VISUAL COGNITION, 2005, *12* (3), 473–494

Psychology Press
Taylor & Francis Group

Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration

Pieter J. A. Unema
Maastricht University, The Netherlands

Sebastian Pannasch, Markus Joos, and Boris M. Velichkovsky
Dresden University of Technology, Germany

The present study focuses on two aspects of the time course of visual information processing during the perception of natural scenes. The first aspect is the change of fixation duration and saccade amplitude during the first couple of seconds of the inspection period, as has been described by Buswell (1935), among others. This common effect suggests that the saccade amplitude and fixation duration are in some way controlled by the same mechanism. A simple exponential model containing two parameters can describe the phenomena quite satisfactorily. The parameters of the model show that saccade amplitude and fixation duration may be controlled by a common mechanism. The second aspect under scrutiny is the apparent lack of correlation between saccade amplitude and fixation duration (Viviani, 1990). The present study shows that a strong but nonlinear relationship between saccade amplitude and fixation duration does exist in picture viewing. A model, based on notions laid out by Findlay and Walker's (1999) model of saccade generation and on the idea of two modes of visual processing (Trevarthen, 1968), was developed to explain this relationship. The model both fits the data quite accurately and can explain a number of related phenomena.

Free eye movements during scene perception have been the subject of study throughout much of the history of psychology, but, despite this, no clear theoretical account of the processes involved has been developed. Below we

---

Please address all correspondence to: P. J. A. Unema, Faculty of Psychology, Maastricht University, PO Box 616, NL 6200 MD Maastricht, The Netherlands.
Email: unema@applied-cognition.org

The cooperation between the universities of Dresden and Maastricht was partially supported by a grant from the German Research Foundation (DFG), No. VE 192/8-1. The authors thank Jens R. Helmert, Fiona Mulvey, Constanze Liebers, and Tilo Wehner for fruitful discussions and support in conducting this study. We also thank two anonymous reviewers for their helpful comments and suggestions.

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280444000409

outline some of the important findings and theoretical implications of eye movement studies of scene perception.

Probably one of the earliest systematic studies of looking at pictures was performed by Guy Buswell (1935). His general method was to have subjects look at pictures as if they were choosing one for themselves, for as long (or short) as they pleased. Studying large samples over many trials, Buswell asked a number of questions: What are the regions that attract attention what is looked at first, what is looked at last, and what is looked at longest. In other words: How is attention deployed? According to Buswell (1935, p. 9), ‘‘Eye movements are unconscious adjustments to the demands of attention during a visual experience.’’ One of his general conclusions was that there are two distinguishable patterns of eye movements: ‘‘One of these consists of a series of relatively short [fixation] pauses over the main portions of the picture’’ (p. 142) and a second ‘‘in which series of fixations, usually longer in duration, are concentrated over a small area of the picture’’ (p. 142). Furthermore, he observed that ‘‘the duration of the fixation pauses increases, as one looks at the picture’’. In numbers: The duration of the first five fixations averaged ~210 ms, increasing to ~366 ms over the last five fixations. He further noted that the increase in fixation duration continued until around the 100th fixation, after which no systematic increase was observed.

Karpov, Luria, and Yarbus (1968) concluded that there is apparently an orienting period during which the physically salient features of the image are scanned, followed by a period of scrutiny in which the more informative areas are fixated. This seems to concur with Buswell’s two patterns, although Buswell did not explicitly conclude that these patterns always occur in that order.

Antes (1974), like Buswell (1935), found that the fixation duration gradually increases over time while at the same time saccadic extent decreased. Friedman and Liebelt (1981) expounded on this and showed that fixations on incongruous objects were longer than fixations on other objects. In contrast to incongruous objects, the duration of fixations on congruous objects increased over time. Unfortunately, they did not study the time course of saccadic amplitude. Scinto, Pillalamarri, and Karsh (1986) proposed that the increase of fixation duration and decrease of saccade amplitude in a visual search task should be considered as a strategic adaptation to the demands of the task. Although these studies suggests that fixation duration and saccade latency are affected in a similar way during the first few seconds of inspection, none of them explains the nature of the relationship between the mechanisms controlling saccade amplitude and fixation duration.

A positive relationship between saccade amplitude and saccade latency has been known to exist for some time, however (Cohen & Ross, 1977; Viviani & Swensson, 1982). Zihl and Hebel (1997) found a weak positive correlation between fixation duration and the size of a subsequent saccade in normal subjects but not in subjects with frontal or posterior parietal lobe damage. Pelz and Canosa (2001), on the other hand, reported a weak but significant negative correlation between fixation duration and the amplitude of subsequent saccades during the execution of the ‘‘complex’’ task of washing hands. Since the task is so drastically different from the more constrained situations used in other studies, it is difficult to say what exactly caused this contradictory result. Viviani (1990) argues that the amplitude–latency relationship is much weaker in picture viewing than in step-tracking (following a single or double stimulus jump) for one or more of the following reasons: (1) The size of saccades used in picture scanning is smaller than those used in studies like those of Viviani and Swensson (1982) and Zihl and Hebel (1997), while the latencies for small saccades are almost constant; (2) fixation durations found in scene inspection are more than just saccade latencies for several reasons: Since all potential targets are available continuously in free viewing, there is no way of determining saccadic latencies unequivocally; also, there is a brief, refractory period at the end of each saccade during which no new saccade can be produced (e.g., Harris, Hainline, Abramov, Lemerise, & Camenzuli, 1988); (3) the task of scene inspection may require other processes than those involved in step-tracking tasks, so that inferences drawn from the latter may not apply to the former.

INFORMATIVENESS

It has been argued by several researchers that the informativeness of objects or regions in an image affects the way in which images are scanned. In this section we review some studies on the effects of informativeness on fixation duration, first placement, and fixation density.

What is looked at longest? In his classical study of eye movements in picture perception, Buswell (1935, p.72) concluded from his observations that ‘‘it appears that the fixations of exceptionally long duration fall, for the most part, on those portions of the picture which, judged by the number of fixations, seem to be the principal centers of interest’’. Friedman and Liebelt (1981) found that fixations on objects with a (more informative) low-rated likelihood were generally longer than on objects with high-rated likelihood. They further found that the duration of fixations on objects with a high-rated likelihood increases over time, while those on objects with a low-rated likelihood did not change. De Graef, Christiaens, and d’Ydewalle (1990) found that *first* fixations on (more informative) ‘‘nonobjects’’ were longer than on other objects, but these longer fixations were not found in the initial looking period. Henderson, Weeks, and Hollingworth (1999) similarly showed that first fixations on (more informative) semantically inconsistent objects were longer than on semantically consistent objects.

