

Coarse-to-fine eye movement behavior during visual search

Hayward J. Godwin · Erik D. Reichle ·
Tamaryn Menneer

Published online: 3 April 2014
© Psychonomic Society, Inc. 2014

Abstract It has previously been argued that, during visual search, eye movement behavior is indicative of an underlying scanning “strategy” that starts on a global, or “coarse,” scale but then progressively focuses to a more local, or “fine,” scale. This conclusion is motivated by the finding that, as a trial progresses, fixation durations tend to increase and saccade amplitudes tend to decrease. In the present study, we replicate these effects but offer an alternative explanation for them—that they emerge from a few stochastic factors that control eye movement behavior. We report the results of a simulation supporting this hypothesis and discuss implications for future models of visual search.

Keywords Eye movement control · Fixation durations · Visual search

There has been considerable debate surrounding the control of eye movement behavior during visual search tasks, especially in relation to fixation durations (Findlay & Gilchrist, 2003). This is important because fixation durations are a crucial component of eye movement behavior and information processing (Rayner, 2009). For example, when objects are similar to the target being searched for, fixation durations increase to enable a detailed inspection of those objects (S. I. Becker, 2011; Luria & Strauss, 1975). In the present study, we focus on how eye movement behavior is influenced by the *time course* of visual search. It has previously been reported that during a single trial of search, fixation durations tend to increase as more fixations are made (Over, Hooge, Vlaskamp, & Erkelens, 2007). This was explained with the suggestion that visual search begins with a “coarse” strategy

when a display is examined, focusing on global aspects of the display (e.g., spatial layout), but then, as the trial progresses, shifts toward a “fine” strategy, involving less rapid (and more careful) processing of individual elements in the display (e.g., stimulus features). In other words, there is a strategic change in behavior as information is gathered about the display. The presence of this coarse-to-fine strategy in visual search was further supported by the finding that saccade amplitudes (i.e., how far the eyes move between successive fixations) began by “peaking” early within a trial, initially increasing, but then decreased as the trial progressed (see also Unema, Pannasch, Joos, & Velichkovsky, 2005), which is consistent with the hypothesis that there is a shift from encoding more global to local aspects of the display. It is important to note that a distinction must be made between the overt coarse-to-fine behavior that was observed and the theoretical explanation for this behavior—namely, that the coarse-to-fine behavior was the result of a deliberate coarse-to-fine strategy.

The presence of coarse-to-fine behavior in eye movements during visual search has important implications for how eye movement data from visual search experiments are analyzed. In the best-case scenario, variability introduced by this behavior will introduce noise into the statistical analyses that are conducted but will be equivalent across experimental conditions and will not, therefore, reduce the validity of any conclusions based on those analyses. However, in the worst-case scenario, this additional variability will lead to inaccurate conclusions regarding the acceptance or rejection of experimental hypotheses. For example, consider a visual search task where participants show a tendency to fixate a given object type earlier in trials rather than a second, alternative object type, thereby causing a difference between fixation durations for the two object types by virtue of the temporal order in which they were fixated.

Therefore, gaining a better understanding of possible coarse-to-fine strategies in visual search is an important

H. J. Godwin (✉) · E. D. Reichle · T. Menneer
School of Psychology, University of Southampton, Highfield,
Southampton, Hampshire SO17 1BJ, UK
e-mail: hayward.godwin@soton.ac.uk

endeavour. The goal of the present study was to critically evaluate whether coarse-to-fine behavior can be explained using an alternative, more parsimonious account. We sought to determine whether coarse-to-fine behavior might not reflect a deliberate strategy but, instead, might emerge from the manner in which eye movement behavior is controlled during visual search and, more specifically, from a small number of stochastic factors that constrain how people move their eyes to effectively search for targets. Consideration of these factors is important for understanding the various influences on eye movement behavior during visual search. For example, it is known that saccadic targeting is not always accurate (e.g., saccades often under-/overshoot their intended targets), and these mislocated fixations are often followed by corrective saccades (see Trukenbrod & Engbert, 2007). Furthermore, the time required to program a saccade is influenced by the length of the intended saccade, typically increasing with length because longer saccades presumably require more accuracy and, thus, more time to program (Bartz, 1962; Hackman, 1940; Kalesnykas & Hallett, 1994). However, these aspects of saccadic programming and execution are generally ignored by theories of visual search.

