Modeling Efficient Serial Visual Search

Bella Z. Veksler (bellav717@gmail.com)
Air Force Research Laboratory, Wright-Patterson Air Force Base
Dayton, OH 45431 USA

Wayne D. Gray (grayw@rpi.edu)
Cognitive Science Department, Rensselaer Polytechnic Institute
Troy, NY 12180 USA

Abstract

Humans perform visual search fairly efficiently, finding targets within only a few fixations. Data from eye-tracked participants was subjected to a fixation by fixation analysis to pinpoint why participants tended to make fewer fixations than would be expected by chance. The goal of this paper is to present a computational model that performs visual search as efficiently as humans. The model varied several components that may have aided visual search: memory, search strategy, and degree of parafoveal vision. Two dependent measures were used to evaluate the model: number of fixations to find the target and the distribution of saccade amplitudes. The best fitting model suggested that the biggest contribution to efficient search came from larger parafoveal vision. Search strategy, however, accounted for the distribution of saccade amplitudes.

Keywords: visual search; model; memory; parafovea

Introduction

Visual search is ubiquitous. Whether we are locating an item in the grocery store, trying to find our car in a busy parking garage, or looking for an important piece of information on a web page, visual search is involved in most every task we perform. In this paper we discuss two critical components of efficient serial visual search, *the number of fixations taken to find a target* and the *strategy* used to move the eyes around the screen. Our emphasis is on *active vision* (Findlay & Gilchrist, 2003) to examine the search strategies used by people as they search for items in their environment. The goal of the current work was to devise a computational cognitive model that was capable of reproducing human visual search efficiency. A set of process models varied different cognitive capacities theorized to affect search efficiency (i.e., deliberate strategy, memory size and parafovea size) to explore the parameter space associated with serial search efficiency.

Visual search as a paradigm has been studied meticulously for the better part of the last 50 years. In that time several notable models of visual search have been proposed (Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Wolfe, 1994). The paradigm itself consists of the detection of a target among a varying number of distractors. Search time has been found to be influenced by number of distractors (set size), similarity of distractors and targets, and number of features used to define a target (Davis & Palmer, 2004; Wolfe, 2003). The ease with which a target can be detected is often varied and response time data is typically used as the dependent measure.

While knowing how quickly visual information is found is important, understanding *how* that information is found is just as important—"vision is a tool, not the task" (Pelz & Canosa, 2001, p. 3588). For this reason, the task our participants performed was not visual search, but a decision making task that, like grocery shopping or finding the car in the parking garage, just happened to require visual search. The vast majority of visual search studies have largely ignored the process of visual search, with a few notable exceptions (Zelinsky, Rao, Hayhoe, & Ballard, 1997; Araujo, Kowler, & Pavel, 2001; Unema, Pannasch, Joos, & Velichkovsky, 2005; Zelinsky, 2008). This is problematic because "visual search is more than the time taken by an observer to detect a target and press a button. It is instead a richly complex behavior having both a spatial and temporal dynamic" (Zelinsky et al., 1997, p. 448). By relying on only response time data, visual search paradigms have essentially thrown out the spatiotemporal contingencies that propel the search process. In recent years, however, a considerable effort has been put forth to connect eye movements with the underlying cognitive process (Liversedge & Findlay, 2000).

Zelinsky (2008) analyzed eye movements from participants who searched for common household items on a tabletop. The display was limited to six stationary locations where objects could appear and on each trial there was either one, three or five items to search through. Results demonstrated that eye movements were directed towards geometric centers of progressively smaller groups of objects. It should be noted that due to the limited search display (only six possible locations and up to five items visible on any given trial) the eye movement sequences were relatively short and, in practice, limited to the first three fixations. Thus, a study which has a more complex object structure and which examines longer sequences of fixations may better elucidate the visual search process. One study that looked at longer fixation traces found that fixations and saccades progress in a coarse-to-fine strategy whereby fixation durations increase while saccade amplitudes decrease as search continues (Over, Hooge, Vlaskamp, & Erkelens, 2007). Over et al. (2007) found that participants initially attended to general properties of the search environment (i.e., the lay of the land) but, as the trial progressed, gradually paid attention to specific, detailed information.

One question we can ask is whether the layout of the display facilitates and/or guides the visual search process. Others have found that external landmarks aid visual search by reducing the number of refixations on previously viewed items (Peterson, Boot, Kramer, & McCarley, 2004; Myers & Gray, 2010). In previous work, we found that segmenting

the visual search display into perceptual clusters provides a starting point for understanding where the eyes may go (Veksler & Gray, 2011). The modeling work presented here utilizes the perceptual segmentation found in our previous work to explore the efficiency of serial visual search within this paradigm. Furthermore, two metrics are used to compare human and model data: number of fixations to locate the target and the distribution of saccades around the screen during search.

