Title
Optimality and flexibility in utilizing predictive spatial cues during visual search

Permalink
https://escholarship.org/uc/item/34v1r91p

Author
Mack, Stephen

Publication Date
2014

UNIVERSITY OF CALIFORNIA

Santa Barbara

Optimality and flexibility in utilizing predictive spatial cues during visual search

A dissertation submitted in partial satisfaction of the
Requirements for the degree Doctor of Philosophy
in Psychology

by

Stephen Christopher Mack

Committee in charge:
Professor Miguel P. Eckstein, Chair
Professor Barry Giesbrecht
Professor Mary Hegarty
Dr. Craig K. Abbey, Researcher

September 2014

The dissertation of Stephen Christopher Mack is approved.

Barry Giesbrecht

Mary Hegarty

Craig Abbey

Miguel Eckstein, Committee Chair

June 2014

Optimality and flexibility in utilizing predictive spatial cues during visual search

by

Stephen Christopher Mack

ACKNOWLEDGEMENTS

Graduate school has been an extraordinary challenge for me which has continually tested my limits as a researcher and a person. Despite the many proverbial valleys, as I look back at the past six years I am exceedingly proud: not just of what has been accomplished, but the relationships I've developed with those around me, the friendships that have stayed strong with thousands of miles in between, and the massive amount of personal and professional growth I've managed to stumble upon. Unfortunately for my ego, I did none of it alone. I am lucky beyond words to have had the friends, family, and colleagues I did around me to push me to keep improving, embrace myself and, above all, help me remain (moderately) sane.

None of this would have ever been possible without my unbelievably supportive parents. They found me in the valleys and showed me how to climb out and were there to meet me and celebrate at every peak. Dave and Kathi Mack are truly two of the most selfless, loving, and thoughtful beings on the planet. Without their guidance, I would never have become the person I am, nor would I be so fortunate as to have something so impressive to aspire to every day.

My sister, Jennie (or Birdman, as she shall be known), is the embodiment of everything that I want to be. Her intelligence and boundless generosity and empathy has brought me comfort more times than I can explain. While being 2000 miles away from her and my sister in law, Yael, is one of the least favorite parts of my life, eventually connecting with her after a week of phone tag is most certainly one of my favorites. I cannot explain how much my sister means to me.

To my friends, from Santa Barbara to the furthest reaches of the globe, I love you. Those in California (Matt Peterson, Ryan Kasper, Eamon Caddigan) and a bit further away (Collin Kreple, Steve Lee, Brendan Nytes). I owe all of you one. Actually, way more than one. I probably owe you roughly twenty.

And (next to) last, but nowhere near least, my beautiful partner Jacoby. You have been everything for me. You've put up with my long hours in the lab, you've rubbed my shoulders even when I'm convinced yours hurt more, and you've belayed me when I was grumpy and had extraordinarily and completely arbitrary preferences about how much slack I wanted (climber stuff). Every victory has been ours to share, and every tear has landed on your shoulder. I have literally no idea what I'd do without you, so I'd prefer we not explore that empirically. I love you.

And finally, I want to thank my advisor, Miguel Eckstein, and my committee of Barry Giesbrecht, Mary Hegarty, and Craig Abbey. The road hasn't always been smooth (in fact, I can't confirm it ever has been), but your knowledge and guidance has pushed me to be the researcher and scientist that I am now (sorry about that). Intellectually, I'm a vastly different person than I was six years ago, and I owe much of that to your dedication.

VITA OF STEPHEN CHRISTOPHER MACK

June 2014

Education

9/2008-4/2014 (expected) University of California-Santa Barbara
Ph.D. candidate in Psychological and Brain Sciences GPA: 4.0/4
9/2004-5/2008 University of Wisconsin-Madison
B.S. in Psychology, comprehensive honors and honors in psychology GPA: 3.98/4

Research Experience

9/2008-Present University of California-Santa Barbara
Graduate Student Researcher | Advisor: Dr. Miguel Eckstein
Vision and Image Understanding Laboratory

1/2007-5/2008 University of Wisconsin-Madison
Research Assistant | Advisor: Dr. Rick Cai
Vision Lab

8/2005-5/2007 University of Wisconsin-Madison
Research Assistant | Advisor: Dr. John Curtin
Addition Research Lab

Teaching Experience

Instructor of Record
Psychology 110A | Intro to Vision *6/2013-8/2013*

Teaching Assistant
Intro to Vision (x 2) | Perception (x 2) | Laboratory in Perception (x 2) | *1/2010-6/2014*
Laboratory in Attention | Statistics (x 2) | Intro to Psychology

Publications

Mack, S.C., & Eckstein, M.P. (*in prep*). Remote assistance: When moving the eyes away from the target aids search performance.

Mack, S.C., & Eckstein, M.P. (*in prep*). Predictive spatial cues can hinder multi-fixation search performance.

Eckstein, M.P., Mack, S.C., Liston, D.B., Bogush, L., Menzel, R., & Krauzlis, R.J. (2013). Rethinking human visual attention: Spatial cueing effects and optimality of decisions by honey bees, monkeys and humans. *Vision Research, 85(7), 5-19.*

Mack, S.C., & Eckstein, M.P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision, 11(9), 1-16.*

Talks and Posters

Mack, S.C., Akbas, E., & Eckstein, M.P. (2013). Optimal search by initially looking away from the target in the presence of remote cues. *Poster presented at Vision Sciences Society Annual Meeting.*

Mack, S.C., & Eckstein, M.P. (2012). Detrimental effect of spatial cues predicted by a foveated Maximum a Posteriori eye movement model. *Poster presented at Vision Sciences Society Annual Meeting.*

Mack, S.C., Zhang, S., & Eckstein, M.P. (2011). Cue capture: when predictive cues hinder search performance. *Talk presented at Vision Sciences Society Annual Meeting.*

Zhang, S., Mack, S.C., & Eckstein, M.P. (2011). The odd human eye movements during oddity search are not suboptimal. *Poster presented at Vision Sciences Society Annual Meeting.*

