UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Bounded Optimal State Estimation and Control in Visual Search: Explaining Distractor Ratio Effects

Permalink

https://escholarship.org/uc/item/5kd2h0bw

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN

1069-7977

Authors

Myers, Christopher
Lewis, Richard
Howes, Andrew

Publication Date

2013

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Bounded Optimal State Estimation and Control in Visual Search: Explaining Distractor Ratio Effects

Christopher W. Myers (christopher.myers.29@us.af.mil)
Air Force Research Laboratory
Wright-Patterson AFB, OH USA

Richard L. Lewis (rickl@umich.edu)
University of Michigan
Ann Arbor, Michigan USA

Andrew Howes (howesa@bham.ac.uk)
University of Birmingham
Birmingham, UK

Abstract

We demonstrate that an ideal observer model bounded by known limitations of the human visual system can explain empirical evidence concerning two effects of distractor ratios on visual search—effects that have previously been explained with salience-based models. The model makes optimal state estimations based on Bayesian estimates of stimuli localization and optimal control decisions of where to fixate in order to maximize task performance. Analysis of the model's behavior under different task strategies and different constraints on the visual system reveal which aspects of the model are responsible for the effects: the distractor-ratio effects on number of fixations is a signature of optimal state estimation in the face of noisy spatial information, and the saccadic-bias effect is a signature of both optimal control and estimation under these same bounds.

Keywords: optimal state estimation, optimal control, ideal observer models, visual salience, distractor ratios

Introduction

Visual search is so ubiquitous that we probably hardly notice ourselves doing it. We search for our car keys on a cluttered desk, for our family at the market, or for a reference in text. The current paper addresses how one adapts during visual search by determining what information to visually inspect. We address each of these questions through the development of a model based on the optimal integration of perceived information given a set of known constraints on the human visual system.

In a review of the literature on eye movements, Kowler (2011) describes two general approaches to modeling visual search processes. First are *map-based* approaches, such as salience maps (Itti, 2006; Itti & Koch, 2000) and activation maps (Pomplun, 2003; Wolfe, 2007), where information is accumulated and processed to produce a topogrpahical map. Peaks in the map represent areas/items that differ from their surround, that contain attributes of the target, or both. Map peaks are used to guide search through the display using some peak selection routine, such as a greedy heuristic (Pomplun, 2003) or winner-take-all algorithm (Itti & Koch, 2000). In general, the map-based approach assumes that saccades are programmed to move the fovea to an area of a stimulus that stand out from its surroundings or that is similar in some way to a search goal.

Alternatively, *visibility models* (Kowler, 2011) such as ideal observer/searcher models (Geisler, 2011; Myers, Gray, & Sims, 2011; Najemnik & Geisler, 2008; Baron & Kleinman, 1969), assume that saccades are programmed to direct foveated vision to areas of impoverished acuity in order to maximize information gained in service of task performance. Najemnik and Geisler (2005) found that the number, and spatial distribution, of saccades to find a target could be predicted by a model in which each saccade was directed to the ideal location (i.e., the highest probability of finding the target). Their model was sensitive to known human constraints on vision (e.g., decreasing contrast sensitivity with increasing eccentricity). Hence, saccadic selectivity could be considered a process that maximizes search performance by considering the effect of the eyes' subsequent fixation location.

In the current paper we build on the ideal observer approach by deriving a *boundedly optimal adaptive visibility model* capable of capturing empirical phenomena that demonstrate adaptation to changes in the proportion of available environmental features. More specifically, we use the model to explain phenomena associated with the distractor ratio paradigm (Bacon & Egeth, 1997; Shen, Reingold, & Pomplun, 2000; Zohary & Hochstein, 1989)—phenomena that have previously been given interpretations in terms of salience maps. Key to this explanation is the incorporation of constraints on the representation of spatial information in the periphery into an ideal observer analysis .