Because the factors that may generate coarse-to-fine behavior are likely to vary in a stochastic fashion, we instantiated our alternative account of coarse-to-fine behavior within the framework of a computational model and then used this model to simulate the key aspects of eye movement data. These simulated data were then compared with experimental data obtained from a visual search task in order to evaluate the explanatory adequacy of our account. This model incorporates well-established principles of visual search that have been used in previous computational models. For example, most models of search assume that objects are selected for detailed examination (i.e., fixation) in parallel, with priority given to objects that share features with (i.e., are similar to) the target, as well as to objects that are closer to the current point of fixation (Findlay & Walker, 1999; Zelinsky & Neider, 2008). Our model incorporates this assumption, choosing between multiple potential objects to inspect in parallel. Similarly, the decisions regarding which object to fixate next proceed until either a target is identified or some number of nontargets have been inspected, consistent with response criterion assumptions that have also been adopted in previous models (Chun & Wolfe, 1996; Thornton & Gilden, 2007). Finally, our model includes very simple assumptions about factors that influence both fixation durations (e.g., saccadic programming latencies) and saccade amplitude (e.g., saccadic error). A complete formal description of our model is presented below.

Finally, it is very important to emphasize that we are not disagreeing with the existence of coarse-to-fine behavior as an empirical phenomenon; we are simply providing an alternative account for why it occurs. Furthermore, because it is difficult to make a priori predictions regarding how several

stochastic factors might interact in visual search to modulate fixation durations and saccade amplitudes, it was necessary to instantiate our account as a computational model to evaluate the explanatory adequacy of our account. The latter point is important because our goal was *not* to provide a comprehensive new model of visual search but was, instead, much more modest—to demonstrate that it is not necessary to incorporate a deliberate coarse-to-fine strategy into future models of eye movement behavior and visual search.

Method

Participants

Sixteen 21- to 25-year-old participants (11 females) from the University of Southampton took part in the study.

Apparatus

Eye movement behavior from the right eye was recorded using an Eyelink 1000 running at 1000 Hz. A nine-point calibration was used and accepted if the mean error was less than 0.5° of visual angle, with no error exceeding 1° of visual angle. Drift corrects were performed before each trial.

Participants were seated, in a dimly lit room, 71 cm from the display with their head stabilized using a chinrest, and responses were collected using a game pad. Stimuli were presented on a 21-in. CRT monitor with a 100-Hz refresh rate and a $1,024 \times 768$ pixel resolution.

Stimuli

The stimuli were *T* and offset *L* shapes, with the target being a *T* shape and the *L*s being distractors. The shapes subtended approximately 1.5° of visual angle. The *T* and *L* shapes were randomly rotated by 0° , 90° , 180° , or 270° and then placed at random upon a virtual 5×4 grid ($29^\circ \times 22^\circ$). On each trial, 16 stimuli were presented. The shapes were displaced $\pm 0.57^\circ$ – 2.57° (sampled from a uniform distribution) up/down and left/right, making their arrangement appear random. The colors of the shapes were selected from 16 points in CIE_xy color space that had previously been used in similar search tasks (e.g., Stroud, Menneer, Cave, & Donnelly, 2012). The number of distractors of each color was equated across the experiment, but the specific colors were randomly selected during each trial.

Design and procedure

The task was to search for a *T* among offset *L*s. The *T* target was of a given color that was consistent from trial to trial. The

experiment consisted of 640 trials, preceded by 20 practice trials. A single target was presented on 320 trials. Trials began with a drift correct procedure, after which participants were presented with a centrally located reminder of the target, which they had to fixate for the trial to begin. Participants had unlimited time to respond.

Because the similarity between the target color and the distractors is an important aspect of this task, we followed precedent (e.g., Stroud, Menner, Cave, Donnelly, & Rayner, 2011) and asked each participant to search for a different-colored target. Thus, our results are not limited to a single color but, instead, generalize to all colors.