Mack, S.C., Liston, D.B., Krauzlis, R.J., Bogusch, L., Menzel, R., & Eckstein, M.P. (2010). Spatial cueing effects in perceptual decisions of humans, monkeys, and bees. *Poster presented at Vision Sciences Society Annual Meeting.*

Mack, S.C., Schoonveld, W., & Eckstein, M.P. (2009). Contextual cues facilitate search in real world 3-D environments. *Poster presented at Vision Sciences Society Annual Meeting.*

ABSTRACT

Optimality and flexibility in utilizing predictive spatial cues during visual search

by

Stephen Christopher Mack

Visual search is a critical and pervasive part of our everyday lives. However, the ease of which we perform search can mask its remarkable computational complexity. Search targets are often difficult to detect and embedded in statistically complex backgrounds. To optimize search, human observers often exploit known statistical properties of the visual environment which provide information about target location. Regularities in the spatial organization of the visual environment (e.g. predictive cues) have shown to be one such type of statistical property which can be leveraged to increase search efficiency. Here, a series of three studies examines how well human observers can exploit spatially predictive cues during multi-fixation search. Further exploration focuses on highlighting human flexibility in altering search strategy to enhance perceptual performance, as well as delineating situations in which predictive information in the environment may actually hurt search performance.

Predictive spatial cues have been shown to improve perceptual performance for a variety of tasks, including visual search, under conditions of forced fixation. However, the potential benefits of predictive cues during multi-fixation search are poorly understood. In the first study, we present a letter identification search task, done in the presence and absence of an array of spatial cues which framed potential target locations. We show that human observers direct their eye movements towards cued locations to improve their search performance compared to when the cues are absent. We also develop a foveated eye movement model, which takes into account the diminishing acuity of the human visual system in the periphery, for the task. Model predictions reveal substantial performance benefits via predictive cues, the size of which are much larger than what is seen in human data.

In the second study, we investigate whether human observers will utilize peripheral predictive cues which reside in display regions which have no chance of containing the target (which we call *remote cues*). In doing so, observers must depart from a commonly used "saccadic targeting" strategy, where eyes are directed to likely target locations. When informed of the predictive nature of the remote cues, observers readily adopt an atypical eye movement strategy which favors non-target locations to enhance task performance. A foveated model which ignores the cues reveals that these performance benefits are a likely result of foveating peripheral predictive information. Interestingly, a version of the remote cue task in which observers were not informed of cue contingencies, reveals that while observers can adopt atypical saccadic strategies to improve performance, they do not readily engage in such behavior without explicit information.

In the final study, we explore a situation in which predictive cues actually hinder search performance. In natural viewing environments and everyday search tasks, predictive cues often do not mark the only target locations, but merely likely ones. We return to the letter identification search task and modify the cues so that they are only partially predictive of target location. With this increased uncertainty, the presence of predictive cues actually leads to worse performance when the target is highly visible, compared to when cues are absent. A control task in which the spatial cues were not predictive rules out the possibility of the effect being driven by exogenous attentional capture, and a foveated eye movement model reveals that cues hindering search performance would not be predicted by a rational observer. Implications for real-life and vocational search are discussed.

In all, we see that human observers are willing and able to utilize predictive elements via intelligent eye movement selection of the environment to enhance search performance. In addition, when cue information is made clear, observers are able to readily adopt even extremely atypical eye movement strategies to optimize performance. However, human efficiency in implementing these strategies often falls short of that predicted by a near-optimal observer, even leading to a decrement in performance in extreme cases. Understanding the interaction between human performance and predictive cues, then, is critical to assessing natural visual search and optimizing vocational and life-critical search displays.

TABLE OF CONTENTS

LIST OF FIGURES

I. Introduction

Visual search is a pervasive and remarkably computationally complex task. From locating car keys on a cluttered coffee table to finding a friend's face in a crowd, we're constantly surveying the visual environment for objects of interest. In natural visual environments, search targets are often difficult to detect, variable in appearance, and embedded in rich and statistically complex backgrounds (Eckstein, 2011; Geisler, 2008; Torralba & Oliva, 2003). Although our experience of search may be effortless and automatic, our continued success with it is an impressive accomplishment.

Part of the reason human observers are able to perform search with such proficiency is that search is not random. There are large amounts of structural and statistical regularities in visual environments which can be learned and exploited by human observers to enhance search accuracy and efficiency. These regularities can be in the form of knowledge about what the target does or should look like, the spatial structure of the environment, or the relationship between the two.

Much previous research has been done regarding the first of these sources of information: target appearance. It is intuitive to think that, when searching for a target of a known (or inferred) appearance, human observers may move their eyes towards things that look like the target. Behavioral research has revealed that this is in fact the case, as human observers preferentially deploy saccades to image regions which share visual properties with the target (Beutter, Eckstein, & Stone, 2003; Findlay, Brown, & Gilchrist, 2001; Findlay, 1997; Malcolm & Henderson, 2010; R. P. N. Rao, Zelinsky, Hayhoe, & Ballard, 2002), an oculomotor strategy deemed "saccadic targeting.' These findings have since led to a number of computational models of search which utilize the known or expected appearance of the target to guide model saccades and enhance search efficiency and accuracy (Eckstein, Drescher, & Shimozaki, 2006; R. P. N. Rao et al., 2002; Wolfe, 1994; Zelinsky, 2008).

While directing the eyes towards image regions which resemble the target is certainly a valuable strategy, it is certainly not a complete description of the strategies used to guide eye movements during search. Take, for example, looking for a bowl in an unknown kitchen. Before entering the kitchen, you know there are places where the bowl is likely to appear (on the counter, in a cupboard) and places that are not so likely (on the ceiling, in the coffee machine). These spatial relationships, learned through visual experience, cannot be captured by any amount of target features, but are clearly a can play a large part in our search strategies.