What is looked at first? Yarbus (1967) noted that his subjects tended to look at the most important elements of a picture during the first ~25 s of their viewtime, while secondary elements were scrutinized later: ‘‘Apparently, the fifth period [of ~5 s] concluded the first cycle of examination of the picture’’.

Scanpaths varied strongly with the instructions according to which a subject scanned the images, but were highly replicable from one day to the next. Antes (1974) found that fixations on informative areas are concentrated toward the initial few seconds of viewing, while less informative areas received a greater proportion of the fixations later. De Graef et al. (1990) found no evidence that the number of fixations on (more informative) nonobjects changed over time, however. Henderson et al. (1999) similarly found that (more informative) semantically inconsistent objects were no more likely to be fixated early than the consistent objects.

What is looked at most often? Mackworth and Morandi (1967) studied the question what is fixated most often by a priori determining the informativeness of regions and registering the distribution of fixations over 2 s intervals of a 10 s registration period. They found that the more informative areas received a higher fixation density than less informative areas. The distribution of fixation density over informative and less informative regions appeared to be stable over time, so it was concluded that the informativeness of an area is appreciated immediately. Loftus and Mackworth (1978) found that viewers, when asked to memorize simple line drawings, tend to refixate semantically inconsistent (more informative) objects more often than semantically consistent (less informative) objects. Similarly, Henderson et al. (1999) found semantically inconsistent objects to be fixated more often than other objects. In contrast, Friedman and Liebelt (1981) found fixation density to be higher for the high-probability objects than for low-probability objects.

Although some caution should be exercised when comparing all the different operationalizations of informativeness, most studies agree that the fixation duration on more informative objects tends to be longer than on less informative objects and does not change systematically over time. Less agreement exists on the question of first fixation placement. The timescale of Yarbus' (1967) study cannot readily be compared with that of Antes (1974). The concentration of fixations on more informative areas in the initial looking period found by the latter could not be reproduced by de Graef et al. (1990) or Henderson et al. (1999). There are a number of differences between Antes' study (subjects were asked to rate their appreciation of scenes taken from the Thematic Apperception Test) and those of de Graef et al. (subjects searched for nonobjects in line drawings) or Henderson et al. (subjects were asked to memorize line drawings) that may have led to the choice of different scanning strategies. It should be noted, however, that the relationship between informativeness and early patterns of viewing is still debated (Henderson, 2003). Similarly, some disagreement seems to exist on the density of fixations on informative objects. Here, too, differences in task parameters and stimulus materials may have caused the differences in eye-movement parameters. Loftus and Mackworth (1978) used simple, and Friedman and Liebelt (1981) and Henderson et al. (1999) used detailed, black-and-white line drawings and asked viewers to memorize the images; Mackworth and Morandi (1967) used half-tone, (photo-)realistic stimuli and asked subjects to rate their appreciation of the images. It thus appears that visual informativeness and semantic informativeness have differential effects on scanning strategies.

MODELS

These observations give rise to the question of what actually controls the time spent fixating an object. The notion of ‘‘informativeness’’ implicitly assumes that attentional processes are involved in the control of eye movements during scene perception. How much attention does programming a saccade actually need? Henderson (1992, 1993) proposed a sequential attentional model of eye-movement control. The model assumes that attention is allocated to the foveal information at the beginning of the fixation. Once sufficient information has been extracted to identify the foveal object, a reallocation of attention commences the programming of a new saccade. This is established on the basis of a preattentive map of potential targets. The attentional system then allocates weights to these targets—presumably based on the task parameters. Finally, the location with the highest weight ‘‘gets’’ the saccade. If the speed of foveal processing is slow, a saccade may be executed before the process of identification is complete—in which case a refixation should occur.

Van Diepen, de Graef, and d’Ydewalle (1995) and Van Diepen, Wampers, and d’Ydewalle (1998) used a moving mask paradigm in which participants explored line drawings of real-life situations. The fixated objects were masked with several durations from fixation onset, and participants were asked to search the scene for ‘‘nonobjects’’. The results showed that information from the foveally present stimulus is usually extracted within the initial 45–75 ms of a fixation; masks applied during this period led to increased fixation durations, while masks presented later did not disrupt the time course of a fixation. In another moving mask study, Van Diepen and d’Ydewalle (2003) showed that peripheral masking disrupts saccade target selection: Saccade amplitude decreases with mask duration while a small but significant increase in fixation duration is found. The sequential attention model cannot explain why not only foveal masking, but also peripheral masking early during fixations, disrupts the normal time course of fixations. The results also contradict the claim that saccade targets are selected late during a fixation. Van Diepen and d’Ydewalle critically point out, however, that what is early or late is relative to the duration of the fixation. For short fixations, the saccadic information available after 70–120 ms may well be late. As has been found in reading research, the target of the next saccade is typically programmed between 25 and 100 ms from the beginning of a fixation (Reichle, Pollatsek, Fisher, & Rayner, 1998). A critical difference with reading, however, is that in scene perception the next target has to be actually selected, whereas in reading the default target is always the next word.

The sequential attention model of eye guidance assumes that there is a close link between covert and overt attention and it suggests that a covert movement of attention precedes the actual saccade. Support for this notion comes from the enhanced discrimination ability at the saccade target location before the saccade is executed (Deubel & Schneider, 1996; Kowler, Anderson, Dosher, & Blaser, 1995).

According to the premotor theory of attention (Rizzolatti, Riggio, Dascola, & Umiltà, 1987; Rizzolatti, Riggio, & Sheliga, 1994) there are many different representations of space that may be activated by an attentional mechanism, depending on the motor action (reaching, grasping, eye movements, etc.) for which it is intended. The selection of a target for a saccade involves activating a representation of oculomotor space. This perspective suggests that attention is a byproduct of any motor response rather than the agent initiating it. Rizzolatti and colleagues argue that there must be a unitary visual attentional system that controls both ‘‘selection for action’’ and ‘‘selection for perception’’. Deubel, Schneider, and Paprotta (1998) concluded on the basis of a study on reaching and eye movements that the preparation of a goal-directed movement binds the processing capacities to the same object, while objects other than the movement target are temporarily excluded from high-level visual analysis. Thus, enhanced processing at the location of a planned saccade may be explained on the basis of a premotor account, but, in contrast to Rizzolatti et al. (1987, 1994), Deubel et al. argue that the selection is object specific rather than spatially selective.