The adaptive visibility model has a simple structure that decomposes visual search into *optimal state estimation* (the integration of perceptual evidence into a task-relevant representation of the external stimulus) and *optimal control* (the choice of overt task responses and information gathering actions; Stengel, 1994). The model incorporates a small number of constraints intended to abstractly characterize important properties of a noisy, foveated vision system (Tanner, 1961). Bayesian state estimation is used to optimally integrate the noisy percepts across fixations in service of two control decisions: 1) where to next fixate and 2) when to issue a task response. Both the estimation and control processes are adapted to the simultaneous constraints of the vision system and the task at hand.

The structure of this model affords the formulation and exploration of a number of interesting theoretical questions concerning visual search phenomena. In this paper, we use modeling to determine whether distractor ratio effects are signatures of optimal state estimation, optimal control, (or both), and to identify the constraints of the visual system that are necessary for the effect to arise. To foreshadow the two key results, the model demonstrates (1) that distractor ratio effects may be understood as adaptation to changes in proportions of task-relevant environment features, and that these effects are signatures of optimal state estimation (not control) in the face of spatial uncertainty in the parafovea; and (2) that saccadic bias may be understood as a signature of both optimal control and optimal state estimation in the face of spatial uncertainty.

In the following sections we first discuss efficient visual search in the distractor ratio paradigm and introduce the boundedly optimal adaptive visibility model. We next discuss model results and their implications.

Distractor Ratio Paradigm

The distractor ratio is the ratio between distractor sets that share features with a target for a fixed number of items on a display. For example, the distractor ratio when searching for a conjunction of a color and a shape (e.g., red O) in a display of 48 items is the number of distractors that are the same color relative to the number of distractors that are of the same shape—same-color:same-shape. Hence, the distractor ratio for Figure 1(A) is 3:45, (B) is 24:24, and (C) is 46:2. Subjects are typically instructed to respond appropriately if they determine that a target is present or absent for each trial.

The distractor ratio paradigm has been used to distinguish between endogenous and exogenous influences on saccadic selectivity processes (Bacon & Egeth, 1997; Zohary & Hochstein, 1989). Exogenous influences are hypothesized to arise from the statistical properties of the visual environment, such as salience (Itti, 2006; Itti & Koch, 2000), whereas endogenous influences are those that stem from knowledge brought to bear on the task through instructions (Yarbus, 1967) or learned during performance (Myers et al., 2011). Regardless of the endogenous/exogenous process distinction, results from distractor ratio experiments demonstrate adaptation to the changing structure of the search environment. Specifically, subjects prefer to actively search through the minority set of distractors that share a common feature with the target. Using eye-tracking, Shen et al. (2000) showed that subjects searching for a target (e.g., red O) preferred the same-color distractors (red X's in Figure 1A), yet adaptively shift this preference to same-shape distractors (i.e., green O) when presented with a distractor ratio where shape was the minority feature (e.g., Figure 1C). Importantly, this adaptation reduced response times and the number of fixations to