Regularities in the spatial structure of the environment can enhance search performance by providing information for the strategic planning of saccadic eye movements to likely target locations. This has been most widely examined in terms of scene context, or the spatial relationship between the search target and the organization of the surrounding scene. Observers consistently tend to bias saccades towards either scene regions (Brockmole & Henderson, 2006; Castelhano & Heaven, 2011; Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Henderson, 2003; Hollingworth, 2009, 2012; Neider & Zelinsky, 2006; Spotorno, Malcolm, & Tatler, 2014; Torralba, Oliva, Castelhano, & Henderson, 2006) or individual objects which tend to spatially co-occur with the search target (Castelhano & Heaven, 2011; Eckstein et al., 2006; Mack & Eckstein, 2011). Moreover, there is growing evidence that spatial scene context information is explicitly represented in the brain and accessed during search (Giesbrecht, Sy, & Guerin, 2013;

Preston, Guo, Das, Giesbrecht, & Eckstein, 2013). These results suggest that observers can effectively internalize and exploit spatial relationships in the environment to guide search.

But how *well* are observers integrating these statistics into their search strategies? Unfortunately, the statistics of natural scenes (Chun, 2000; Eckstein, 2011; Geisler, 2008) and object co-occurrence (Bar, 2004; Mack & Eckstein, 2011) are poorly understood. Without knowing the underlying statistics, it becomes difficult to objectively assess observer strategies. The use of synthetic spatial cues and images, however, provides a way of knowing (and controlling) the statistics of visual displays, allowing the formulation of computationally tractable models of visual search to objectively assess the efficiency of how human observers integrate information about spatially predictive elements of the scene into search behaviors and performance.

The use of synthetic spatial cues is common in visual perception research. The presence of spatially predictive cues nearly invariably improves performance on a variety of tasks such as simple detection (Bashinski & Bacharach, 1980; Eckstein, Shimozaki, & Abbey, 2002; Luck et al., 1994; Posner, Snyder, & Davidson, 1980; Posner, 1980), discrimination (Henderson, 1991, 1996; Lee, Koch, & Braun, 1997), measurement of contrast sensitivity (Cameron, Tai, & Carrasco, 2002; Pestilli & Carrasco, 2005), letter identification (Talgar, Pelli, & Carrasco, 2004), and, critically, visual search (Droll, Abbey, & Eckstein, 2009; Eriksen & Yeh, 1985; Nakayama & Mackeben, 1989; Vincent, 2011; for review see Carrasco, 2011). However, the effect of predictive cues on search performance has primarily focused on the deployment of covert visual attention, utilizing simple search displays with targets of known constant visibility, a relatively small number of target locations, and paradigms which do not include eye movements. Given the ubiquity of eye movements in everyday search, an assessment of how spatially predictive cues impacts more naturalistic multi-fixation search seems sorely needed.

Predictive spatial relationships provide valuable information which can be utilized to enhance search performance. Scene context literature provides evidence that human observers can integrate information about predictive elements of the visual environment into their search strategies to enhance performance. However, without knowing the underlying statistics of scenes, objective assessment of these strategies is not possible. The use of experimenter controlled synthetic cues and images provides an environment in which to objectively assess search strategies, but previous literature has focused on single-fixation search, which fails to represent the vast majority of naturalistic search scenarios. Here, three studies are described which attempt to begin to fills these gaps in the literature and provide an objective assessment of human ability, inability, and flexibility in effectively utilize spatial cues during multi-fixation search.

Chapter II addresses the question of whether human observers can effectively integrate predictive cue information into their eye movement and perceptual decisions to increase search performance. We develop a free eye movement letter identification search task in which observers are to identify which one of five letters is present in a noisy image. The signal strength of the target varied and was unknown to the observer, simulating the variable detectability of targets in natural search. Critically, the task was performed in the presence and absence of an array of predictive cues which were fully predictive of target location when present. In this paradigm, variations in perceptual performance and eye movement selection can provide valuable insight into the nature of and degree to which predictive spatial cues influence search. To supplement this behavioral work, a Bayesian foveated model observer is developed to serve as a benchmark against which to compare human performance and provide objective assessment of expected performance gains. The model simulates the diminishing acuity of the human retina with retinal eccentricity as it implements a near-optimal maximum *a posteriori* (MAP) decision rule, in which the eyes are directed to the portion of the visual display containing maximum evidence for the target. In tandem behavioral data, the implementation of the foveated MAP searcher can shed light onto the rationality and quality of predictive cue integration into human observer eye movement strategies and perceptual decisions.

In Chapter III, the investigation turns to observer flexibility in utilizing predictive cues. In Chapter II, as well as the vast majority previous cueing literature, the predictive cues spatially co-occur with the target (or potential target locations). Thus, search guidance via the cues can often be difficult to dissociate from that driven by the target. In Chapter III, we outline a novel search task in which observers were to indicate the presence of a target embedded in noise at one of ten potential target locations. The orientation of three peripheral spatial cues, which we call remote cues, indicated where the target would appear if it were present, and observers were explicitly informed of their predictive nature. Critically, the region of the image which contained the peripheral cues had no chance of containing the target, so investigating them would necessarily draw the eyes away from potential target locations. Variants of the task manipulate the visibility and availability of remote cue information to determine the relative impacts of remote cue information and eye movement strategy on perceptual performance. A foveated Bayesian eye movement model, which can only move its eyes to potential target locations, is developed to provide a baseline of expected performance in the absence of a remote-cue-driven eye movement strategy. A version of the task in which observers are given no information about the predictive nature of the remote cues explores whether observers are willing to adopt uncommon eye movement strategies when benefits are not readily apparent. In all, these studies highlight human flexibility in utilizing predictive information in the visual environment.

The work described above focuses heavily on describing the benefits that may arise from the effective integration of predictive cue information into search strategies. However, is it necessarily so that predictive cues always lead to enhanced performance? In Chapter IV, we return to the letter identification search task and describe a situation in which performance in the presence of predictive cues actually suffers compared to when cues are absent. Eye movement patterns reveal possible causes for this unexpected reversal in performance. A control study, in which cues are present, but not at all predictive of target location, seeks to rule out the alternative explanation of our data via exogenous attentional capture, and the foveated MAP developed in Chapter II is implemented to determine whether such an anomalous result would be predicted by a rational observer. Additionally, the implications of understanding human-cue interactions in vocational and life-critical search tasks are discussed.