Results

Data processing

We removed any fixations with a duration of less than 60 ms or greater than 1,200 ms (5 % of raw data set fixations) as outliers. We also removed incorrect-response trials and any fixations that coincided with a response. In addition, to avoid including fixations that were elongated or modulated by response preparation, we removed fixations that landed on a target or that were preceded by or directly followed by a fixation on the target. It is worth noting that previous studies of coarse-to-fine behavior included the first fixation on displays. However, first fixations were excluded here because they typically began during the presentation of the target reminder preceding each trial, causing only part of the fixation to occur on the display itself. These fixations tended to be elongated, and it was not possible to determine whether they were elongated as a result of coarse-to-fine behavior or because of saccadic inhibition, which involves the lengthening of fixation durations after display changes (Reingold & Stampe, 2002). Finally, because there were fewer observations for ordinal fixation numbers ten and above, we limited our analyses to the first nine ordinal fixation numbers. The final data set consisted of 28,541 fixations.

Analytic approach

Response accuracy was high (97 %), so our analyses focused on replicating the pattern of results that have been interpreted as evidence for a “coarse-to-fine” strategy in search. Specifically, this pattern involves increasing fixation durations coupled with decreasing saccade amplitudes across a trial. We then conducted a simulation, using our model to determine whether this pattern of results could also be explained using our alternative account.

Coarse-to-fine behavior: Fixation durations and saccade amplitudes

We began by examining whether fixation durations increased and saccade amplitudes first increased and then decreased as a function of ordinal fixation number (Fig. 1), using two separate within-subjects ANOVAs. Both analyses indicated that there was an effect of ordinal fixation number [fixation durations, $F(8, 120) = 4.5, p < .0001$; saccade amplitudes, $F(8, 120) = 9.8, p < .0001$]. To minimize the number of post hoc comparisons, we then conducted a series of Bonferroni-corrected paired t -tests comparing both dependent variables at the first, fifth, and ninth ordinal fixation numbers. For fixation durations, these contrasts indicated that mean fixation durations increased between the first and fifth ordinal fixation numbers, $t(15) = 5.9, p < .001$, and between the first and ninth ordinal fixation numbers, $t(15) = 3, p = .024$, but showed no difference between the fifth and ninth ordinal fixation numbers ($t < 1$). Contrasts also indicated that saccade amplitudes increased between the first and fifth ordinal fixation numbers, $t(15) = 20.4, p < .0001$, and showed a trend for a decrease in saccade amplitudes between the fifth and ninth ordinal fixation numbers, $t(15) = 2.5, p = .069$, resulting in no difference between the first and ninth ordinal fixation numbers, $t(15) = 1.2, p = .66$.

Our analyses thus replicate the pattern of findings (e.g., Over et al., 2007) that have been taken as evidence for a coarse-to-fine strategy in visual search, with fixation durations increasing as a function of ordinal fixation number and saccade amplitudes first increasing but then showing a trend toward decreasing later in the trial. Furthermore, despite the fact that the number of fixations per trial was lower than that in previous research examining coarse-to-fine behavior (e.g., Over et al., 2007), our stimuli were similar to those used in the majority of visual search studies, making the theoretical implications of our results applicable to those studies.

Simulating our alternative account

Because our results replicate the basic pattern of findings that have been interpreted as evidence for a coarse-to-fine search strategy (Over et al., 2007), we can now provide an alternative account of those findings, using our results to evaluate this account. To do this, we first describe the four basic principles of our account (which are independently motivated from several other areas of visual-cognitive research and formally implemented as a Monte Carlo simulation model). These principles identify six basic parameters that are widely accepted as determining eye movement behavior in visual search and are thus founded in previous research. These parameter values are therefore estimated to be the factors necessary to effectively simulate visual search behavior.