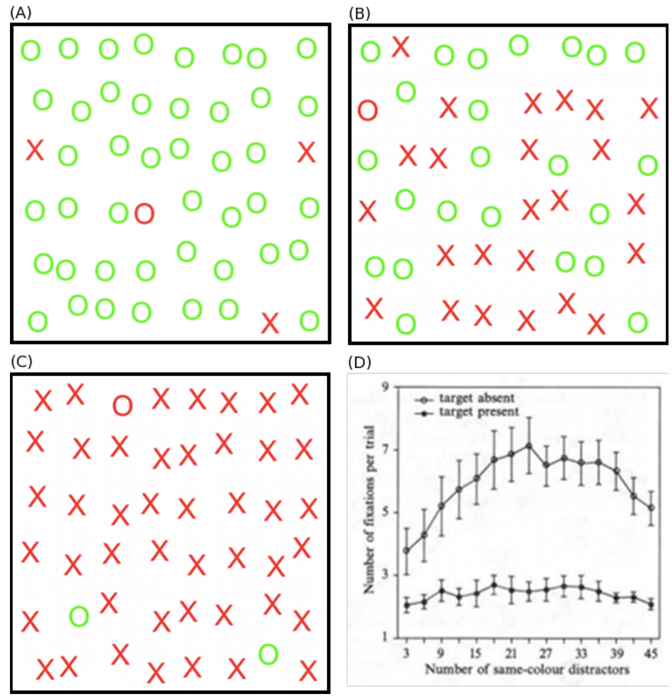

Figure 1: Distractor ratio stimuli when searching for a red O, and results from Shen et al. (2000). Panel (A) is a stimulus containing three distractors that share the same color feature as the target. Panel (B) has equal number of like-color and like-shape distractors. Panel (C) has two like shape distractors. Panel (D) demonstrates a ∩-shaped curve associated with an increasing number of same-color distractors for target absent (open circles) and target present (filled circles) trials, and represents the distractor-ratio effect.

locate the target (see Figure 1D), improving search efficiency (Bacon & Egeth, 1997; Zohary & Hochstein, 1989).

Saccadic selectivity in the distractor ratio paradigm demonstrates rational adaptation from the standpoint that subjects minimize their time to locate a target (c.f. Gray, Sims, Fu, & Schoelles, 2006). Hence, response times and the number of fixations are minimal when a search stimulus has a minority feature (e.g., color or shape; see Figure 1A) relative to when the distractor ratio is equal to one (see Figure 1B) for target-present and target-absent trials (see Figure 1D). Interestingly, Shen et al. (2000) report that saccadic selectivity favoring the minority feature occurred as early as the very first saccade in a trial.

One potential explanation for the distractor ratio effect is that saccadic selectivity is exogenously influenced through stimulus salience (Theeuwes, 1993). In Figure 1A, the red X distractors stand out from the surrounding green O distractors. The reverse is true for Figure 1C. Hence, the salience approach predicts saccadic selectivity favoring the red X's in Figure 1A and the green O's in Figure 1B. Importantly, the inclusion of some inhibition of return mechanism (IOR; Klein,

2000) is a required addition to salience-based models in order to eliminate endlessly fixating the most salient areas of the display, which are not guaranteed to contain the target. Importantly, the IOR and salience mechanisms would be capable of not only reproducing an important hallmark of adaptive search in the distractor ratio paradigm (the ∩-shaped curves depicted in Figure 1D), but also another hallmark: saccadic bias in favor of the minority distractor set.

While the salience+IOR approach provides a potential explanation of adaptive search in the distractor ratio paradigm, we sought an explanation where the observed effects are a consequence of ideal adaptation to noisy encoding processes in the fovea and parafovea. In the following section we describe a reduced complexity version of the distractor ratio paradigm for testing our model.

Horizontal Array Distractor-Ratio Paradigm

We reduced the task environment complexity from a two dimensional array (see Figure 1A, B, & C) to a one-dimensional array. This reduction in complexity facilitated the running of a large number of model trials while maintaining the critical components of the distractor ratio paradigm. The reduced complexity version used for testing the model was a set of seven objects arranged horizontally, with 8.3° of visual angle between adjacent items. The model searched through both target-present and target-absent trials for the same target throughout. Distractors were a conjunction of the same color as the target and a different shape, or vice versa. The model was tested over seven different distractor ratios (6:0, 5:1, 4:2, 3:3, 2:4, 1:5, 0:6; see Figure 2).

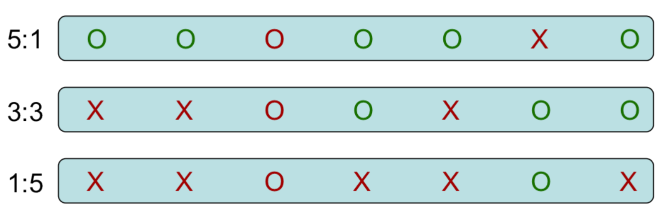

Figure 2: Three trials from the horizontal distractor-ratio paradigm. The target is a red O. Trial 5:1 corresponds with Figure 1A, 3:3 corresponds with Figure 1B, and 1:5 with Figure 1C.