Visual search is a deeply intricate computational undertaking. Fortunately, there is a wealth of spatial structure in the visual environment which can be extracted to aid observers in this task. The research outlined here provides bridges the gap between scene context and spatial cueing work to provide an objective assessment of human cue use during multi-fixation search. While observers are willing to integrate such information into their oculomotor strategies and demonstrate surprising flexibility in their search behaviors to optimize performance in the face of changing task demands, it is also clear that the sheer presence of predictive information is not necessarily sufficient to enhance task performance. With the increasing popularity of spatial cues meant to aid performance in vocational search tasks (Baker et al., 2003; Philpotts, 2009; V. M. Rao et al., 2010; Willhauck, Schneider, De Kok, & Ammer, 2000), having an intimate understanding of human-cue interaction (including when cues may actually hurt performance) is critical to designing effective search displays and aids.

II. Predictive cues aid multi-fixation visual search

Spatially predictive cues have nearly invariably been shown to improve perpetual performance on a variety of tasks under conditions of forced fixation, including visual search. However, the effects of predictive cues on eye movement selection and multi-fixation search performance are poorly understood. Although evidence exists from the scene context literature that human observers direct their eyes to predictive elements of the environment during search, it is unclear how *well* they do this or the benefits of predictive cues on perceptual decision accuracy.

Here, we develop a letter identification search task to assess whether human observers can effectively leverage information from an array of predictive spatial cues to enhance performance compared to when cues are absent. Search performance is markedly improved when predictive cues are present, particularly at low contrasts. This performance enhancement is reliably related to a pattern of eye movements which systematically targets cued locations, indicating the integration of cue information into eye movement strategies.

To compare the gains in perceptual performance afforded by the cues with an objective benchmark of performance, a foveated maximum *a posteriori* (MAP) eye movement model is developed. The model also shows a cue-driven pattern of eye movements when cues are present which, as seen in human data, leads to enhanced performance. However, the gains in perceptual performance predicted by the model far exceed those of human observers and human eye movement patterns depart significantly from model predictions, indicating suboptimalities in human search strategies and/or perceptual performance.

A. Introduction

Uncertainty about the spatial location of a target often hinders its detection and identification (Burgess & Ghandeharian, 1984; Swensson & Judy, 1981). Yet, cues predictive of the spatial location of a target nearly invariably improve performance on a variety of tasks such as simple detection (Bashinski & Bacharach, 1980; Eckstein et al., 2002; Luck et al., 1994; Posner, 1980), discrimination (Henderson, 1991, 1996; Lee et al., 1997), measurement of contrast sensitivity (Cameron et al., 2002; Pestilli & Carrasco, 2005), letter identification (Talgar et al., 2004), and, critically, visual search (Droll et al., 2009; Eriksen & Yeh, 1985; Nakayama & Mackeben, 1989; Vincent, 2011); for review see Carrasco, 2011). However, the effect of predictive cues on search performance has primarily focused on the deployment of covert visual attention, utilizing simple search displays with targets of known constant visibility, a relatively small number of target locations, and paradigms which involve short presentation times and no eye movements.

Typically, real world search involves both covert deployments of visual attention as well as active eye movements which direct the high resolution fovea to potential regions of interest in the scene (Henderson, 2003; for review see Eckstein, 2011), yet few studies have investigated the role of predictive spatial cues in multiple fixation search.

Additionally, although most studies investigating predictive spatial cues utilize targets of known and constant discriminability or detectability (e.g. luminance or orientation difference with respect distractors and/or the background; (Eckstein et al., 2002; Eriksen & Yeh, 1985; Posner, 1980), in most real world scenarios the visibility of search targets varies across instances and is not known to the observer. Furthermore, in many applied settings of human visual search with various imaging modalities (e.g. medical images: (Gur & Sumkin, 2006; V. M. Rao et al., 2010); satellite and aerial images: Willhauck et al., 2000; air traffic control: Metzger & Parasuraman, 2006; Wickens, Conejo, & Gempler, 1999), search images are presented with highly visible cues indicating the locations an image analyst or computer vision algorithm considers likely to contain a target.

Here, we ask whether spatial cues aid perceptual performance in more complex search scenarios in which there are variations in target discriminability and observers are allowed free eye movements. We implemented a letter identification search task in which observers were allowed free eye movements to locate and identify letters of varying detectability in a noisy display. We compared letter identification performance in the cues present condition, where an array of four color coded predictive cues indicated target locations, to that of a cues absent condition in which the stimuli remained identical except for the absence of the predictive spatial cues. We show that although the spatial cues improved identification performance overall, and markedly so at lower signal contrasts. Analysis of observers' fixation selections suggests that underlying this enhancement in performance was related to the preferential foveation of the predictive cues, indicating the effective integration of cue information into eye movement selection in multi-fixation search.

We also develop a foveated maximum *a posteriori* (MAP) eye movement model, which simulates the decrease in acuity across the retina, to provide a near-optimal benchmark of performance against which to compare human data. The foveated MAP, which directs eye movements towards the location in the image most likely to contain the target, also shows marked improvement in performance when predictive cues are present. As with human observers, this enhancement in search performance is accompanied by an increase in cue-driven eye movements. The predicted benefits afforded by spatial cues, however, are much larger than what is seen in human data, indicating suboptimalities in human eye movement strategies and perceptual decisions.

B. Observer study

1. Method

Four naïve observers (3 female; ages 19-22) completed a letter identification search task. The observers' task was to determine which one of five letters (A-E, 1.1 x 1.1º), at one of five randomly selected contrasts (10-18%) embedded in a Gaussian luminance noise field (25 ± 4.9 cd/m², 22.2 x 22.2º), was present in the display. The signal to noise ratios (SNRs) of the letters ranged from 6.5 to 14.

A letter was always present in the stimulus image. In cues present sessions, four cue circles (diameter 2.36º) were overlaid on the images whose color (red, green, blue, or yellow) indicated the probability of the letter appearing in that region. The four cued regions always contained the target when present, with individual probabilities of 50, 30, 10, and 10% and a randomized color-probability pairing for each observer. On each trial, the locations of the high probability regions were rotated about the center of the image by a random angle so that location itself was not a predictor of target location. In cues absent sessions, the structure of the images and statistics governing target location were identical (i.e. there were still four locations that could contain the target), but no cue circles were present. Prior to starting the task, all participants were explicitly informed of the probabilistic structure of target location and the associated cue probabilities. Observers completed ten 100 trial sessions of the cued condition and eight 100 trial sessions of the uncued condition.