The role of memory within visual search has also been greatly debated. In some instances, researchers have inferred from response time data that memory is not utilized during search (Horowitz & Wolfe, 2003; Melcher & Kowler, 2001). In other instances, it has been shown that visual search is guided by memory for previously viewed items (Korner & Gilchrist, 2007; Peterson, Beck, & Wong, 2008, 2001), that there is some memory for the search path (Dickinson & Zelinsky, 2007), and that more new locations are searched as opposed to old (Beck, Peterson, Boot, Vomela, & Kramer, 2006; McCarley, Wang, Kramer, Irwin, & Peterson, 2003). The current work also explores the role of memory within visual search, by varying the number of previously seen items that the model avoids re-fixating during search.

In summary, we use human data and computational modeling to explore the combination of components that contribute to efficient serial visual search. The components explored include search strategy, amount of memory for previously seen items, and the effective field of view. Previewing our conclusions, the major contribution to search efficiency comes from a larger parafovea. Memory plays an important role in this task, though not as an important role as we might have expected. Search strategy was explored as human data indicated that participants did not move their eyes around the screen in a random fashion, but rather transitioned across clusters of items on the screen.

Experiment

We explore the allocation of attention during visual search when search is a subtask of a larger decision making task. The larger task was composed of the following on each trial,

1. 20 targets (represented as two-digit numbers) appeared on a *radar* screen at random locations (left-side of Figure 1). Each two-digit target subtended a 0.62° of visual angle.

2. Participants were provided with a list of six targets (right-side of Figure 1) and told to determine which target had the highest threat value.

3. Participants had to locate each one of the targets on the radar screen (visual search) and click on it with the mouse.

4. The target's threat value appeared next to the target. The delay between clicking and appearance varied between groups – 1, 2, 4, or 8 seconds.

5. Participants held the number and threat value of the target with the "highest threat value so far" in memory as they continued to locate other targets in the list of six.

6. When they decided that they had found the target with the highest threat value (usually, but not always, after an exhaustive search), they selected that target (with the mouse) in the list on the right hand side, and clicked on the Choose button (at the bottom right of Figure 1).

Although participants searched through the display for multiple targets on any given trial, for purposes of this paper, only the first search through the display (until the first search target is found) will be reported and modeled.

Method

Participants were divided into four conditions which varied the duration of how long they had to wait before information (threat value of target) appeared (1, 2, 4, or 8 s). All other aspects of the task remained the same across all participants.

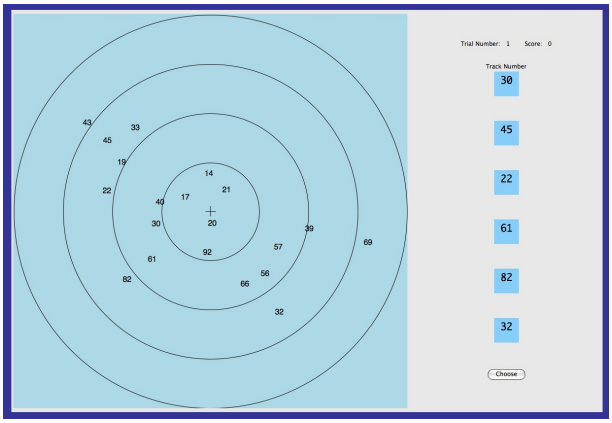

Figure 1: Task environment.

Participants A total of 88 participants from Rensselaer Polytechnic Institute were run during the study. Of those, 12 were excluded because their valid eye data fell below 90%, and two were excluded because their accuracy scores fell below 3 standard deviations of the group mean resulting in 74 participants included in the analyses (57 males). There were 19 participants in two of the conditions and 18 in the other two. The mean age of all participants was 18.8, *SD*=0.85.

Apparatus The experiment was presented using a computer running Mac OS X on a 17 inch flat-panel LCD monitor set to 1024x768 resolution, 39° x 25° of visual angle at the distance at which participants sat from the screen.The software used for the experiment was written in LispWorks 5.0. An LC Technologies eye tracker was used to collect eye data during the study at a rate of 120Hz. A chin rest was used to help ensure the accuracy of recorded eye data. Eye data quality was checked after every block of 10 trials to ensure the eye tracker was functioning and participants remained calibrated.