Findlay and Walker (1999) proposed a model of saccade generation based on the mutually competitive inhibition of ‘‘move’’ and ‘‘fixate’’ activity that avoids the concept of attention altogether. Four levels of control over both ‘‘where-’’ and ‘‘when-systems’’ influence ‘‘the balance of power’’ between fixate and move responses. During the fixation, a salience map develops that represents the spatial locations of potential saccade targets. The salience map (a concept originally introduced by Koch and Ullman, 1985) is a two-dimensional network formed on the basis of the retinal image, in which the visual field is spatially mapped in a retinotopic way. The level of activity of any point on the map is determined by a combination of intrinsic salience, search selection, and spatial selection. Proximity to the present fixation location also increases an item’s intrinsic salience (Motter & Belky, 1998a, 1998b). The choice is ultimately made on the basis of a ‘‘winner-takes-all’’ strategy. The time at which this would happen during picture perception depends mainly on the way the fixate–move balance is controlled by *automated* processes, i.e., the kind of operative processes that can be overridden by voluntary decisions, but habitually control eye movements without interference from volition—lower level, *automatic* processes should be of lesser influence. The model shows that it is possible to describe a large number of eye-movement related phenomena through the processes involved in the planning and execution of saccades.

Following a more neurophysiological tradition, a distinction between two visual pathways was made by the authors of a special issue of *Psychologische Forschung* in 1968 (cf. Held, Ingle, Schneider, & Travarthen, 1967). Trevarthen (1968, p. 299) coined the terms *ambient* and *focal* as labels for ''anatomically distinct brain mechanisms'' that serve parallel functions of spatial orientation (the subcortical pathway) and visual identification (the cortical pathway). The original approach was later modified by the suggestion that both systems are dominated by distinct cortical regions (e.g., Milner & Goodale, 1995; Milner et al., 1991; Ungerleider & Mishkin, 1982; Velichkovsky, 1982). The ventral stream sends information from the occipital lobes to the inferotemporal regions. It processes mainly foveal signals at a relatively slow pace. The dorsal stream relays ambient visual information rapidly, but at a lower spatial resolution to the posterior parietal cortex. This functional distinction establishes the duality of two modes of processing that are concerned with the questions ''what'' and ''where/how'', respectively. Recently, Velichkovsky, Dornhoefer, Pannasch, and Unema (2000) and Velichkovsky, Rothert, Kopf, Dornhoefer, and Joos (2002) reported a relationship between these modes of visual processing and eye movements: Fixations of shorter durations are mostly followed by saccades of longer amplitudes (*ambient* processing) and longer fixations are often followed by shorter saccades (*focal* processing). A possible basis of this interplay is the fact that the eye primary-motor area has both dorsal stream and ventral stream input (Kaufer & Lewis, 1999).

It may be speculative to apply this perspective to the phenomena of the gradual change of fixation durations and saccade amplitudes during picture perception. However, a cautious translation into these terms could be that, initially, a larger proportion of processing is apparently dedicated to the ambient task of determining the whereabouts of the objects in the scene, since the duration of initial fixations seems to be no longer than the minimal time to produce visually guided saccades.

Several studies (e.g., Antes, 1974; Buswell, 1935; Friedman & Liebelt, 1981) have shown that saccade amplitude and fixation duration reach an asymptotic level after some time. This gives rise to the speculation that both saccade amplitude and fixation duration are governed by a common mechanism. If such a mechanism exists, their respective time courses ought to be similar and factors affecting one should also affect the other. In the experiment presented here, we manipulated the number of objects viewed in the scene in the expectation that this would affect the average saccade amplitude. Thus, if the number of objects in an image affects the time course of saccade amplitude, this should also be the case for fixation durations and vice versa. Furthermore, we carefully examined the relationship between saccade amplitude and fixation duration following previous studies associating this relationship with the distinction between ambient and focal vision.

METHOD

Subjects

Fifteen healthy females with normal or corrected to normal vision gave their written informed consent to participate in the experiment. Subjects' ages ranged from 18 to 28 years with a mean age of 22 (*SD* = 2.57); all subjects were right-handed. They were paid for their participation.

Apparatus

Eye movements were recorded at 250 Hz using the SR Research Ltd EyeLink eye tracking system with a spatial resolution of around 15 min of arc and an approximate accuracy of 1°. Fixation onset was detected online and transmitted to the presentation system with a delay of approximately 12 ms. Images were displayed using an nVidia/TNT2 64 pro graphic chip and a CRT display (17-inch, Dell P990) at 1024 × 768 pixels at 256 colours and a frame rate of 100 Hz. Viewed from a distance of 60 cm, the screen subtended a visual angle of ~31° horizontally and ~26° vertically.

Stimuli and design

Each stimulus consisted of an image of the interior of a room, rendered by a freeware software package for interior design. Forty-eight 1024 × 768 pixel images with a 24 bit colour depth were constructed, half of which had an object density of 16 objects; the other half had 8 objects. ''Objects'' were defined as small, movable objects, excluding doors, shelves, floors, windows, and the like. ''Background objects'' were kept in low saturation colours. The objects were distributed according to a (4 × 4) rectangular grid (see Figure 1) and had a size range of ~1° to ~3.5° in either direction.

Figure 1. Example of the computer-generated scenes with 8 (left) and 16 objects (right). The white lines represent the grid according to which objects were placed.

Images were presented for 20 s each, after which a statement was presented in a pop-up window. The statements could either concern the global distribution of objects (‘‘there were more objects in the left half of the picture’’) or the presence/absence of a particular object (‘‘there was a white frying pan on the counter’’). Half of the statements were true; the other half were false. Subjects were asked whether they thought the statement was true or false. Thus there were four conditions: Eight or Sixteen objects and ‘‘where-’’ or ‘‘what-type’’ questions. Conditions were presented in one of four fixed orders.

Procedure

Participants were seated in a sound attenuated, dimly lit room. A chinrest was used to stabilize the posture and to keep the viewing distance constant. Each participant was informed of the purpose of the study before signing their consent. Recording took place in two sessions, each lasting approximately 15 min. Participants were instructed not to move unduly, but if necessary, to change position during the intervals between the image presentations. The total duration of the experiment, including the preparation for recording and calibration, was around an hour per subject.