This one-dimensional version of the distractor-ratio paradigm facilitated computationally tractable Monte-Carlo evaluation of the model, while retaining the relevant features of the paradigm. In the following section we describe the model in detail and present results from the model evaluated on the one-dimensional version of the task.

Adaptive Visibility Model

The goal of this modeling endeavor is to explain the phenomena associated with strategic adaptation observed in the distractor ratio paradigm, (i.e., the ∩-shaped curve and saccadic bias in favor of minority features) as adaptation to perceptual noise. To achieve this goal we differentiate between state estimation and control (Stengel, 1994). The model we present optimally estimates the state of the visual environment given noisy input, and controls responses based on the optimal state estimation.

The process of active, effortful visual search can be decomposed into two key control decisions: 1) determining if the target is in the stimulus (i.e., the stopping rule), and 2) determining where next in the stimulus to inspect (i.e., saccadic selectivity). All search models must contain a stopping rule and a saccade location selection process.

Toward this end we first identified physiological constraints on the visual search process. Next, we assumed that subjects in distractor ratio experiments intended to minimize the time to complete a trial. This assumption has been used in other models as a subjective utility function when an objective utility function is not provided to subjects (Gray et al., 2006; Myers et al., 2011). Third, we determined a set of strategies that could be performed in the task environment. Finally, we used Monte Carlo simulations of the model to determine if the bounded optimal model could explain the distractor-ratio hallmarks of adaptive search. Further, we investigate which model constraints were critical to adaptive visual search observed in humans when performing distractor ratio tasks. We cover each of these steps in more detail in the following sections, and provide a walkthrough of the model process before presenting the model results.

Constraints on Visual Search

The model begins a trial with a representation of whether the shape and color feature at each of the seven stimulus locations contains the same feature as the target. The model adopts a simple feature-vector coding of the display in which each of the seven locations is represented by 2 real-valued features (one for color, one for shape), where the value 1 is arbitrarily chosen as the target value for each dimension, and 0 as the non-target value. Thus, the true state of the display can be represented as a 14-element vector of 1s and 0s.

There are two constraints on the model, each of which limit the accuracy of the perceived information for each fixation. First, visual acuity decreases with increasing eccentricity from the fovea; we capture this constraint in the model with feature noise. Second, information located in the parafovea is subject to localization error (Levi, 2008; Neri & Levi, 2006), such that objects encoded in the parafovea may erroneously combine features from different objects at different locations (*illusory conjunctions*; Põder & Wagemans, 2007). Each percept obtained by the model is simply the true 14-element vector representing the display, corrupted by these two noise sources: feature noise and location noise.

The feature noise added to each true percept is a 14-element vector of values sampled from independent normal distributions with mean zero and standard deviations that increase as a step-function based on distance from the fovea. Standard deviations for determining feature noise within the fovea were set to 0.1 and 10 for outside the fovea (the qualitative results presented below do not depend on the precise shape of this acuity function). Localization noise was added to the model's percept by allowing the feature value for each position to be sampled from nearby positions with some probability. This probability was set to a low value for the fovea (representing an assumption of good feature binding in the focus of attention) and higher values for parafoveal positions (again, the qualitative results presented below do not depend on the precise values). The result from introducing these constraints was a model with a foveated visual system susceptible to illusory conjunctions. For each location, we sample all objects and obtain noisy feature information from these objects. For the fixation position this will very often be the true object but features will intrude for other positions in the periphery.

Optimal State Estimation & Control

The model uses Bayes' rule to optimally estimate the state of the display by integrating noisy perceptual information derived from each fixation. For each given noisy perceptual sample, the model computes the likelihood that the sample was generated from each of the possible target-absent and target-present displays and features at locations within those displays. This is accomplished as follows. First, the likelihoods of observing the perceptual sample at the feature level are computed (using the feature noise model). Second, the likelihoods that a sampled object at a particular location in the display has a specific feature value for each of the possible displays is computed (using the spatial noise model). Third, the probability that the percept was sampled for each display type is computed. Finally, the posterior probability over all the display types is computed using Bayes' rule.