Procedure Participants were run separately. After signing informed consent forms, participants were given task instructions, calibrated to the eye-tracker and asked to keep their chin in the chinrest throughout the duration of the experiment. They also had to fixate a fixation cross prior to each trial to ensure the eye-tracker's accuracy.

Participants completed six blocks of 10 trials (60 trials total). A mandatory 60s break was included halfway through the study. A practice block of 5 trials was included prior to the 60 experimental trials during which time the experimenter remained in the room to ensure that participants understood how to do the task and that eye data remained valid. The experiment took $\approx$ 40 minutes to complete. Each trial proceeded as described in the beginning of the Experiment section.

Results

The majority of participants tended to search for targets in the order presented on the right hand side of the display (top to bottom). Participants tended to locate the first target they were searching for after $\approx$ 8 fixations on radar items. Since the first search in a trial was not biased by any memory effects of having found a target on a previous search, it will be used for comparison to simulation model results.

Number of Fixations to Find Target Table 1 summarizes the average number of fixations to locate the target, by condition in the study. A one-way ANOVA was conducted and indicated that there was not a significant effect of condition on either the total number of fixations to find the target, $F(3, 69) = 1.27$, $p = .29$, or the number of unique fixations to find the target, $F(3,69) = 0.63$, $p = .60$. Importantly, of the fixations shown in Table 1, roughly one target is fixated twice. This pattern suggests that participants were not necessarily maintaining all of the searched items in memory.

Table 1: Average number of total and unique fixations on radar items prior to finding first target in a trial.

Condition	N	Mean Total (SD)	Mean Unique (SD)
1	18	8.38 (1.6)	7.66 (0.86)
2	19	7.57 (1.06)	7.24 (0.75)
4	19	8.2 (1.48)	7.47 (0.93)
8	18	8.0 (1.19)	7.48 (0.87)

Collapsing over conditions, Figure 2 plots the cumulative probability of finding the first target selected. A two-way ANOVA (number of fixations as a repeated factor) was run to determine whether the lockout condition influenced search efficiency. There was a significant main effect of number of fixations, $F(43, 2967) = 2847.23$, $p < .001$, no interaction, $F(129, 2967) = 0.95$, $p = .63$, and no main effect of condition, $F(3, 69) = 0.54$, $p = .66$ on the probability of finding the target within that number of fixations. Therefore all data from the different conditions was collapsed to be used for comparison with the models.

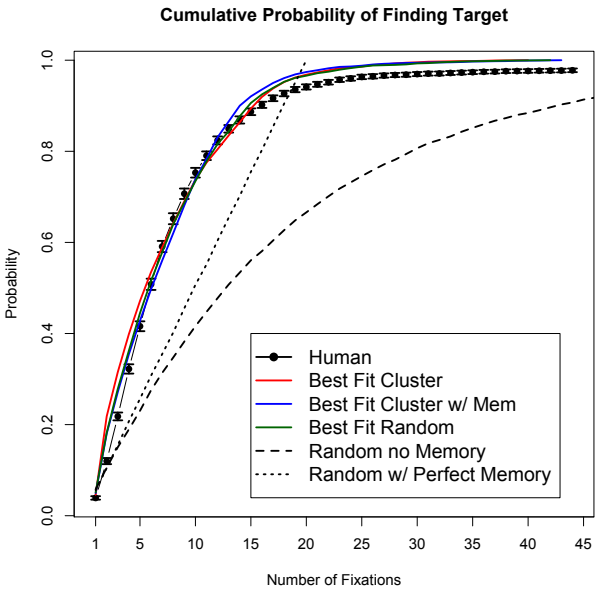

Figure 2: Cumulative Probability of Finding the Target Within Number of Fixations.

Figure 2 shows the cumulative probability of finding the initial target within N fixations, aggregated across all 74 participants. This figure also suggests that about 50% of the time participants located the target within 8 fixations. For comparison purposes, Figure 2 also shows what would be expected by chance in a model that randomly searched the radar with either no memory (dashed line) for previously seen items or perfect memory (dotted line). As can be seen, participants find the target in fewer fixations (8 on average) than would be expected by chance (10 or half of the items on the screen). This suggests that accounting only for the amount of items held in memory (so as not to refixate them) during search is insufficient to model this efficiency.