Data analysis

All analyses involved SPSS 11.0 and MATLAB 6.0 software running on a Windows 2000 PC. Fixations and saccades were analysed offline using the EyeLink software. Saccades were defined after a velocity criterion ($\geq 30^\circ . s^{-1}$) or an acceleration criterion ($\geq 8000^\circ . s^{-2}$). Fixations were detected if a sequence of at least four consecutive samples fulfilled the criterion of not belonging to a saccade. Blinks were removed using a dedicated program in order to ensure that no eyeblinks were mistaken for saccades. Fixations immediately preceding and following eyeblinks and those lying outside the screen area were also discarded. Saccades starting or landing outside the screen area were excluded from the analysis. Only fixations and saccades starting after the image appeared and finishing before the image disappeared were used for further analysis.

RESULTS

Behavioural data

The type of question (‘‘where-type’’ questions versus ‘‘what-type’’ questions) had a significant influence on the number of questions answered correctly. The ‘‘what-type’’ questions were answered correctly about 69% of the time, whereas ‘‘where-type’’ questions scored about 80%, $F(1, 14) = 8.31$, $p = .012$. The number of objects in the picture also had a significant effect on the number of correct responses: 78% correct answers were given when the scene contained 8 objects, 72% when the scene contained 16 objects, $F(1, 14) = 6.91$, $p = .020$.

Furthermore, there was a significant interaction between the number of objects and the type of questions: The number of objects seemed not to play a role when ''where''-questions were asked (79% correct for the 8 objects condition, 80% for the 16 objects condition), whereas a significantly lower percentage (63% in the 16 objects condition, 76% in the 8 objects condition) of the ''what''-questions was answered correctly, $F(1, 14) = 14.76$, $p = .031$.

Fixation durations

The number of objects had a small but highly significant effect on fixation duration (fd). In the 8-objects condition, the average fixation lasted for 250 ms, in the 16-objects condition it was 240 ms, $F(1, 14) = 15.29$, $p = .002$. In order to analyse the time course of fixation durations, all fixations commencing within 500 ms bins were averaged for each subject. The last fixation of each trial was excluded if it continued after the end of the 20 s period.

In order to test the time course of fixation durations, the average fixation durations commencing within 500 ms bins entered a nonlinear regression analysis. As expected, there was a significant steady increase in fixation duration over the first few seconds (Figure 2). We tested the model:

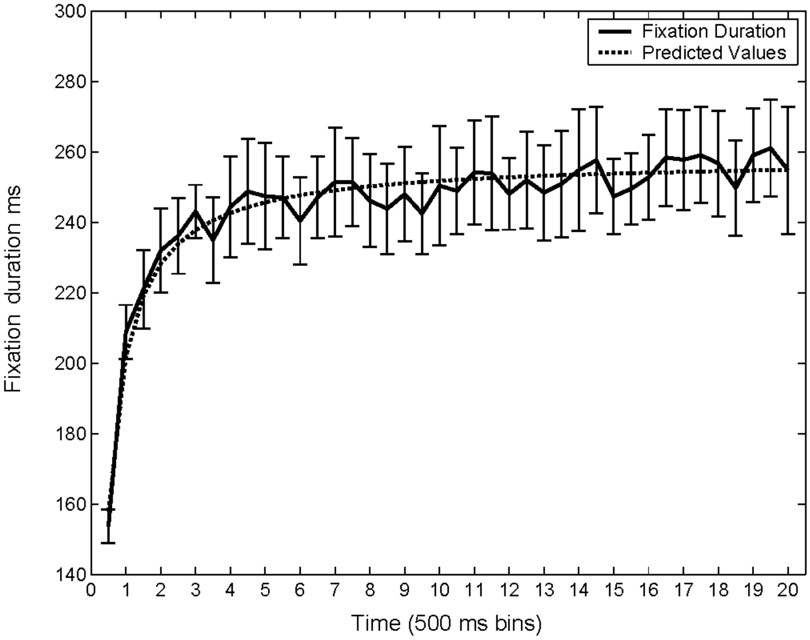

Figure 2. Fixation duration as a function of viewing time. The error bars represent the 95% confidence interval. The dotted line represents the predicted values (see text for details).

$$fd = b \cdot e^{(a/t)} \tag{1}$$

with b the asymptotic value and a the acceleration rate, in an nonlinear regression analysis. The regression function that best fit the data is described by:

$$fd = 252 \cdot e^{(-0.22/t)} \tag{2}$$

where t is the time at a resolution of 0.5 s, resulting in an R^2 = .94 (over the whole range of 20 s, i.e., 40 data points). The 95th percentile of the asymptotal fixation duration is reached after ~3.4 s, which is after roughly 11 fixations. The parameters were estimated using the nonlinear regression function of the software package.

The number of objects had a small but significant effect on both the asymptote (264 ms in the 8 objects condition, 250 ms with 16 objects) and on the acceleration ($-.27$ in the 8 objects condition and $-.21$ in the 16 objects condition), $F(1, 14) = 22.88$, $p = .001$ (see Figure 3).

In terms of time needed to reach 95% of the asymptotic value, this means that despite its lower asymptotic value (250 ms in the 16 objects case versus 264 ms

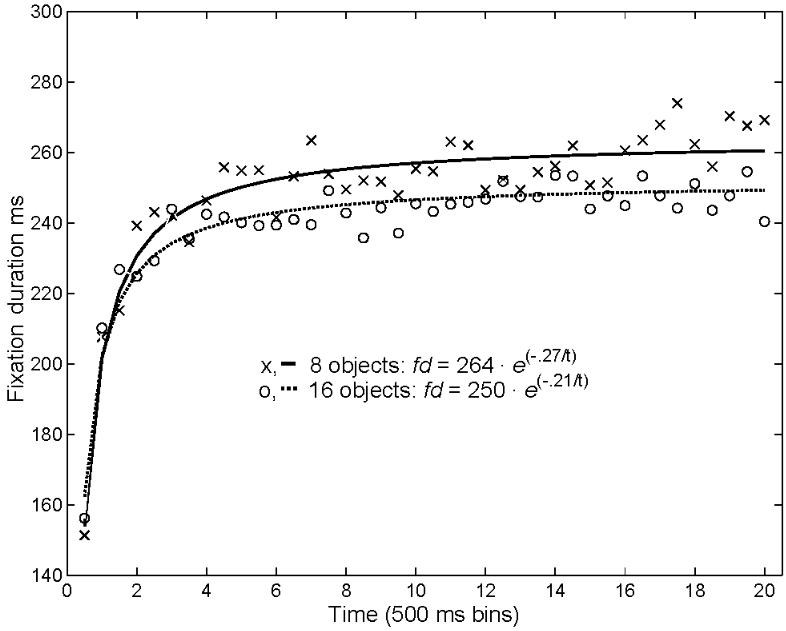

Figure 3. Predicted (lines) and observed (scattered x and o) average fixation durations as a function of viewing time.

in the 8 objects case) it is reached later (at 4.3 s) than in the 8 objects condition (3.11, see Fig. 3).