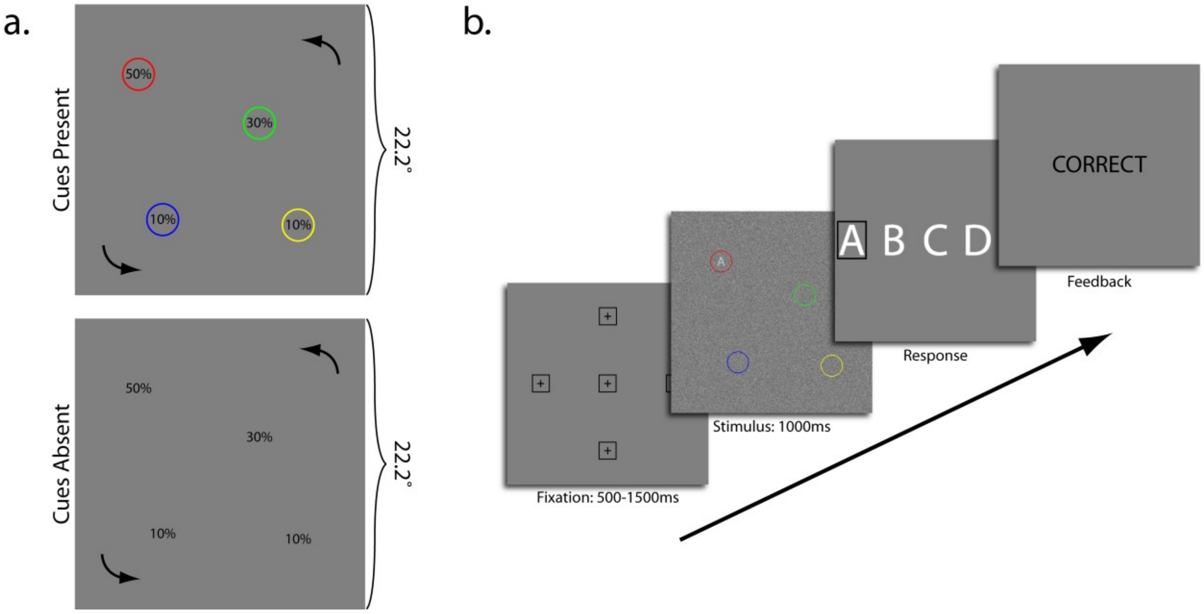

Figure 1. Trial structure of the letter identification task. Panel (a) shows the probabilistic distribution of search target location. Note that the relative spatial relationship between the high probability regions was maintained across trials, although their absolute spatial location rotated about the center of the image by a random angle for each trial. (b) shows the temporal structure of a trial.

Trials were initiated by the observer fixating on one of five fixation points distributed in a cross and pressing the space bar. Initial fixation locations were rotated at the same angle as their paired stimulus image so that there were only five starting fixation locations with respect to the target locations. If fixation was broken during the interval in which the fixation point was present (randomized 500-1500ms), the trial was aborted and restarted from the beginning. Immediately following the fixation period, the stimulus image appeared on the screen for 1000ms, allowing between 2-4 saccades. While the stimulus was present, observers were free to move their eyes to identify which letter was presented in the image. After stimulus presentation, a response screen appeared on which the observer indicated via mouse click which letter they believed to be present in the image. If observers were uncertain, they were instructed to make their best guess. Following their response, observers were provided with binary feedback (correct/incorrect). If they were incorrect, they were not told which letter was present on that trial (see Figure 1b for trial structure).

Gaze was monitored using an Eyelink 1000 infrared eye tracker (SR Research) providing a monocular left eye track at a sampling rate of 250 Hz. Changes in eye position that exceeded a velocity of 35°/s and an acceleration of 9500°/s/s were considered to be saccades. The first eye movement which deviated more than 2° from initial fixation was considered the first saccade for purposes of analysis. A chin rest was used to stabilize head position.

2. Results

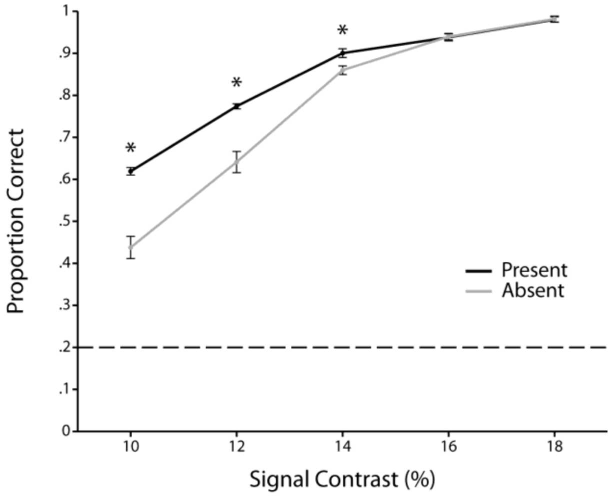

Figure 2. Identification performance as a function of signal contrast. Perceptual performance was enhanced in the presence of cues overall. The size of this benefit was largest at low signal contrasts and decreased with increasing signal strength. Error bars represent ±SEM.

Perceptual performance was calculated as a function of signal contrast for both cueing conditions. Overall performance in the cues present condition exceeded that of the cues absent across all signal contrasts (*F(1,3)*=50.98, *p*=.006; see Figure 2), indicating overall facilitation of identification performance in the presence of predictive cues. A significant interaction between cueing condition and contrast also emerged (*F(4,12)*=37.56, *p*<.001). Pairwise comparisons revealed that performance was superior for cued sessions at the three lowest signal contrasts (paired t, *p*<.05), with the magnitude of the difference decreasing with increasing contrast to the point where identification performance converged at the two highest signal contrasts (paired t, p>.4). These results are consistent with previous research which shows that as the target becomes more visible, the spatial guidance the cues provide becomes less valuable for locating it, decreasing the performance benefit afforded by cues (e.g. Eckstein et al., 2013).