Search Strategies

There are four potential strategies for locating the target in the distractor ratio paradigm. First, one could choose to make no eye movements at all during a trial, continuing to stare straight ahead. We rule out the use of this strategy as its utility in a search environment such as the distractor-ratio paradigm is very low. The remaining strategies were *random* search, *sequential* search and the *maximum a posterior* (MAP) searcher of Najemnik and Geisler (2008), which we label here the *look-for-targets* strategy, which simply looks at the location most likely to contain the target.

The random search strategy was implemented by uniformly sampling a location with replacement from all the possible locations in the reduced complexity paradigm until a response was made. Consequently, the model could choose to re-fixate a location it just acquired a percept from. The sequential search strategy was implemented by starting in the middle location and searching from right to left, and back around until a response was made.

The MAP strategy took advantage of the posterior probabilities after each fixation. The model chose the next fixation location based on the posterior likelihood of containing a target. In the next sections we provide a walkthrough of the model's process for completing a trial followed by results from each of the three strategies just described.

Model Walkthrough

The model begins each trial with all possible displays being equally likely; consequently, the initial values for the target-present and target-absent decision variables equal 0.5. Once a trial is presented to the model, it begins by fixating a location, obtaining a noisy percept from the fixated location, optimally integrating the noisy percept with previously acquired information from the trial, and calculating decision variables (i.e., target-absent and target-present) based on the optimally integrated percept. If neither of the decision variables reaches a decision threshold (arbitrarily set to 0.85 in the simulations, but which could be optimized to maximize utility in the face of imposed speed-accuracy tradeoffs), then the model selects a new location to fixate. If one of the decision variables is greater-than the threshold, then the model responds appropriately. A maximum number of fixations was set to 30 to prevent the model from infinitely fixating locations in the trial. To be clear, the model is not learning across trials, but is adapting to each trial, independently.

Model Results

The model was run for 20,000 trials for each of the random, sequential, and look-for-targets strategies. Each trial completed by the model was randomly selected with replacement from all possible trials. Surprisingly, all strategies produced the ∩-shaped curve for target-absent and target-present trials, indicating that the distract-ratio effect may arise from optimal state estimation in the face of noisy perception, independent of the saccadic control strategy. We investigate this finding in more detail below.

Less surprisingly, the random strategy required, on average, more fixations to respond ($M_{Target-Present}$ = 4.54; $M_{Target-Absent}$ = 5.13) than the sequential strategy ($M_{Target-Present}$ = 3.84; $M_{Target-Absent=4.53}$), which in turn took more fixations to respond than the look-for-targets strategy ($M_{Target-Present} = 2.94; M_{Target-Absent} = 3.77$).

The frequency of saccades directed toward objects containing a minority feature in a trial was evaluated to determine if it differed from what would be expected by chance (i.e., saccadic bias in right column of Figure 3; Shen et al., 2000). The analysis revealed that the look-for-targets strategy produced saccadic bias for target-present and target-absent trials whereas the random and sequential strategies did not. The results from the search efficiency and saccadic bias analyses demonstrate that the look-for-targets strategy produces both hallmarks of adaptive search within the distractor ratio paradigm.

To determine which perceptual constraint was required to yield the effects, we ran another round of simulations without location noise (one of two constraints in our ideal observer model). To make this determination we ran two sets of simulations: 20,000 trials for no-location-noise/high-feature-noise and 20,000 trials for no-location-noise/low feature-noise. The