Saccade amplitudes

As expected, the number of objects affected the average saccade amplitude, $F(1, 14) = 22.64$, $p < .001$: The average saccade size was 6.3 degrees in the 8 objects scenes and 5.6 in the 16 objects scenes. Similar to fixation durations, a gradual change of average saccade amplitude was found to occur within the first few seconds (see Figure 4).

The same model as the one in (1) was tested for saccade amplitude:

$$sa = b \cdot e^{(a/t)} \tag{3}$$

Again with b as the asymptotic value and a as an acceleration parameter. The regression function:

$$sa = 5.26 \cdot e^{(0.10/t)} \tag{4}$$

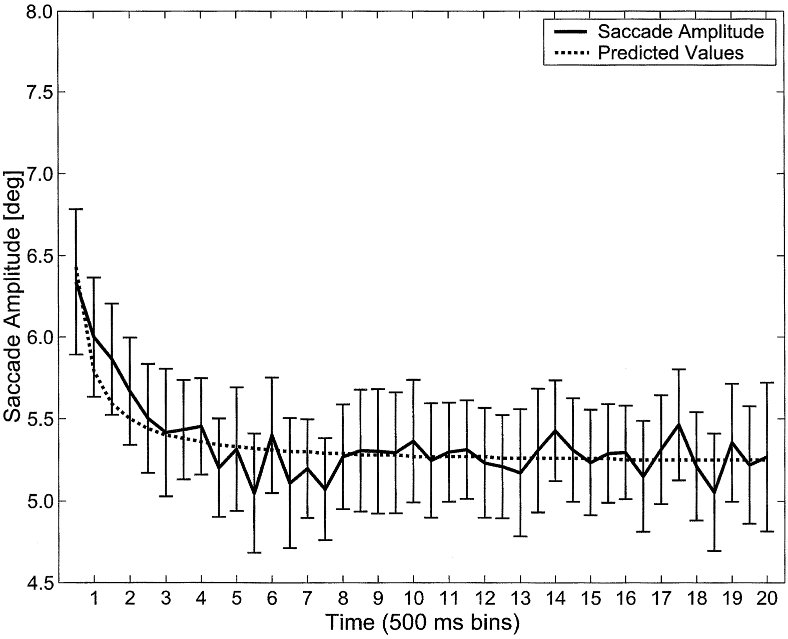

Figure 4. Saccade amplitudes as a function of viewing time. The error bars represent the 95% confidence interval. The dotted line represents the predicted values (see text for details).

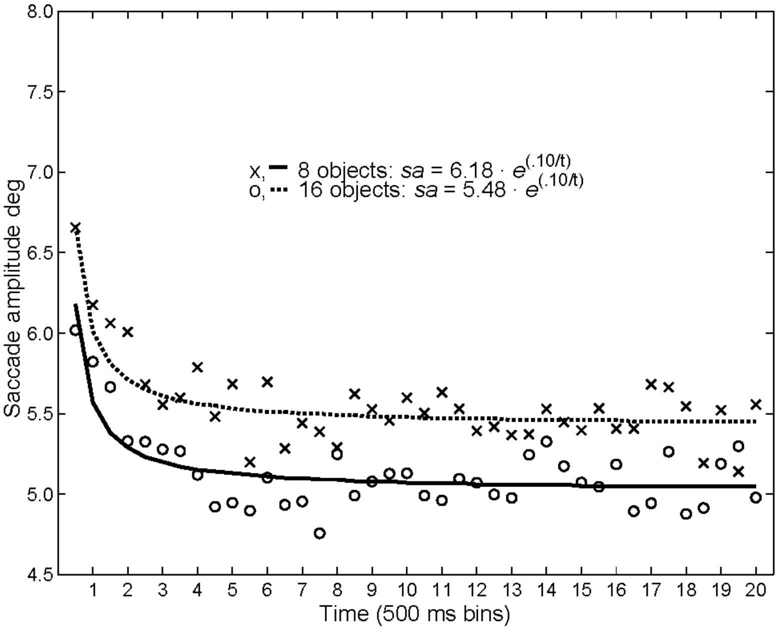

Figure 5. Predicted and observed average saccade amplitudes as a function of viewing time.

best fit the data and explained 78% of the variance. Saccade amplitude reaches the 95th percentile somewhat earlier than fixation duration, on average after ~2 s or ~8 saccades. Although there is a considerable difference in the asymptotic value between 8 and 16 objects (6.16 versus 5.48 degrees), $F(1, 14) = 46.07$, $p < .001$, there was no difference in the acceleration rate (in both cases 0.104), $F(1, 14) = 0.0$ (see Figure 5).

Landing positions

Participants' first saccade went to the left half of the scene more often (80%) than to the right half (20%), between the upper (58%) and lower (42%) halves there was less of a difference. All subjects but one showed this preference to start scanning in the left half of the scene, although they did not do so on all images. We calculated that ~63% of the first saccades landed on any object during first 2.5 s, while ~64% of saccades during the rest of the 20 s period landed on objects. The difference was not significant. Since scanning usually involves a number of saccades that do not immediately land on the intended target (e.g., Findlay, Brown, & Gilchrist, 2001; Viviani & Swensson, 1982), we were interested in the number of corrective saccades. Since there were no external criteria to determine whether a saccade should classified as being corrective (i.e., there were no ‘‘wrong’’ landing positions), we defined those that were preceded by short fixations ($\leq$180 ms) and had a maximal amplitude of $\leq 2.5^\circ$ as corrective saccades; 7.4% of all saccades met these criteria. We divided the total time per picture into eight 2.5 s periods to see if corrective saccades prevailed during any particular period. A repeated measures ANOVA showed that there was no overall influence of period on the occurrence of corrective saccades, $F(7, 98) = 0.625$, $p = .712$.

Saccade amplitude as a function of fixation duration

In order to study the relationship between fixation duration and saccade latency, saccade amplitudes were averaged over all fixations within 20 ms bins. Figure 6 shows the average subsequent saccade amplitude for all fixations up to 600 ms. Longer fixation durations were left out of further analyses because of their relative rarity. As is clearly visible, a subset of fixations of $\leq$180 ms come with significantly larger than average subsequent saccades. A repeated measures MANOVA was performed on the post hoc factor ‘‘fixation duration’’ with the levels $\leq$180 ms versus all other fixation durations.