Eye Movements and Clusters In prior work, we derived a perceptual clustering algorithm which utilized human judgments of clusters to segment the display (Veksler & Gray, 2011). Participants in that study judged items to be part of the same cluster if they were within 3.28° of visual angle of each other. The algorithm adds items to a single cluster if they are less than 3.28° of visual angle apart, further adding more items that fall within 3.28° of all the items in the cluster already until no more can be added. The segmented displays generated using this algorithm were then used to determine whether search is based on clusters of targets.

Figure 3 illustrates the probability within the human eye data of a participant remaining in any given cluster given the size of that cluster. Given the eye fixation transitions, three equations were derived to fit the transition probabilities in the human data and to be later used in the model that moves its

eyes around the screen.

- Likelihood of staying in a cluster given the size of the cluster is:

$$P(Stay\ In\ Cluster) = .3292 * ln(clustersize) - .0266 \quad (1)$$

- If participants stay within a cluster, the likelihood of them looking at the closest item to the current fixation within the cluster is:

$$P(Go\ To\ Closest\ In\ Cluster) = 1.4324 * (clustersize)^{-.776} \quad (2)$$

- If participants move their gaze outside of the cluster, the likelihood of them looking at the closest item outside of the cluster is:

$$P(Go\ To\ Closest\ Outside\ Cluster) = .1888 * e^{(.0577 * clustersize)} \quad (3)$$

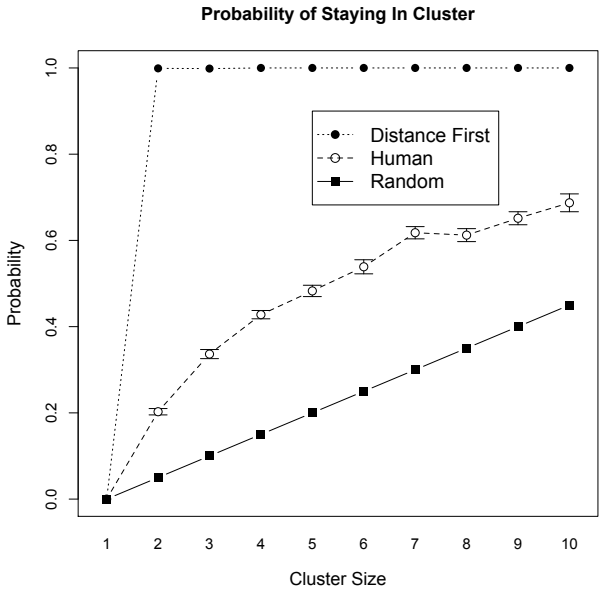

Figure 3: Probability of staying in the cluster on subsequent fixation. Distance First: prediction if participants always saccaded to closest item to current. Random: prediction if participant randomly saccaded around the screen.

Distribution of Saccade Amplitudes In addition to looking at the number of fixations that participants made to find the target, we also looked at the distribution of saccades (distances traveled by the eye between fixations). Figure 4 illustrates the distribution of saccade amplitudes in the human data (solid black line). As can be seen, the majority of saccades span about 2.26° of visual angle indicating participants moved their eyes to locations fairly close to each other. There is however, a smaller second mode around 17° indicating that participants occasionally swept their eyes across larger areas of the screen. It is beyond the scope of this paper to address these larger sweeps or when they tended to occur.

Model

Several visual search models were explored and simulated in order to model the efficiency of human serial visual search. There were three parameters that were manipulated in the modeling of the visual search process in this task. The first was the degree to which memory for previously seen items was used in the search process. The memory component essentially avoids shifting gaze to a target if it has been previously fixated within the last *N* fixations. The number of items held in memory was varied between 1(no memory)-19(perfect memory). It should be noted that even though we only looked at the first search within the trial, there may still be memory operating during search, particularly for previously searched locations.

The second parameter that was explored was the effective Field of View (FOV) that the model has. The model is able to shift its gaze to the target it is searching for if it notices it within its parafovea, typically about 2 to 6 degrees of visual angle around the current fixation (Reis & Judd, 2000). While the fovea is the high acuity region of the retina, up to about 2° of visual angle, the parafovea is a region in which acuity is not as high, with decreasing acuity as the eccentricity from the fovea increases. We explored values of 1,2,2.5, and 3 degrees of visual angle around the current fixation point providing an effective fovea+parafovea region (FOV) of 2, 4, 5, and 6 degrees, respectively.

The final manipulation had to do with the actual search strategy used. Three search strategies were explored: cluster-based, cluster-based with memory for clusters and random.

