Comparative Search Reveals the Tradeoff between Eye Movements and Working Memory Use in Visual Tasks

Shantanu Inamdar (inamdar@cs.umb.edu)
Marc Pomplun (marc@cs.umb.edu)
Department of Computer Science, University of Massachusetts Boston
100 Morrissey Blvd., Boston, MA 02125, USA

Abstract

The experiments reported here provide an insight into how the use of working memory is influenced when eye movements become 'costlier' in a visual task. In our comparative search paradigm, each half of the screen contains a column of simple geometrical objects of three sizes and forms. Participants have to detect whether the two halves of the screen are exactly identical or contain a difference. The eye movement data recorded from two experiments provides evidence for a working memory versus eye movement tradeoff. Visual working memory use – within capacity limits - is flexibly adapted to optimize task performance for varying costs of eye movements.

Introduction

The capacity of visual working memory is surprisingly small. This has been impressively shown by research on change detection in a flicker paradigm, revealing a phenomenon termed change blindness (see Simons & Levin, 1997; Simons, 2000, for reviews). In such experiments, participants are presented with two almost identical images A and B, and their task is to detect the only local difference between them; for example, one of the items shown in image A is missing in image B. However, A and B are not shown at the same time, but they alternate with short blanks separating them in time. Participants have been found to be strikingly insensitive to the changes presented to them, indicating a remarkable capacity limitation of visual working memory.

A similar finding in a different paradigm was obtained by Ballard, Hayhoe and Pelz (1995). These researchers used a task in which participants had to copy a shown pattern of colored blocks by moving blocks from a source to a workspace area. In one experiment, the task was performed on a computer screen, and in a second experiment, it was performed with real blocks on a surface. In both experiments, it was found that participants made only minimal use of working memory; for example, they tended not to simultaneously memorize the color and the position of a block. Instead, they performed additional eye movements towards the model and back to the workspace or source to acquire information only immediately before it was needed.

The interpretation of both the change blindness and block matching results is in line with models of task performance with minimal memory demands (Ballard, 1991; Brooks, 1986). The basic idea underlying these models is that the effort of building a comprehensive internal representation can be avoided if the required information is easily accessible by sensors. In other words, why should we invest time and effort in filling up our visual working memory if instead we can just 'grab' the relevant information from the visual field when we need it? Loosely speaking, it is more efficient to use the visual scene as an external memory instead of internalizing a substantial part of it. Eye movements (saccades) are very quick and therefore 'inexpensive' as compared to expensive working memory use.

If this view of visual task performance - balancing the use of working memory and eye movements based on their costs to optimize efficiency – is correct, then making eye movements more expensive should result in an increased use of working memory to reduce the number of saccades during task performance. This is exactly what we did in Experiment 1. We employed the paradigm of comparative visual search task, which has been shown to yield insight into working memory performance (Pomplun, 1998; Pomplun, Sichelschmidt, Wagner, Clermont, Rickheit & Ritter, 2001; Pomplun, Reingold & Shen, 2001). Participants had to compare two columns (hemifields) of items to determine whether there was a difference between them or not. By varying the distance between the hemifields, we varied the cost of eye movements in this task. Participants' eye movements were recorded to reveal working memory use. In Experiment 2, we tested the capacity limitation of working memory in the present context. After each switch between hemifields, the features of items were hidden for a varying amount of time, thereby artificially and drastically changing the cost of inter-hemifield saccades.

Experiment 1

To investigate the influence of the distance between object columns on working memory use, we devised a comparative visual search task employing the gaze-contingent window technique. The stimulus displays showed object columns with three levels of distance between them. This created three levels of required amplitude for inter-hemifield saccades and thus three levels of costs for eye movements. To perform the task efficiently, participants had to memorize some information from one hemifield, then switch their gaze to the other one, compare the information given there with their memory content, then memorize another chunk of information, and so on. By identifying inter-hemifield saccades in participants' gaze trajectories we were therefore able to determine the amount and time course of working memory use.

Method

Participants. Eight students from UMass Boston participated in the experiment. They were paid for the participation and did not have any information about the nature of the study.

Apparatus. Eye movements were recorded with the SR Research Ltd. EyeLink-II system, which operates at a sampling rate of 500 Hz and measures a subject's gaze position with an average error of less than 0.5 degrees of visual angle. Stimuli were presented on a 21-inch Dell Trinitron monitor with a refresh rate of 85 Hz and a screen resolution of 1152 by 864 pixels.