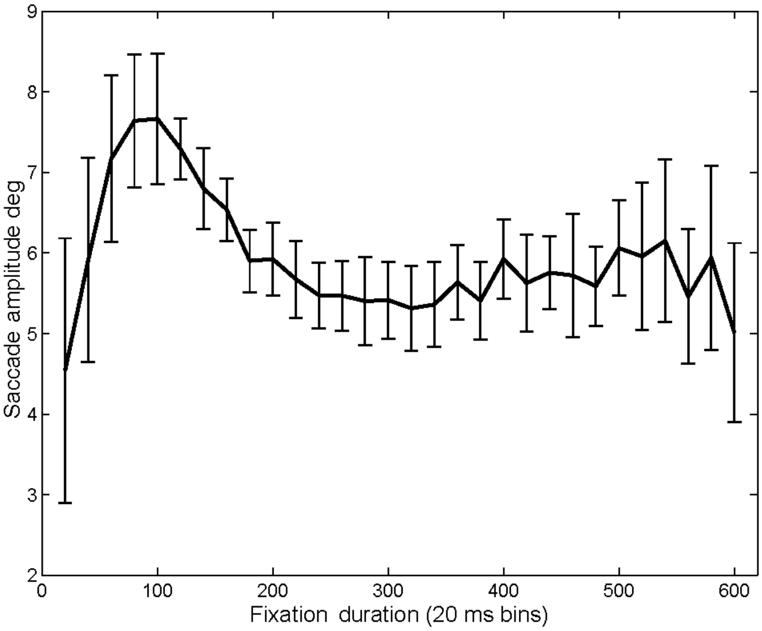

Figure 6. Saccade amplitude as a function of fixation duration.

The effect was strongly significant, $F(1, 14) = 64.76$, $p < .001$: Saccade sizes are on average 1° larger in the fixation duration range ≤ 180 ms than elsewhere. This effect seems to be omnipresent, as it is true for all subjects and is stable over time, both within the individual scene and over the whole set of images. Further analyses of this relationship follows in the discussion section.

Long-term changes

In contrast to the relatively short-term changes that occur within the inspection of a single picture, there were no significant long-term changes. The average fixation duration and saccade size did not change significantly over the total of 48 pictures. To test this, the 48 pictures were grouped in sets of 12. The effect of picture set on fixation duration was not significant, $F(3, 42) = 0.90$; nor was there an effect of picture set on average saccade size, $F(3, 42) = 0.46$.

DISCUSSION

As pointed out in the introduction, the phenomenon of gradual change of fixation duration and saccade amplitude during the first few seconds of scene inspection has been found by a number of other researchers, starting with Buswell in 1935. Explanations for this involve an orienting phase (Karpov et al., 1968), the extraction of information from the more informative regions over time (Antes, 1974; Friedman & Liebelt, 1981), and strategic task adaptation (Scinto et al., 1986). Irwin and Zelinsky (2002) reported a continuous increase in fixation duration over a 15 fixations long period. They found that, apart from the first two fixations, a close estimate of the fixation duration is given by $236.7 + 5.5 \times$ (fixation number). Although this allowed the data to be fit, it is not clear what theoretical basis justifies omitting the first two fixations. Furthermore, Irwin and Zelinsky do not offer an explanation for the increase in fixation duration. It has repeatedly been suggested that, although there are strict limitations to the information retained across saccades, some sort of representation is gradually built up over time (e.g., Irwin, 1992; O'Regan & Lévy-Schoen, 1983). Since the primary task of the visual system in initial viewing periods should be to allow visually guided behaviour (e.g., Norman, 2002), it seems plausible that in this period the fastest mechanism prevails. The ambient visual system, with its low spatial resolution, may therefore have a more prominent role in the generation of saccades in this initial period than later on. Since a low spatial frequency automatically implies larger distances, this should cause the average saccade size to be larger during the initial period.

The point in time at which the initial inventory of the spatial layout of the image is complete might be given by the moment at which the fixation durations and saccade amplitudes approach their asymptotes. The fact that the asymptotic values depend on the number of objects suggests that higher cognitive processes may play only a limited role in setting these values. Our data suggest that the mechanisms underlying the development of fixation duration and saccade amplitude over the period of inspection are similar, though not identical. Both fixation durations and saccade amplitudes develop according to a simple exponential growth curve with two parameters—the asymptote and the acceleration rate. The parameter of interest in the model presented above is the acceleration rate, since the acceleration rates of both functions do not agree with one another. Saccade amplitude appears to stabilize earlier (within ~2 s or 8 saccades) than fixation duration, which does not approach within 95% of the asymptote until 3.4 s or ~11 fixations have passed. In terms of the competitive inhibition model of Findlay and Walker (1999) of saccade control, the where-system seems to adapt to the new visual environment faster than the when-system.

Increasing the number of objects in the scene leads to an increase of competition for dominance and an increase of overall salience in the salience map. Since the average distance between the objects diminishes proportionally to the number of objects in the scene, the chance of an object winning the race for the next saccade on the basis of intrinsic salience increases as well. We should thus expect the average saccade amplitude to be less and the average fixation duration to be shorter, when more objects enter the scene. Our data show indeed that the asymptotic values of both saccade amplitude and fixation duration are affected accordingly by the number of objects in the picture. It is striking that this is true for both saccade amplitudes and fixation durations. The former seems quite obvious because of the proximity effect, but the fact that fixation durations are shorter may demand a somewhat different explanation. In terms of the fixate–move balance, one could argue that the change is either due to more activity on the ‘‘move-side’’ or to less activity on the ‘‘fixate-side’’ of the balance. The move-side obviously ‘‘pushes’’ more when there are more objects in the scene. On the other hand, higher object-density could also result in a different scanning strategy and a decrease in the level of fixate-activity might thus ensue. On the basis of this study we cannot determine which of the two possibilities offers the better explanation, however.

The second phenomenon relating saccade amplitude to fixation duration is characterized by the surprising shape of the distribution of saccadic amplitudes over fixation durations (see Figure 6). During the first 100 ms of the fixation, an increase in the likelihood of large amplitude subsequent saccades can be observed, whereas an exponential decrease toward an asymptotic level that may depend on the spatial extent of the scanned area. Therefore, it seems appropriate to consider this relationship as a ‘‘grow-and-shrink’’ process of saccade facilitation. Similar processes of activation have been described by several authors in the field of perception (e.g., Hagenzieker, van der Heijden, & Hagenaar, 1990; McClelland & Rumelhart, 1981).