The cluster-based search model first segments the screen into several clusters based on prior empirical work (Veksler & Gray, 2011). These clusters are then used to guide the model's eye movements based on the cluster transition probabilities as per Equations 1-3. The model decides on each fixation whether or not it wants to shift attention away from the current cluster. It then decides with a certain probability to shift its gaze to either the closest item within the cluster or the closest item outside of the current cluster. The cluster-based model with memory for clusters also maintained memory for clusters it has already searched. Thus, when transitioning out of a cluster, it avoided looking to targets within previously searched clusters.

The random model search strategy is used as a baseline model. This model disregards the placement of the items on the screen and randomly chooses a target from the set of targets in the radar. The memory component was varied from a random model with no memory to one with perfect memory for targets already seen. One limitation of this model is that because the model disregards placement of items on the screen, its shifts of gaze can span long distances resulting in inefficient eye movements.

There were 19(memory store) × 4(FOV angle) × 3(strategies) models run on the radar targets used by participants in the study. Each model was run on each of the trials of human data and the number of fixations along with saccade amplitudes that were made prior to finding the first target were recorded to be compared with human data from the same set. In all, each model was run on 4559 trials.

Results

The simulations were run to determine which models could find the targets in the radar using the same number of fixations that participants used. The cumulative likelihood of finding the first target in a trial within *N* fixations was derived for each model and the human data and then compared. As an additional dependent measure, fixation transitions were recorded for each model and the distribution of saccade amplitudes was compared with human data.

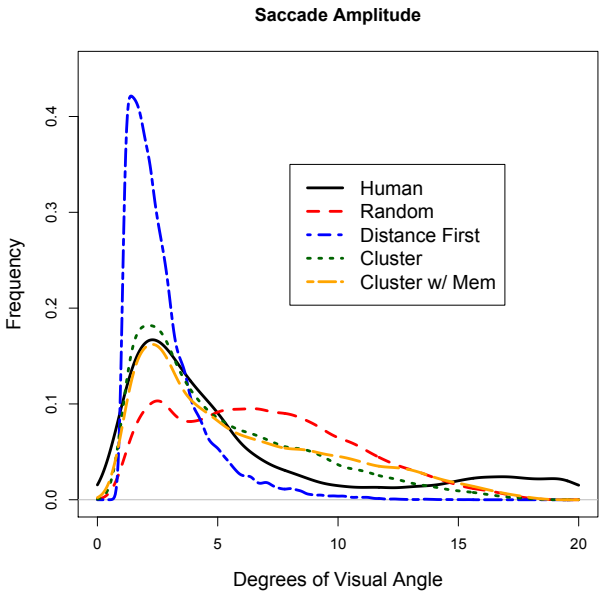

Figure 4: Distribution of saccade amplitudes across all eye data for humans and models. Models depicted are best fitting.

Search efficiency was greatly improved by the inclusion of a parafovea in all of the models (a FOV of 4, 5 or 6 degrees of visual angle). Without a parafovea (57 models), the best fit that can be achieved between human and model data has an RMSE=0.11 and an R^2=.92. The model that achieves this is the cluster search model with cluster memory and memory for 16 individual targets. If we include a parafovea, 74% of the parafovea-included models surpass this fit. Therefore, the models that were next compared all had varying degrees of a parafovea.

Based on RMSE, the top 15 models all had an effective FOV of 5 degrees. The top cluster-based search model that utilized cluster memory had a memory of 4 items, RMSE=0.018 and an R^2=.99. For comparison, the top cluster-based search model that did not utilize cluster memory needed to remember 15 items and required a FOV of 6 degrees to attain good fit, RMSE=0.0252 and an R^2=.99. The top random search model needed a memory for 14 items and a FOV of 5 degrees, RMSE=0.017, R^2=.99. Table 2 summarizes the results of the model comparisons along with baseline comparison to the two models depicted in Figure 2. For conciseness only the best fitting models are reported.

Table 2: Best fitting simulation model in each search strategy, comparing cumulative number of fixations. FOV: effective field of view in degrees of visual angle.

Search Strategy	Memory	FOV (°)	RMSE	R^2
Random	1	2	0.19	0.83
Random	19	2	0.04	0.90
Random	14	5	0.017	0.99
Cluster	15	6	0.025	0.99
Cluster w/ Mem	4	5	0.018	0.99

Table 3: Simulation models' results comparing saccade amplitude distributions. FOV: effective field of view in degrees of visual angle.