Materials. The stimulus displays showed two columns of simple geometrical objects on a white background. The objects were in three different forms (triangles, circles, and squares) and three different colors (fully saturated blue, green and red). They were evenly spaced avoiding item overlap and contiguity with diameters of approximately 0.95 degrees and a distance of 1.91 degrees between neighboring objects.

Each stimulus image consisted of two hemifields separated by a black line. There were 20 objects in each hemifield, which were equally balanced for form and color. The columns of objects in each hemifield were identical except for one difference (target), which were present in half of the displays. This difference could consist in the color or form of one of the objects. The two columns were at the distances of approximately 15, 30 and 45 degrees from each other for the small, medium and large distance conditions, respectively. The corresponding objects in each hemifield were connected by a black line to help participants make precise eye movements when switching between the halves and not lose track of the current row(s) in reference.

The two hemifields were presented employing the gaze-contingent moving window paradigm (Pomplun, Reingold and Shen, 2001). Only in the hemifield containing the current gaze position the objects' features were visible; gray blobs were used to mask the actual form and color of the objects in the other hemifield. As soon as the participant's gaze crossed the midline of the display, a display change was initiated and completed within a maximum duration of 14ms. This manipulation was required because, when the two columns were very close to each other it would not have been necessary for participants to switch their gaze between the columns. Instead, participants could have perceived both columns at once accurately enough to perform comparisons using covert shifts of attention. Figure 1 illustrates the gaze-contingent window manipulation.

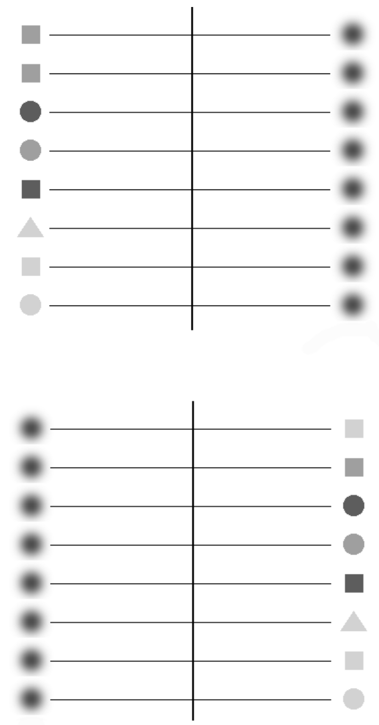

Figure 1: Sample display as seen by the participant when the gaze position is in the left hemifield (top) or in the right hemifield (bottom). The first row of items contains a color difference. Note that the actual displays contained 20 objects per hemifield.

Procedure. Prior to the start of the experiment the participants were instructed to decide for each display if it contained a difference between the two hemifields. Participants were to press a designated button on a game pad if there was a difference and press another one if there was none. They were asked to fixate a marker in the upper left region of the display before each new trial. The marker was placed at the first position of the object in the display stimulus to appear. On pressing a button, the display stimulus was displayed. This procedure served for the recalibration of the system during the experiment and the standardization of participants' initial gaze position in each trial.

After the instruction of a participant and the initial setup of the system, there were 24 practice trials to get the participant well versed with the system. These trials were followed by four blocks of 48 stimuli each. Participants were free to take breaks in between blocks if they so desired. Before continuing, after a break, the system was recalibrated to reduce error. A block consisted of 16 consecutive stimuli for each of the three distances. Out of these 16 stimuli, eight had a target and eight did not have one. The targets, if present, were at any of the 20 objects in the stimulus. The position of targets as well as the order of stimulus types were counterbalanced across blocks and participants.

Results and Discussion

Only target-absent trials with correct – negative – response were included in the data analysis for Experiment 1. There were two reasons for this restriction: First, since in target-present trials the target could sometimes be found within a few saccades, including these trials would have added substantial noise to the data. Second, it is known from previous research (Pomplun et al, 2001; see also Zelinsky, 1996) that verifying a suspected target induces eye-movement patterns that are substantially different from the ones generated during the preceding search process.