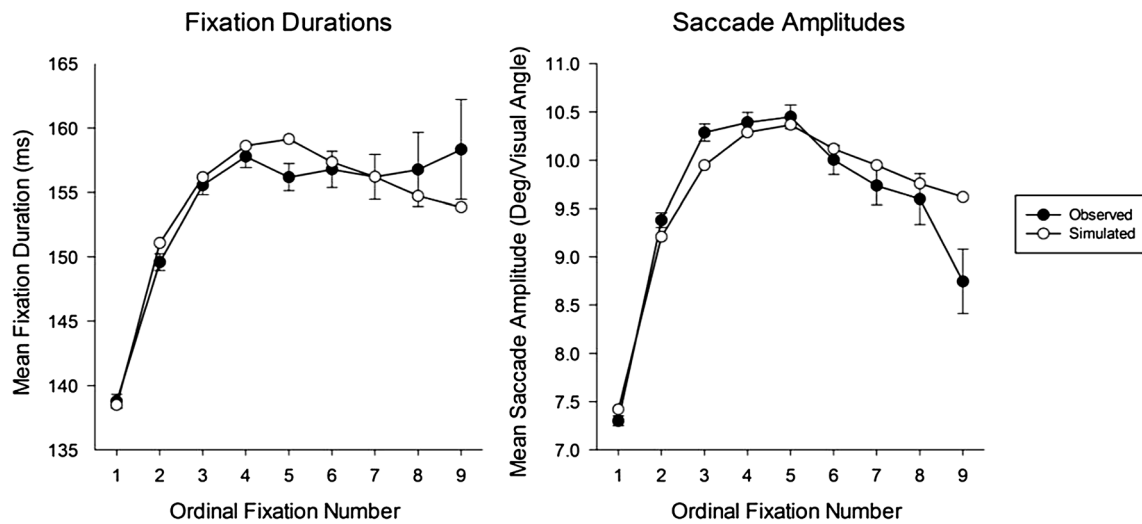


Fig. 1 Mean observed and simulated fixation durations (left panel) and saccade amplitudes (right panel, in degrees of visual angle) for distractors as a function of ordinal fixation number. Error bars represent $\pm SEM$

The first principle is that stimuli are selected to be saccade targets probabilistically, with the probability of any given stimulus y being selected as the target of the next saccade being specified by

$$p(\text{stimulus } y = \text{saccade target} \mid \text{fixation location } x) \quad (1)$$

$$= \frac{\Delta_{\max(x,i)} - \Delta_{x,y}}{\sum_{i \in N} (\Delta_{\max(x,i)} - \Delta_{x,i})}$$

where Δ is the Euclidean distance (measured in degrees of visual angle) between any two locations, x is the current fixation location, and i is an index of the set of N “eligible” saccade target locations. (What determines whether a stimulus is eligible as a saccade target is discussed below.) Saccade targets are therefore selected using a version of the Luce (1959) choice ratio wherein the probability of any given stimulus being selected as a saccade target is both inversely related to its distance from the current fixation location and proportional to the relative distances of all other eligible saccade target locations. This effectively means that stimuli near the current fixation location will have a higher probability of being selected than those far from the current fixation location, but with the most distant stimulus having a $p = 0$ probability of being selected.

The second principle is related to the eligibility of stimuli as saccade targets: To be included in the set of N possible saccade targets, a stimulus must not be in the set of κ previously fixated stimuli, under the assumption that the locations of these stimuli are either actively maintained in a short-term visual buffer (Pylyshyn, 2003) or actively inhibited (e.g., via inhibition of return to previously fixated locations; Zelinsky & Neider, 2008) to prevent the same stimuli from being the

targets of repeated saccades. However, because saccades are prone to motor error (which was simulated by sampling from a two-dimensional Gaussian distribution with a standard deviation of σ degrees of visual angle), only fixations that land within a certain distance of a stimulus (specified by a radius of θ degrees of visual angle) are considered to be “fixated” and, thus, included as one of the κ locations being maintained in the short-term visual buffer. Fixations outside of this radius result in another saccade being initiated, with a bias to move the eyes toward the original saccade target because of its close spatial proximity to the mislocated fixation, but with the actual target being selected probabilistically (as described by Eq. 1). When the number of previously fixated locations increases to the point of equaling the visual buffer capacity (i.e., κ), the oldest location is removed from the buffer (i.e., forgotten or “overwritten” by the currently fixated stimulus) so that stimuli that were previously fixated during the early part of a trial have some probability of being refixated.