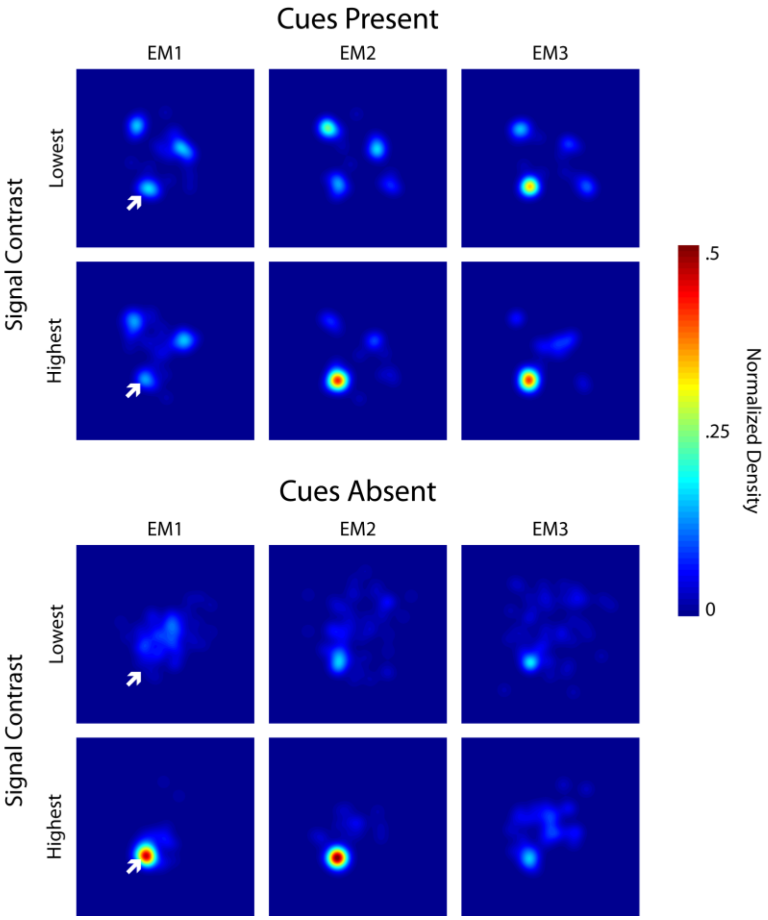

Figure 3. Normalized eye movement distributions for the letter identification task. Eye movement distributions are shown for a single target location (indicated by the white arrow) for the highest and lowest contrast across three eye movements for cues present and cues absent conditions. When cues are present, eye movements cluster around cued locations. When cues are absent, eye movements are dispersed throughout the image until it convenes on the target location.

Analysis of eye movement selection clearly revealed that this enhancement in performance in the presence of predictive cues was related to a cue-driven saccadic strategy when cues were present. Figure 3 shows normalized distributions of the first three eye movements for a single target location at the lowest and highest signal contrast for both cues present and absent. When cues were present, eye movements grouped not only around the target location (indicated by the white arrow), but also around the other cued locations, indicating a cue-driven eye movement strategy. When cues were absent, eye movements were dispersed rather evenly throughout the stimulus if not at the target location, reflecting the lack of predictive information in the image.

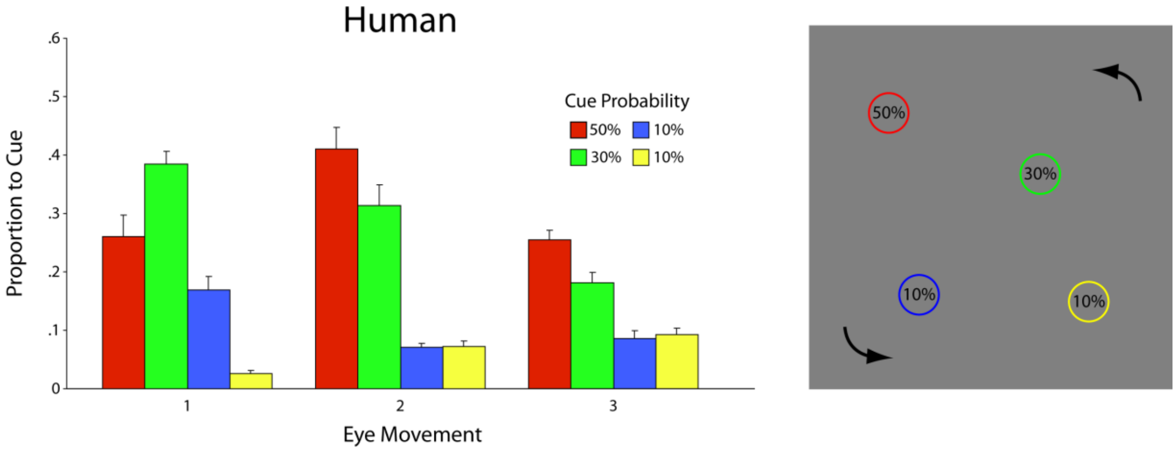

Figure 4. Proportion of eye movements to each cue. Interestingly, the largest proportion of first eye movements is directed to the 2nd highest probability cue, potentially a result of it being the closest cued location from fixation. In general, however, higher probability cues receive more fixations than lower probability cues. The image on the right shows the correspondence between cue location and probability.

Although human eye movements are clearly driven towards the cues, it is an open question how observers prioritized the cue information in saccadic selection. Figure 4 shows the proportion of saccades towards each cue in cues present trials (labeled by its probability of containing the target) as a function of eye movement. Overall, the proportion of eye movements towards the cues decreased as trials progressed ($F(2,6)=25.06$, $p=.001$). Additionally, the cue probability substantially impacted eye movement selection ($F(3,9)=53.13$, $p<.001$), with higher probability cues (50% and 30%) gaining significantly more fixations than low probability cues (paired *t*, all $p<.01$). Critically, a significant interaction between eye movement and cue ($F(6,18)=13.15$, $p<.001$) revealed that the preference for observers to target certain cues changed over the course of a trial. Of particular interest is the propensity for observers to direct the largest proportion of first saccades to the 2nd highest probability cue (even though it was 20% less likely to contain the target than the most predictive cue) but prefer the highest probability cue in subsequent eye movements. One potential explanation for this finding lies in the physical setup of the task. Human observers have been shown to prefer shorter saccades, regardless of potential costs in doing so in terms of information gathering (e.g. Araujo, Kowler, & Pavel, 2001). In our task, the 2nd highest probability cue location also happened to be the shortest mean distance from all fixations (*M*=7.23º). Thus, the preference for the 2nd highest probability location in the first eye movement may simply represent a strategy to initially make a short eye movement.