The error rate, that is, the proportion of participants' incorrect responses, did not vary significantly with the distance between the hemifields, $F(2;14) = 3.72$, $p > 0.05$. Its relatively low value (2.34%, 4.29%, and 1.36% for small, medium, and large distance, respectively) indicated that participants performed their task accurately. Interestingly, while average response time was shorter for small and medium distance (8.47s and 8.38s, respectively) than for large distance (8.73s), this difference did not reach significance, $F(2;14) = 1.28$, $p > 0.3$ (see Figure 2, top panel). As expected, however, the duration of saccades switching between hemifields did depend on the distance between them, $F(2;14) = 146.65$, $p < 0.001$. Saccade duration increased significantly with increasing distance across all its levels - small, medium, and large (58.99ms, 72.58ms, and 101.40ms, respectively) - all $t(7) > 11.23$, $p < 0.001$ (see Figure 2, center panel). Thus, eye movements indeed became more time-consuming or 'expensive' with growing distance between the hemifields. How did participants adapt to this change in costs?

First of all, the duration of participants' processing intervals – the average time that their gaze remained in the same hemifield before switching to the other one – depended significantly on the inter-hemifield distance, $F(2;14) = 13.65$, $p < 0.001$. Larger distance led to longer processing intervals (439.4ms, 466.4ms, and 531.3ms for small, medium, and large distance, respectively). While the differences between large distance and the other two distances was significant, both $t(7) > 3.03$, $p < 0.05$, the difference between small and medium distance only showed a tendency, $t(7) = 2.17$, $p = 0.065$. Regardless of the variance in the empirical data, it is evident that participants reacted to the increased cost of eye movements by spending more time on processing between switches. The difference in processing time even clearly exceeded the difference in saccade duration (see Figure 2, center panel). However, the longer processing time may also have been caused by more difficult matching (despite the horizontal lines in the displays) or increased memory decay during switching. In order to find evidence for an actual and successful increase in working memory use due to more expensive eye movements, we needed to analyze how much information was actually memorized between switches. This information was obtained by measuring the number of inter-hemifield saccades per trial.

The distance between the hemifields exerted a significant effect on the number of gaze switches between them, $F(2;14) = 13.78$, $p < 0.001$ (see Figure 2, bottom panel). With growing distance, participants switched less often between the hemifields while completing their task (17.61, 16.09, and 14.37 switches for small, medium, and large distance, respectively). There were significant differences in the number of switches across all three distances, all $t(7) > 2.44$, $p < 0.05$. Fewer inter-hemifield saccades for larger distances indicate that participants must have stored a larger amount of information in working memory in order to complete the task with the same accuracy.

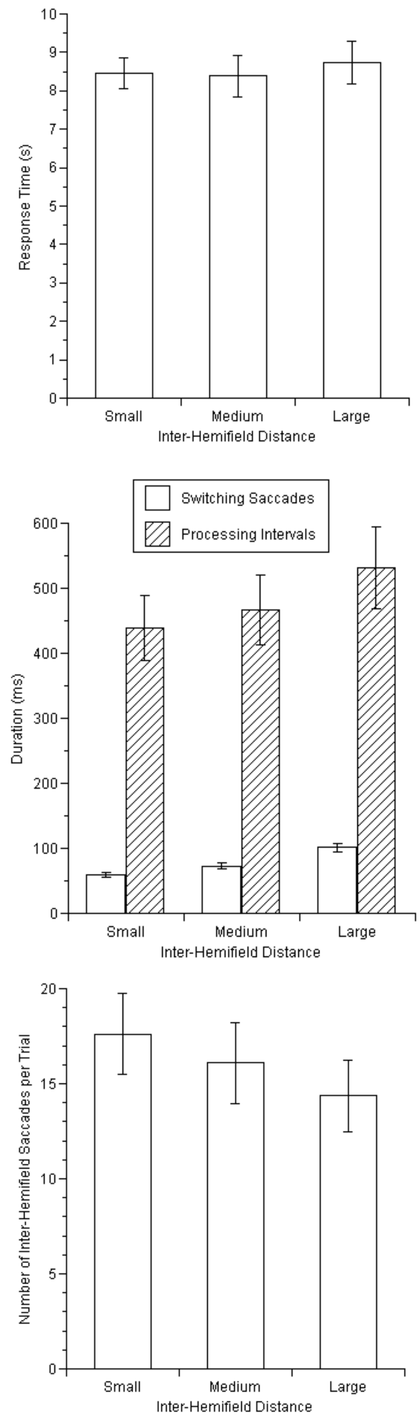

Figure 2: Response time (top), duration of processing and switching between hemifields (center), and number of inter-hemifield saccades per trial (bottom) in Experiment 1.