The third principle is that fixation durations directly reflect the amount of time that is required to both encode a stimulus (τ_1) and program a saccade (in milliseconds, as a function of τ_2) and that the latter time increases linearly with the length of the intended saccade. This first part of this principle is consistent with the well-established fact that, even in a complex task like reading, very little time (e.g., 60 ms) is necessary to extract visual information from a display (Rayner, Liversedge, White, & Vergilino-Perez, 2003). The second part of this principle is consistent with empirical observations that fixation duration increases with the length of the next saccade (Bartz, 1962; Hackman, 1940) and the hypothesis that longer saccades require more time to program because the targets of those saccades are further away, thereby

necessitating greater saccade accuracy. Thus, in our model, fixation durations are a function of the intended length as specified by

$$\text{fixation duration} = t_1 + t_2(\text{intended saccade length}). \quad (2)$$

Finally, the last principle is simply that participants in our visual search task do not necessarily inspect all of the stimuli in a display but, instead, typically examine a subset of those stimuli (specified by the parameter N_{stim}) because some of the stimuli are so dissimilar to the target that they can be ruled out as targets without inspection (e.g., a red distractor when looking for a blue target; Treisman & Gelade, 1980) and/or because the participants adopt an increasingly liberal criterion for responding “target absent” as a trial progresses and an increasing number of distractors are inspected without locating a target (Thornton & Gilden, 2007).

The model as described above was used to simulate our visual search experiment after multiple grid searches of the model’s parameter space were completed to identify the parameter values that minimized the mean absolute percentage difference between each of the observed and simulated means displayed in Fig. 1. The final simulation results shown in Fig. 1 are thus based on 10,000 simulated trials and the following parameter values: $\kappa = 4$, $\sigma = 0.9$, $\theta = 1.9$, $\tau_1 = 87$, $\tau_2 = 7$, and $N_{\text{stim}} = 7$. As the figure shows, the model accurately predicted the observed means, with a mean absolute percentage observed-simulated deviation of 1.74 %. It is also important to note that the best-fitting parameter values are concordant with prior estimates of those parameters. For example, the value of κ is in close agreement with previous research suggesting that a short-term visual buffer can maintain four or five stimuli (Pylyshyn, 2003). Similarly, the values of τ_1 and τ_2 produce saccadic latencies of 110–260 ms, consistent with the observed range in visual search (Rayner, 2009). Finally, the values of σ , θ , and N_{stim} are consistent with observations that participants occasionally fixate between stimuli and rarely examine all stimuli prior to responding (Findlay & Gilchrist, 2003).

Discussion

Eye movement behavior during visual search exhibits a tendency for fixation durations to increase and saccade amplitudes to initially rise to a “peak” and then decrease as a trial progresses (Over et al., 2007). One explanation for this is the adoption of a coarse-to-fine strategy in which participants first gather information about global aspects of the display and, as the trial progresses, then shift toward gathering detailed aspects of the display by inspecting individual elements (Over et al., 2007). This pattern of data has important implications for the analysis of visual search experiments that have

recorded eye movements and any theoretical conclusions based on those analyses, thus motivating our present attempt to understand this behavior in more detail.

Our goal was to determine whether the apparent coarse-to-fine pattern of behavior that was previously reported could, instead, emerge from a small number of well-established properties of eye movement behavior. However, as was indicated in the introduction, because it is difficult to predict how several stochastic factors interact with one another to produce such complex behavior, it was necessary to first conduct a visual search experiment to replicate the coarse-to-fine behavior and to then use the data from that experiment to evaluate a more parsimonious account that does not require a shift in strategy on behalf of the participants. This alternative account was formally instantiated as a computational model.