Search Strategy	Memory	FOV (°)	RMSE	R^2
Random	13	6	0.0009	.44
Distance First	16	6	0.0015	.75
Cluster	4	6	0.0004	.90
Cluster w/ Mem	8	6	0.0004	.86
Random	14	5	0.0010	.36
Cluster	15	6	0.0004	.88
Cluster w/ Mem	4	5	0.0005	.82

These results suggest that for a model to be able to search as efficiently as human participants, it needs to have some amount of a parafovea and either a large memory for individual items or a small memory for individual items along with some memory for clusters searched.

Next we compared the distribution of saccade amplitudes over the course of the search in each of the models. As an added baseline, a distance-first model was run to show what would happen if the model always saccaded to the closest item to its current point of gaze. Figure 4 depicts the human data along with the best fitting models using each of the search strategies. Table 3 provides statistics for both the best fitting models (top panel) as well as the best fitting models from the cumulative number of fixations comparison (bottom panel). In terms of modeling the distribution of saccade amplitudes, both of the cluster-based search models fit the human data well. The random search and distance first model, however, have much poorer fits.

Discussion

This work was intended to provide a computational model of the efficiency of serial visual search found in humans. Two dependent measures were used to evaluate the models generated: efficiency of search (number of fixations to locate a target) and the distribution of saccade amplitudes (how far the eye moved between fixations). It was found that incorporating a larger parafovea contributed a great deal to the efficiency with which the model was capable of finding the target. The inclusion of a memory for clusters allowed the model to have less of a need for a larger memory store for individual items searched. The cluster-based search model was also much better able to reproduce the distribution of saccade amplitudes found during human visual search, suggesting the efficacy of a search strategy based on segmentation of a display into clusters.

One limitation of the current cluster-based model and direction for future work is accounting for the longer spanning saccades as when human participants transition out of a cluster (i.e. moving to the opposite side of the screen). Another is addressing the discrepancy between the best fitting models according to the two dependent measures.

Acknowledgments

The work was supported, in part, by grant N000141010019 to Wayne Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer. Preparation of this document was performed while the main author held a National Research Council Research Associateship Award at Air Force Research Lab.

References

Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: the costs of choosing the optimal path. *Vision Research*, *41*(25-26), 3613–3625.

Beck, M. R., Peterson, M. S., Boot, W. R., Vomela, M., & Kramer, A. F. (2006). Explicit memory for rejected distractors during visual search. *Visual Cognition*, *14*(2), 150–174.

Davis, E. T., & Palmer, J. (2004). Visual search and attention: an overview. *Spatial Vision*, *17*(4-5), 249–255.

Dickinson, C. A., & Zelinsky, G. J. (2007). Memory for the search path: evidence for a high-capacity representation of search history. *Vision Research*, *47*(13), 1745–1755.

Duncan, J., & Humphreys, G. W. (1989). Visual-search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision*. Oxford Univ. Press.

Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? *Visual Cognition*, *10*(3), 257–298.

Korner, C., & Gilchrist, I. D. (2007). Finding a new target in an old display: evidence for a memory recency effect in visual search. *Psychonomic Bulletin & Review*, *14*(5), 846–851.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14.

McCarley, J. S., Wang, R. X. F., Kramer, A. F., Irwin, D. E., & Peterson, M. S. (2003). How much memory does oculomotor search have? *Psychological Science*, *14*(5), 422–426.

Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, *41*(25-26), 3597–3611.

Myers, C. W., & Gray, W. D. (2010). Visual scan adaptation during repeated visual search. *Journal of Vision*, *10*(8.).

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, *47*(17), 2272–2280.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25-26), 3587–3596.

Peterson, M. S., Kramer, A. F., Wang, R. X. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, *12*(4), 287–292.

Peterson, M. S., Boot, W. R., Kramer, A. F., & McCarley, J. S. (2004). Landmarks help guide attention during visual search. *Spatial Vision*, *17*(4-5), 497–510.

Peterson, M. S., Beck, M. R., & Wong, J. H. (2008). Were you paying attention to where you looked? the role of executive working memory in visual search. *Psychonomic Bulletin & Review*, *15*(2), 372–377.

Reis, H., & Judd, C. (2000). *Handbook of research methods in social and personality psychology*. Cambridge Univ Press.

Treisman, A. M., & Gelade, G. (1980). Feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494.

Veksler, B. Z., & Gray, W. D. (2011). A tale of two problems: human judgments of visual clusters and data collection via the web vs. paper. In *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting*. Human Factors and Ergonomics Society. Las Vegas, NV.

Wolfe, J. M. (1994). Guided search 2.0 - a revised model of visual-search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, *7*(2), 70–76.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835.

Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, *8*(6), 448–453.