All in all, the results of Experiment 1 demonstrated that participants adapted their use of visual working memory to the varying cost of eye movements. Figure 3 shows sample eye movements of one of the participants for each of the three distances.

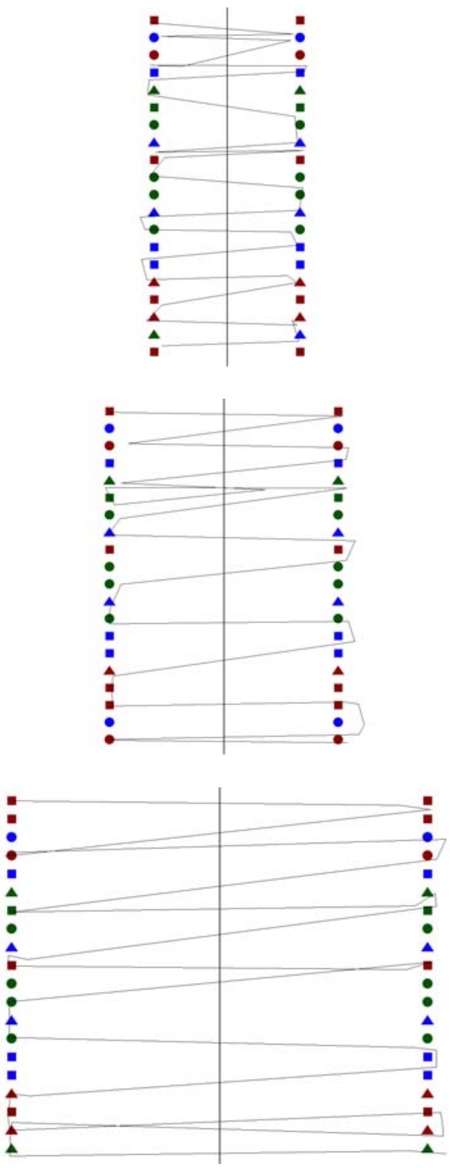

Figure 3: Eye movements of one of the participants across the three levels of costs for eye movements: Small distance (top), medium distance (center) and large distance (bottom). The lines connect successive fixations.

Experiment 2

The distance manipulation in Experiment 1 was sufficient to induce higher eye movement costs and, in turn, successful adaptation of working memory use. This was possible because the required memory load was within the capacity of visual working memory. However, what happens if the cost of eye movements is drastically increased so that working memory is used to its limit? This question was tackled in Experiment 2. Obviously, to further increase the cost of eye movements, a different approach had to be taken. We decided to introduce a variable delay between the gaze crossing of the midline and the unmasking of the objects in the currently attended hemifield. Participants knew that after the crossing of the midline they would have to wait for a certain duration until they could compare their memorized information with the one shown in the current hemifield. Consequently, each inter-hemifield saccade delayed task performance by this duration. This manipulation, although very artificial, had the advantage of allowing us to establish any desired cost (delay) of inter-hemifield saccades. We used two levels of visibility delay plus a no-delay condition to vary the cost of eye movements in the comparative search task by substantial amounts.

Method

Participants. Eight students from UMass Boston participated in Experiment 2. They were paid for the participation and did not have any information about the nature of the study.

Apparatus. The same apparatus as in Experiment 1 was used.

Materials. The stimulus display showed only the images with medium distance from Experiment 1. The two hemifields were presented employing the gaze-contingent moving window paradigm. Whenever participants switched from one hemifield to the other, objects in both hemifields were masked, and the objects in the attended hemifield appeared after a delay of 0, 500 or 1000ms.

Procedure. The procedure was the same as in Experiment 1 except that there were 12 practice trials followed by three blocks. A block consisted of 24 stimuli with eight stimuli for each of the 0ms, 500ms, and 1000ms delay conditions. Out of the eight stimuli, four had a target and the other four had no target. The position of targets as well as the order of stimulus types were counterbalanced across blocks and participants.