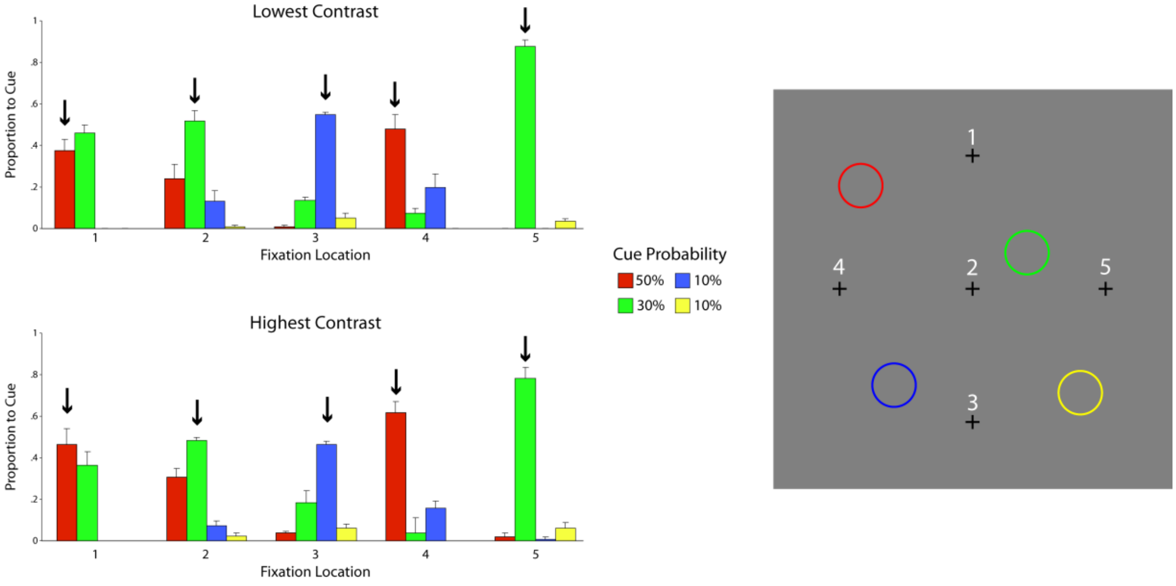

Figure 5. Assessment of distance-based strategy in first saccade selection. Regardless of contrast, observers tend to direct their first eye movement to the cued location which is closest to the starting fixation point (indicated by a black arrow for each fixation).

To assess this claim, Figure 5 shows the distribution of *first* saccades to each cue, broken down by fixation location for the lowest and highest contrast. A distance based strategy is clearly present, as the cue to garner the most first saccades is nearly always the one nearest the point of fixation (indicated by the arrow), regardless of signal contrast. It is, however, worth nothing that the proportion of first eye movements towards the highest probability cue increased significantly at the highest contrast ($t(3)=10.82$, $p<.001$), suggesting that stronger visual evidence may have provided observers enough information to override a distance-based eye movement selection strategy.

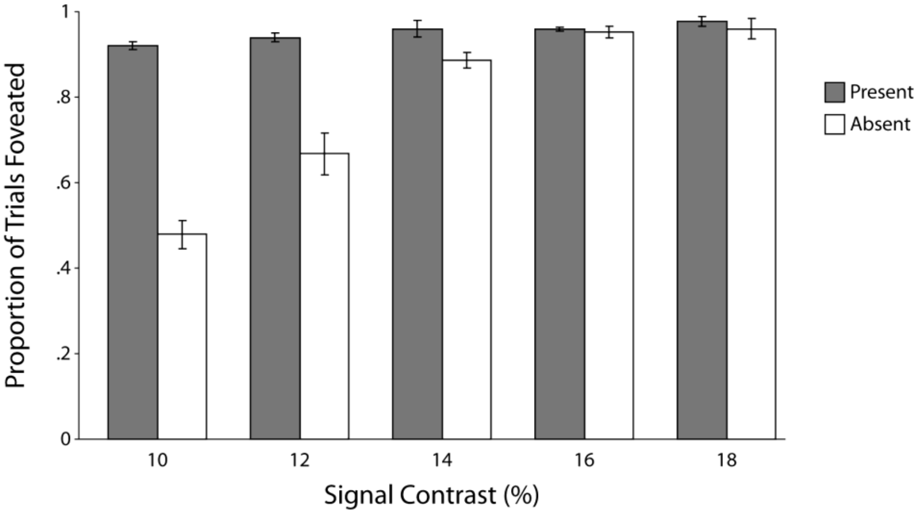

Figure 6. Proportion of trials that the target was foveated. In the presence of predictive cues, observers were far more likely to foveate the target, particularly at low signal contrasts. The disparity is largest at low signal contrasts and decreases rapidly with increasing signal strength, mirroring trends in identification performance. Error bars represent ±SEM.

Observers clearly used the cues to guide search, but what were the functional consequences of this cue-driven eye movement strategy when cues are present? A two (cueing condition) by five (contrast) repeated measures ANOVA revealed that, as a result of this cue-driven eye movement strategy, the target was foveated (i.e. eye position fell within 2° of target) on substantially more trials when cues were present than when they were absent (*F(1,30)*=127.01, *p*=2.63e-12; see Figure 6). A significant cueing by contrast interaction (*F(4,30)*=34.6, *p*=7.70e-11) indicated that while this discrepancy in ability to foveate the target was quite large (~10-40%) and highly significant (paired t, all *p*<.01) at the lowest three signal contrasts, cue presence did not significantly impact observers' propensity to fixate the target at the two highest signal contrasts (paired t, *p*>.2), mirroring trends in perceptual performance. These converging lines of evidence suggest that the cue driven eye movement pattern displayed when cues were present enhanced observer probability of foveating the target, leading to augmented identification performance.