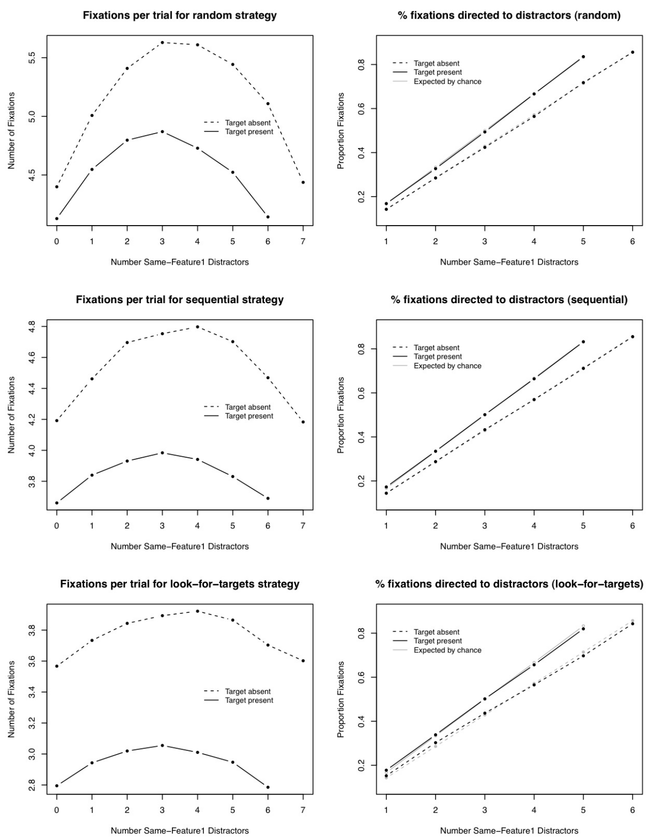

Figure 3: Hallmarks of adaptive search in the distractor ratio paradigm for the random (top), sequential (middle), and look-for-targets strategies (bottom). The left column demonstrate search efficiency in the paradigm and correspond to the human data in Figure 1D. The right column demonstrates saccadic bias in the look-for-targets strategy and the absence of the bias in the other strategies.

removal of location noise eliminated the presence of the ∩-shaped curve, whereas high feature noise only contributed to greater fixations to respond relative to low feature noise (see Figure 4). Consequently, we argue that the ∩-shaped curve observed in distractor ratio tasks results from the potential for illusory conjunctions in the parafovea.

Discussion & Conclusions

The preliminary analysis presented above contrasted a well-known salience based theory and a novel ideal observer based theory of distractor ratio phenomena. Despite the fact that the salience theory is widely accepted and that there is no previous ideal observer analysis of distractor ratio phenomena, we found that it offered a comprehensive explanation of the available empirical findings. Importantly, the different behaviors seen in people as a consequence of varying the statistical structure of the task environment emerge from a model that computes optimal state estimation and makes optimal control decisions given the constraints imposed by the human visual system.

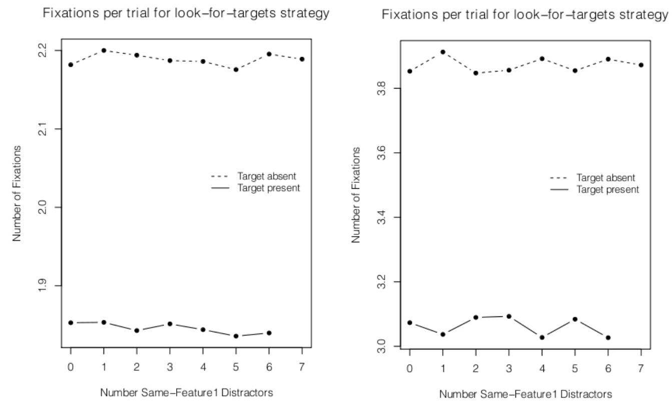

Figure 4: Search efficiency results without location noise when feature noise was low (left) and when feature noise was high (right) for the look-for-targets strategy.

These preliminary findings are promising because the ideal observer, by virtue of the combination of optimal state estimation and control, offers the potential of a deeper explanation than the mechanistic salience model. The ideal observer combines a theory of the information processing mechanisms with an analysis of optimal state estimation and control. Furthermore, the estimation and control decomposition permits the exploration of specific hypotheses concerning the locus of the explanation for a given search phenomenon. Here, we determined that distractor ratio effects are signatures of optimal state estimation in the face of spatial noise in the periphery, while the saccadic bias effects are signatures of both optimal estimation and control.

