. and Davids, K. (1998). Visual search strategy, selective attention and expertise in soccer. Research Quarterly for Exercise and Sport, **69**, 111-128.

Williams, A. M., Davids, K. and Williams, J. G. (1999). Visual perception and action in sport. Taylor & Francis, London.

Yarbus, A. L. (1967). Eye movements and vision. (English translation by L. A. Riggs), Plenum Press, New York

Zelinsky, G. J. (2001). Eye movements during change detection: implications for search constraints, memory limitations and scanning strategies. Perception and Psychophysics, **63**, 209-225.