Results and Discussion

In analogy to Experiment 1, only target-absent trials with correct response were analyzed in Experiment 2. The error rate was comparable to the one obtained in Experiment 1 (4.16%, 2.29%, and 4.16% for visibility delays of 0ms, 500ms, and 1000ms, respectively) and was not significantly influenced by the delay, $F(2;14) < 1$. However, unlike in Experiment 1, there was a significant effect of the delay on participants' response time, $F(2;14) = 88.35$, $p < 0.001$. Response time increased significantly with longer delay across all of its three levels (12.30s, 22.68s, and 30.47s for 0ms, 500ms, and 1000ms delays, respectively), all $t(7) > 5.95$, $p < 0.01$ (see Figure 4, top panel). This demonstrates that participants were unable to completely compensate for the artificially imposed long delays between processing intervals.

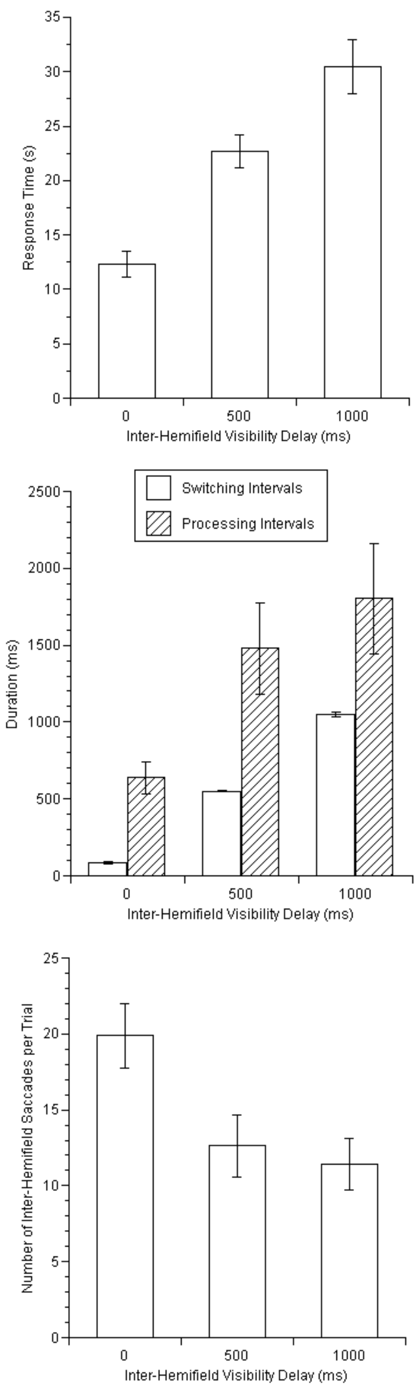

Figure 4: Response time (top), duration of processing and switching between hemifields (center), and number of inter-hemifield saccades per trial (bottom) in Experiment 2.

Since in Experiment 2 the distance between hemifields was not varied, it is not surprising that the duration of inter-hemifield saccades (84.98ms) was not significantly affected by the delay, $F(2;14) = 1.57$, $p > 0.2$. Nevertheless, it is still important to consider saccade duration in the data analysis as it determines the actual duration of the switching interval, that is, the time between two processing intervals. In the two delay conditions, the time for the delay begins to count as soon as the participant's gaze crosses the midline of the display. Assuming an approximately symmetric trajectory of saccades, the crossing occurs after about 42.5ms and is sent to the display computer with an average delay of 2ms. Due to the latency of the monitor screen, unmasking the objects after the delay takes an average of 6ms, so all in all the actual switching intervals are approximately 85ms, 550ms, and 1050ms.

These delays exerted a significant effect on the duration of processing intervals, $F(2;14) = 43.07$, $p < 0.001$. In Experiment 2, a processing interval was defined as the time from the unmasking of a hemifield until the onset of the next saccade switching to the other hemifield. Processing intervals increased with longer visibility delays (638ms, 1479ms, and 1803ms for delays of 0ms, 500ms, and 1000ms), with significant differences between all three levels, all $t(7) > 5.3$, $p < 0.05$ (see Figure 4, center panel). Interestingly, the 500ms delay caused an additional 841ms in processing time as compared to the no-delay condition, whereas the 1000ms delay increased processing time only by another 324ms relative to the 500ms delay condition. It therefore seems that participants adapted their behavior to the 500ms condition by dramatically increasing their processing intervals, while there was only little extra effort when this delay was doubled, assumedly due to the capacity limit of visual working memory. If this assumption is correct, we would expect a substantial increase in memory load between the no-delay and 500ms delay conditions and only a small increase between the 500ms and 1000ms conditions.