One reason for a gradual increase, as opposed to an immediate maximum probability, is that the information necessary to calculate the parameters for the direction and amplitude of saccades is not available immediately, but needs to build up. Several studies have shown that there is a positive relationship between saccade amplitude and saccade latency (e.g., Cohen & Ross, 1977; Viviani & Swensson, 1982). Although this finding is robust, it has not been clear so far whether this is the result of motor planning or perceptual processing (Viviani, 1990). We will argue here that the nonlinear relationship between saccade amplitude and fixation duration results from two parallel processes. One involves the simple race for the most salient object in the scene; the other involves an inhibitory process that allows spatial selection and search selection to attenuate the salience map. Let us therefore assume that the build-up of activity can be described by:

$$sa = b \cdot e^{(-a/t)} \tag{5}$$

with b the asymptotic value, a the acceleration rate of the function, and t the time elapsed since the beginning of the fixation. Let us suppose this function reflects a set of events occurring during the course of each fixation. This means that with increasing time, the likelihood of saccades to more distant targets increases towards an asymptotic level. Let us now turn to the shrink portion of the curve. The fact that we do not find a monotonously growing relationship between saccade amplitude and fixation duration implies that some regulatory process must be in effect. After 100 ms, saccade amplitudes decline toward an asymptotic level which may depend on the density of potential saccade targets within the scanned area. We argue here that the shrink portion of the curve is actually the result of a self-inhibitory process that, too, gradually builds up its activity. Let us now assume that the final balance is achieved by an inhibitory process. For the sake of simplicity we will assume that this is an autoinhibitory process (although it is of course always possible that the same result is achieved by a heteroinhibitory process[1]) because of the notion put forth by (among others) Findlay and Walker (1999) that this is the result of some lateral inhibition process. Let us further assume that this process does not kick in immediately, but after some delay c:

$$I = b \cdot e^{-a/(t-c)} \tag{6}$$

If we further assume that the inhibition is not complete, but ends at a constant level, presumably dictated by the object density, we can predict the total activity sa of a saccade generating unit by:

$$sa = b \cdot e^{(-a/t)} - (b \cdot e^{-a/(t-c)}) + d \tag{7}$$

---

[1] Autoinhibition: a simple negative-feedback loop; heteroinhibition: negative feedback provided by an independent unit.

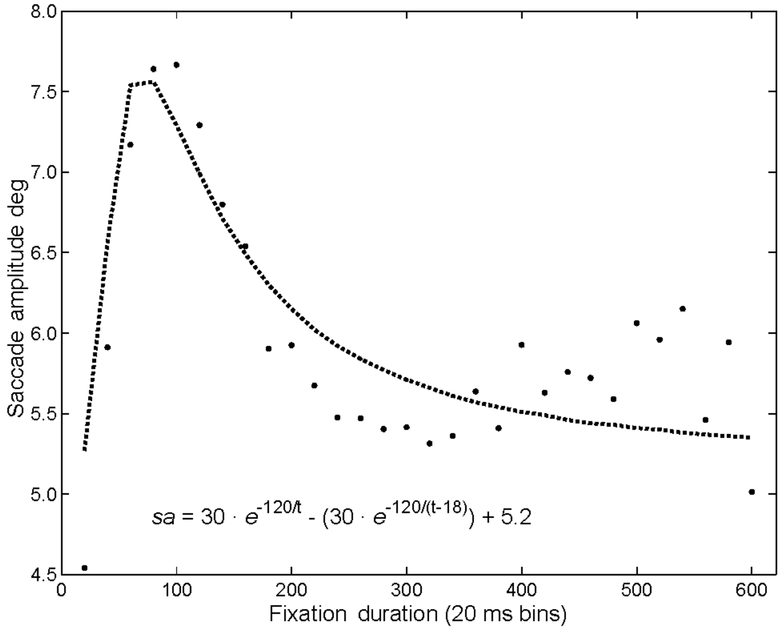

Figure 7. Actual (dots) and predicted (dotted line) values of saccade amplitude as a function of fixation duration.

Using the following parameter values:

$a = 120$, representing the smoothness of the function
b, the upper asymptote = 30°
c, the delay = 18 ms
d, a constant of = 5.2°

one finds a very reasonable fit with the actual data with an explained variance of 68% ($r = .82$). The observed and predicted saccade amplitudes on the basis of the parameters above are shown in Figure 7. Note that the upper asymptote of 30 degrees nicely agrees with maximal extent of the video display. As stated earlier, the grow-and-shrink curve is stable during the whole 20 s of picture viewing: We selected two periods of 2.5 s from the beginning and the middle of the 20 s viewing period. The predicted values correlated significantly (albeit somewhat lower than with the whole period) with both periods ($r = .59$ and $r = .69$, respectively; both $p < .001$). Similarly, we compared the amplitude-duration function of the first 12 images with that of the last 12 images and again we found somewhat lower but significant correlations with the predicted values ($r = .56$ and $r = .67$, respectively; both $p < .001$).

Apparently then, saccadic eye movements are unconstrained until the inhibitory activity is larger than the excitatory activity—which, in our data, occurs after around 120 ms from the beginning of a fixation. In terms of the model of Findlay and Walker (1999), this means that the fixate–move balance does not immediately favour fixation but after a 120 ms delay. This implies that, during this period, saccades are based on intrinsic salience rather than search selection or spatial selection. Van Diepen and d'Ydewalle (2003) found that the presentation of a peripheral visual mask in an early phase following fixation onset leads to an increase in fixation duration and a decrease in saccadic amplitude. Our model suggests that the lack of visual information necessary to produce intrinsic salience results in a decrease in the asymptotic value of saccadic amplitude. We would therefore expect that the average saccade amplitude following fixations of less than ~180 ms duration would be smaller as compared to the unmasked condition. This logic bears some resemblance to the work of Carpenter (Carpenter, 1981; Carpenter & Williams, 1995; Hanes & Carpenter, 1999) on saccade latencies. According to Carpenter's LATER-model, a race between a ''move'' signal and a ''stop'' signal towards a trigger threshold determines whether a saccade is made or not. It should be noted though, that in our case the initial saccade activity reflects the likelihood of making large amplitude saccades as compared to that of making moderate size saccades, rather than a race towards making a saccade *per se*.