Our model accurately simulated the data that we obtained from our experiment. This is important because it demonstrates how three simple principles for selecting saccade targets (i.e., probabilistic selection biased toward proximal stimuli, inhibition of return to a small number of previously fixated stimuli, and bias against inspection of dissimilar stimuli) in conjunction with basic findings about saccadic programming and execution (fixation duration increases with next saccade length; Bartz, 1962; Hackman, 1940) might explain empirical phenomena that were originally interpreted as evidence that visual search is mediated by a complex coarse-to-fine strategy (Over et al., 2007). But perhaps more importantly, our account provides a simple theoretical framework for thinking about and simulating the patterns of eye movements that are observed during visual search tasks. One merit of this theoretical framework is that it is parsimonious, being based on simple principles that are widely applicable across a variety of visual-cognitive task domains (e.g., reading; Reichle, Pollatsek, & Rayner, 2012). Another merit is that the framework is specified such that each of its assumptions can be directly related to psychological processes and/or behaviors that can be measured (e.g., saccadic programming latencies; W. Becker & Jürgens, 1979).

Finally, it is worth mentioning that, although we have developed an alternative account of coarse-to-fine behavior in visual search, the original patterns (Over et al., 2007) still stand and have been replicated here, demonstrating robust systematic modulations of fixation durations and saccade amplitudes as a function of ordinal fixation number during visual search. This is important to consider because it has theoretical implications for our understanding of eye movement behavior during visual search. The pattern of data also has implications for the development of models of visual search in terms of how they simulate this empirical pattern.

Author Note The authors would like to thank Robin Walker and two anonymous reviewers for their comments on an earlier version of this article. H.G. was supported by funding from the Economic and Social

Sciences Research Council (grant ref. ES/I032398/1). E.D.R. was supported by funding from the National Institutes of Health (R01HD053639).

References

- Bartz, A. E. (1962). Eye-movement latency, duration, and response time as a function of angular displacement. *Journal of Experimental Psychology*, *64*(3), 318–24.
- Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity of perceptual difficulty? *PLoS One*, *6*(3), 1–5. doi:10.1371/journal.pone.0017740
- Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*(9), 967–983. doi:10.1016/0042-6989(79)90222-0
- Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*(1), 39–78. doi:10.1006/cogp.1996.0002
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.
- Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *The Behavioral and Brain Sciences*, *22*(4), 661–74. doi:0140-525X/99
- Hackman, R. B. (1940). An experimental study of variability in ocular latency. *Journal of Experimental Psychology*, *27*, 546–558.
- Kalesnykas, R. P., & Hallett, P. E. (1994). Saccades and the Latency of Eye, *34*(4), 517–531. doi:10.1016/0042-6989(94)90165-1
- Luce, R. (1959). On the possible psychophysical laws. *Psychological Review*, *66*, 81–95. doi:10.1037/h0043178
- Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception and Psychophysics*, *17*(3), 303–308. doi:10.3758/BF03203215
- 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), 2272–80. doi:10.1016/j.visres.2007.05.002
- Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461
- Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science*, *14*(4), 385–8. doi:10.1111/1467-9280.24483
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in non-reading tasks. *Psychological Review*, *119*, 155–185. doi:10.1037/a0026473
- Reingold, E. M., & Stampe, D. M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *Journal of Cognitive Neuroscience*, *14*(3), 371–88. doi:10.1162/089892902317361903
- Stroud, M. J., Menneer, T., Cave, K. R., & Donnelly, N. (2012). Using the dual-target cost to explore the nature of search target representations. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 113–122. doi:10.1037/a0025887
- Stroud, M. J., Menneer, T., Cave, K. R., Donnelly, N., & Rayner, K. (2011). Search for multiple targets of different colours: Misguided eye movements reveal a reduction of colour selectivity. *Applied Cognitive Psychology*, *25*(6), 971–982. doi:10.1002/acp.1790
- Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, *114*(1), 71–103. doi:10.1037/0033-295X.114.1.71
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi:10.1016/0010-0285(80)90005-5
- Trukenbrod, H. A., & Engbert, R. (2007). Oculomotor control in a sequential search task. *Vision Research*, *47*(18), 2426–43. doi:10.1016/j.visres.2007.05.010
- 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. doi:10.1080/13506280444000409
- Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, *16*(5), 553–566. doi:10.1080/13506280802000752