Accordingly, we analyzed the number of inter-hemifield saccades and found a significant effect on it by the delay, $F(2;14) = 9.11$, $p < 0.01$. While there were significantly more switches in the no-delay condition (19.88) than in the 500ms condition (12.64) and the 1000ms condition (11.43), both $t(7) > 2.73$, $p < 0.05$, the difference between the 500ms and 1000ms conditions did not reach significance, $t(7) = 1.80$, $p > 0.1$ (see Figure 4, bottom panel). This finding suggests that in the 500ms delay condition participants filled up their working memory in each processing interval to a large extent, which could not be significantly increased in the 1000ms delay condition. Figure 5 shows the eye movements of one participant for the three levels of visibility delay.

Conclusions

All in all, the present study provides evidence for a working memory versus eye movement tradeoff in visual tasks, supporting the point of view that visual scenes are used as an 'external memory' to an extent that optimizes task performance (Ballard, 1991; Ballard, Hayhoe & Pelz, 1995; Brooks, 1986). Experiment 1 demonstrates that increasing the cost of eye movements by demanding longer saccades leads to increased memory use by the participants. Memorizing more information at a time enables participants to compensate for the increased saccade duration without performing their task significantly less efficiently.

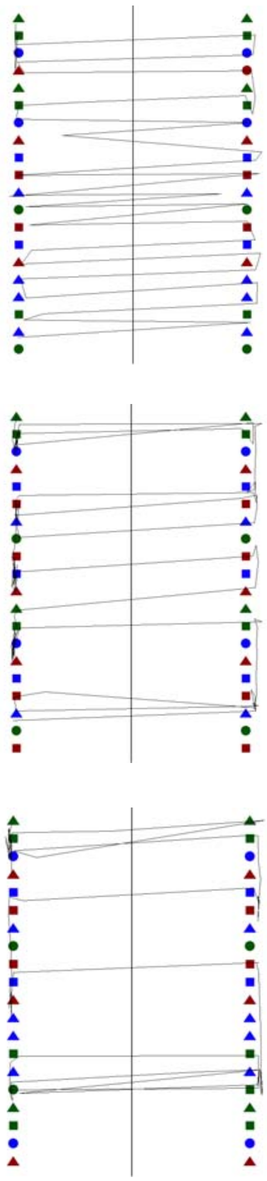

Figure 5: Eye movements of one of the participants across the three levels of visibility delay: No delay (top), 500ms delay (center), and 1000ms delay (bottom). The lines connect successive fixations.

The results of Experiment 2 show that participants adapt to the imposed 500ms visibility delay by dramatically increasing their working memory load and reducing the number of expensive inter-hemifield saccades. Adding another 500ms to the visibility delay, however, causes participants to increase their working memory load only very little. Obviously, due to the limited capacity of visual working memory, it would take participants a disproportionate amount of time and effort to memorize more information at a time in order to avoid costly inter-hemifield saccades. The flexibility of memory use for optimizing efficiency that determines the results of Experiment 1 reaches its limits under the condition of extremely expensive eye movements in Experiment 2.

According to the present data, the employment of visual working memory can be flexibly adapted to optimize task performance as long as the creation of internal representations does not exceed an estimated duration of roughly one second. Beyond this duration, the limited capacity of working memory will dramatically reduce a person's efficiency in completing a visual task.

Motivated by the present results, our future research will focus on developing a quantitative model of the working memory versus eye movement tradeoff, which was infeasible for the current data due to its substantial variance. Moreover, we will apply the current comparative visual search approach to investigate the role of memory decay in visual tasks.

References

Ballard, D.H. (1991). Animate vision. *Artificial Intelligence Journal*, 48, 57-86.

Ballard, D.H., Hayhoe, M.M. & Pelz, J.B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7, 66-80.

Brooks, R.A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2, 14-22.

Pomplun, M. (1998). *Analysis and models of eye movements in comparative visual search.* Göttingen: Cuvillier.

Pomplun, M., Reingold, E.M. & Shen, J. (2001). Investigating the visual span in comparative search: The effects of task difficulty and divided attention. *Cognition*, 81, B57-B67.

Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G. & Ritter, H. (2001). Comparative visual search: A difference that makes a difference. *Cognitive Science*, 25, 3-36.

Simons, D. J., ed. (2000), *Change Blindness and Visual Memory: A Special Issue of the Journal Visual Cognition*, (Philadelphia, PA: Psychology Press).

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1, 261-267.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research*, 36, 2177–2187.