Although these findings are encouraging, the model requires a number of important revisions before we can be fully confident that it provides a rigorous demonstration of the implications of the hypothesized visual processing constraints for behavior. In particular, we did not explore the full strategy space for directing saccades; although the *look-for-targets* (MAP) strategy may be close to optimal in this task, we must derive the optimal strategy in the full space and confirm that its predictions are consistent with those of the simple look-for-targets strategy.

Furthermore, we must conduct new human experiments that systematically test predictions of the ideal observer that differ from those of the salience model. We envisage that the new data will be collected using a utility maximization paradigm similar to those used by Trommershäuser and colleagues (Stritzke, Trommershäuser, & Gegenfurtner, 2009; Trommershäuser, Maloney, & Landy, 2003) in the explorations of perceptual motor control, and Lewis, Shvartsman, and Singh (to appear) in the exploration of eye-movements in linguistic tasks. Bounded optimal control models naturally predict differences in performance that arise when the payoff is changed but the task and stimuli are otherwise identical, while salience-based models do not naturally predict such differences. A key advantage of these explicit-payoff paradigms is that assumptions regarding what subjects are maximizing during the experiment are grounded in the external payoffs, which are then used as the subjective utility functions in the optimal control models.

Acknowledgments

This research was supported by a grant awarded from the Air Force Office of Scientific Research to the three authors (AFOSR #12RH05COR) and grants awarded by NASA and the FAA to the second and third authors (NNX12AB08A and E2020272).

References

Bacon, W. J., & Egeth, H. E. (1997). Goal-directed guidance of attention: evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 948–961.

Baron, S., & Kleinman, D. L. (1969). The Human as an Optimal Controller and Information Processor. *IEEE Transactions on Man-Machine Systems*, *10*(1), 10–17.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*(3), 461–482.

Itti, L. (2006). Quantitative modelling of perceptual salience at human eye position. *Visual Cognition*, *14*(4-8), 959–984.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147.

Kowler, E. (2011). Eye movements: the past 25 years. *Vision Research*, *51*(13), 1457–1483.

Levi, D. M. (2008). Crowding An essential bottleneck for object recognition : A mini-review. *Vision Research*, *48*, 635–654.

Lewis, R. L., Shvartsman, M., & Singh, S. (to appear). The adaptive nature of eye-movements in linguistic tasks: How payoff and architecture shape speed-accuracy trade-offs. *Topics in Cognitive Science*.

Myers, C. W., Gray, W. D., & Sims, C. R. (2011). The insistence of vision : Why do people look at a salient stimulus when it signals target absence ? *Visual Cognition*, *19*(9), 1122–1157.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, *8*(3), 4.1–14.

Neri, P., & Levi, D. M. (2006). Spatial resolution for feature binding is impaired in peripheral and amblyopic vision. *Journal of Neurophysiology*, *96*(1), 142–153.

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, *7*(2), 1–12.

Pomplun, M. (2003, April). Area activation: a computational model of saccadic selectivity in visual search. *Cognitive Science*, *27*(2), 299–312.

Shen, J., Reingold, E. M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. *Perception*, *29*(2), 241–250.

Stengel, R. (1994). *Optimal Control and Estimation*. New York: Dover.

Stritzke, M., Trommershäuser, J., & Gegenfurtner, K. R. (2009). Effects of salience and reward information during saccadic decisions under risk. *Journal of the Optical Society of America A*, *26*(11), B1–B13.

Tanner, W. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, *89*, 752–765.

Theeuwes, J. (1993). Visual selective attention: a theoretical analysis. *Acta Psychologica*, *83*(2), 93–154.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003, July). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America. A, Optics, image science, and vision*, *20*(7), 1419–33.

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: Oxford University Press.

Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum.

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception*, *18*(2), 191–200.