Since the initial viewing period is dominated by growth towards asymptotic values of fixation durations and saccade amplitudes (Figures 2 and 4), it is tempting to consider this as the gradual transition from low-level, *automatic* control to higher level, *automated* control of eye movements. This level of control should draw upon stored spatial information as well as retinotopic representations. Our behavioural data show that memory for spatial information is not affected by the number of objects in the scene, as the percentage of correct answers to the ''where''-type questions does not change with object density. Such an ''ambient'' visual memory would be typical for a function of the dorsal ''where/how'' system. The data also show that short fixations have a greater likelihood of being followed by large saccades than longer fixations, which agrees with our previous findings (Velichkovsky et al., 2000, 2002). The way in which the transition from short-fixation/large-saccades to long-fixations/small-saccades is established suggests that two qualitatively different competitive processes, an ambient and a focal process, negotiate whether to keep fixating or to go on to the next salient object.

The present parallel model fits well with currently prevailing views of saccade control and describes how the relationship between saccadic amplitude and fixation duration in picture perception is realized. Although some consequences of the model have not been studied, e.g., how saccadic amplitudes are affected by onset transients within the grow and the shrink phase, respectively, and what the parametric consequences are of perceptually less challenging tasks such as reading, we trust that further research will show the viability of this approach.

REFERENCES

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

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

Carpenter, R. (1981). Oculomotor procrastination. In D. Fisher, R. Monty, & J. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 237–246). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Carpenter, R., & Williams, M. (1995). Neural computation of log likelihood in the control of saccadic eye movements. *Nature*, *377*, 59–62.

Cohen, M. E., & Ross, L. E. (1977). Saccade latency in children and adults: Effects of warning interval and target eccentricity. *Journal of Experimental Child Psychology*, *23*, 539–549.

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

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

Deubel, H., Schneider, W. X., & Paprotta, I. (1998). Selective dorsal and ventral processing: Evidence for a common attentional mechanism in reaching and perception. *Visual Cognition*, *5*(1–2), 81–107.

Findlay, J., Brown, V., & Gilchrist, I. (2001). Saccade target selection in visual search: The effect of information from the previous fixation. *Vision Research*, *41*(1), 87–95.

Findlay, J., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–721.

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

Hagenzieker, M. P., van der Heijden, A. H., & Hagenaar, R. (1990). Time courses in visual-information processing: Some empirical evidence for inhibition. *Psychological Research*, *52*(1), 13–21.

Hanes, D., & Carpenter, R. (1999). Countermanding saccades in humans. *Vision Research*, *39*, 2777–2791.

Harris, C., Hainline, L., Abramov, I., Lemerise, E., & Camenzuli, C. (1988). The distribution of fixations in infants and naive adults. *Vision Research*, *28*(3), 418–432.

Held, R., Ingle, D., Schneider, G., & Trevarthen, C. (1967). Locating and identifying: Two modes of visual processing. A symposium. *Psychologische Forschung*, *31*, 42–43.

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

Henderson, J. M. (1993). Visual attention and saccadic eye movements. In G. d'Ydewalle & J. van Rensbergen (Eds.), *Perception and cognition: Advances in eye movement research* (pp. 37–50). Amsterdam: North-Holland.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.

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

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics*, *64*(6), 882–895.

Karpov, B. A., Luria, A. R., & Yarbus, A. L. (1968). Disturbances of the structure of active perception in lesions of the posterior and anterior regions of the brain. *Neuropsychologia*, *6*, 157–166.

Kauffer, D., & Lewis, D. (1999). Frontal lobe anatomy and cortical connectivity. In B. Miller & J. Cummings (Eds.), *The human frontal lobes* (pp. 27–44). New York: Guilford Press.

Koch, C., & Ullman, S. (1985). Shifts in visual attention: Towards the underlying circuitry. *Human Neurobiology*, *4*, 219–222.

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

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 565–572.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, *2*(11), 547–551.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

Milner, A. D., & Goodale, M.-A. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., Bettucci, D., Mortara, F., Mutani, R., & Terazzi, E. (1991). Perception and action in visual form agnosia. *Brain*, *114*(1), 405–428.

Motter, B., & Belky, E. (1998a). The guidance of eye movements during active visual search. *Vision Research*, *38*, 1805–1815.

Motter, B., & Belky, E. (1998b). The zone of focal attention during active visual search. *Vision Research*, *38*, 1007–1022.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*(1), 73–144.

O’Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, *23*, 765–769.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in a complex task. *Vision Research*, *41*, 3587–3596.

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

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across vertical and horizontal meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*, 31–40.

Rozzolatti, G., Riggio, L., & Sheliga, B. (1994). Space and selective attention. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 231–265). Cambridge, MA: MIT Press.

Scinto, L. F. M., Pillalamarri, R., & Karsh, R. (1986). Cognitive strategies for visual search. *Acta Psychologica*, *62*, 263–292.

Trevarthen, C.-B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *31*(4), 299–337.

Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. Goodale, & R. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Diepen, P. M. J., & d’Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition*, *10*(1), 79–100.

Van Diepen, P. M. J., de Graef, P. & d'Ydewalle, G. (1995). Chronometry of foveal information extraction during scene perception. In J. Findlay, R., Walker, & R. Kentridge (Eds.), *Eye movement research: Mechanisms, processes, and applications* (pp. 349–362). New York: Elsevier.

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

Velichkovsky, B. M. (1982). Visual cognition and its spatial-temporal context. In F. Klix, J. Hoffmann, & E. Meer (Eds.), *Cognitive research in psychology*. Amsterdam: North-Holland.

Velichkovsky, B. M., Dornhoefer, S. M., Pannasch, S., & Unema, P. J. A. (2000). Visual fixations and level of attentional processing. In A. Duhowski (Ed.), *Eye tracking research and applications*. Palm Beach Gardens, NY: ACM Press.

Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhoefer, S. M., & Joos, M. (2002). Towards an express diagnostics for level of processing and hazard perception. *Transportation Research, Part F*, *5*(2), 145–156.

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (Ed.), *Reviews of oculomotor research: Vol. 4: Eye movements and their role in visual and cognitive processes* (pp. 353–383). Amsterdam: Elsevier.

Viviani, P., & Swensson, R. G. (1982). Saccadic eye movements to peripherally discriminated targets. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 113–126.

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

Zihl, J., & Hebel, N. (1997). Patterns of oculomotor scanning in patients with unilateral posterior parietal or frontal lobel damage. *Neuropsychologia*, *35*(6), 893–906.

Manuscript received November 2003
Manuscript accepted May 2004