3. Discussion

In a novel letter identification search task, human observers were able to utilize an array of predictive spatial cues to enhance multi-fixation search performance. Moreover, this improvement in performance was associated with a cue-driven pattern of eye movement selection when cues were present, enhancing observers' ability to foveate and identify the target. There was, however, evidence that strategies unrelated to the cue information (such as a minimum distance first saccade) may have also been implemented into human eye movement select.

Although cue presence led to enhance search performance via cue-driven saccadic selection, it is unclear how *well* observers utilized this information. Perhaps larger gains in search performance could be attained by more efficient use of cue information in both perceptual decision making and eye movement selection. To objectively assess the quality of human observer perceptual performance and saccadic strategy, a foveated eye movement model, which takes into account inhomogeneities in visual acuity across the retina, was developed to provide a benchmark of rational and near-optimal behavior.

C. Foveated maximum a posteriori (MAP) eye movement model

Here, we develop a foveated Bayesian maximum *a posteriori* (MAP) eye movement model for the task, which directs eye movements towards locations most likely to contain the target, to provide a quantitative measure of near-optimal eye movement selection and task performance. A foveated model accounts for the well-documented fact that human visual acuity is highest at the fovea and rapidly falls off in the periphery, leading to severe declines in perceptual performance for most visual tasks at peripheral locations (Najemnik & Geisler, 2005; Peterson & Eckstein, 2012, 2013; Zhang & Eckstein, 2010). Therefore, any complete model of active search performance should explicitly account for this inhomogeneity in acuity. Comparison of foveated model predictions with human eye movement patterns across the various tasks can provide a formal basis to identify and characterize suboptimalities in the usage of cue information, and provide a set of results, in terms of both eye movements and perceptual performance, which would be expected from a rational observer.

1. Model formulation

A foveated Bayesian maximum *a posteriori* (MAP) model was applied to provide a benchmark against which to compare human eye movement selection and search performance. A MAP model considers the likelihood ratio that the target is at each location in the image by calculating the likelihood the visual evidence at that location reflects both signal and noise distributions, and weights that ratio by the prior probability that the target will appear there. The next eye movement is directed towards the location which is most likely to contain the target (highest weighted likelihood ratio). When there are inhomogenieties in processing across the visual field (as in a foveated system), MAP models have been shown to produce eye movement patterns that diverge from an optimal search model that plans eye movements to maximize decision accuracy for a localization task (Najemnik & Geisler, 2005, 2008). However, in many cases, MAP models are a reasonable approximation of optimal eye movements (Najemnik & Geisler, 2008; Zhang & Eckstein, 2010), lead to similar decision accuracy (Zhang & Eckstein, 2010), and are good predictor of human eye movement distributions (Beutter et al., 2003; Findlay, 1997; Verghese, 2012).

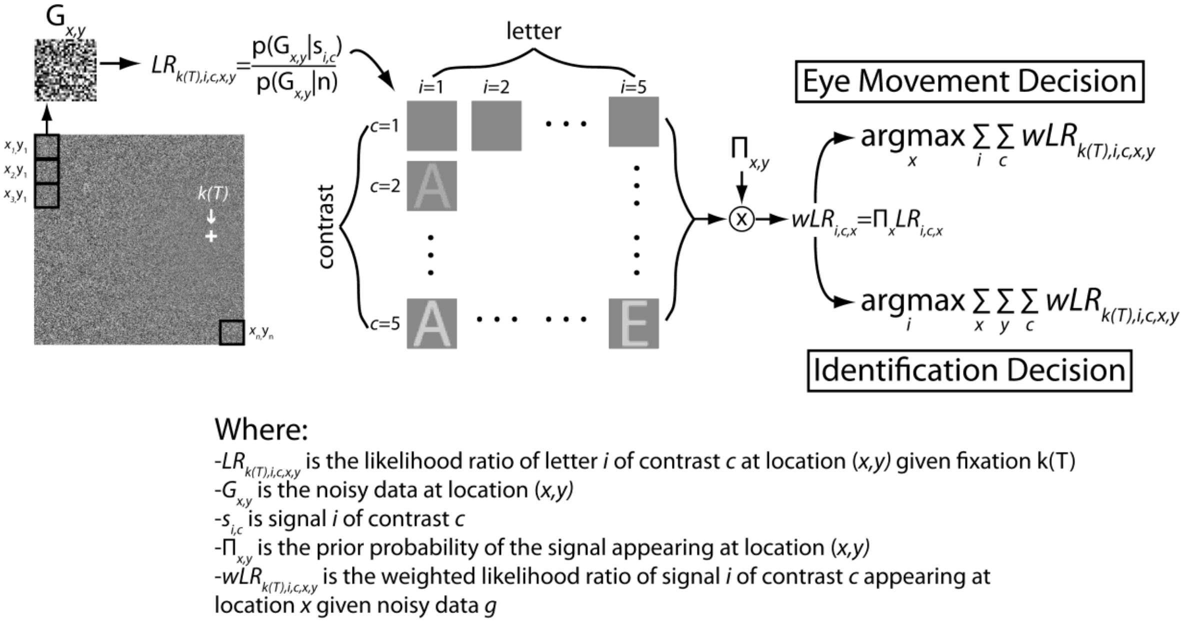

Figure 7. Flowchart of the foveated MAP model. A likelihood ratio is calculated for each possible signal (5 letters x 5 contrasts), which is weighted the prior probability of the signal appearing at that location. Note that the likelihood ratio calculations which underlie the eye movement and ID decision are identical. The decisions simply differ in what dimensions likelihood ratios are summed across.

The model receives a noisy image (*G*, the data) in which one of the five letters is embedded at a randomly selected one of five contrasts. The noisy image remained the same throughout each simulated trial, reflecting the static image noise utilized in the observer study. A likelihood ratio for all signals (25 in all) is calculated for all possible signal locations. The likelihood ratio that a single letter at a single contrast appears at one location is given by:

$$LR_{k(T)i,c,x,y} = \frac{p(G_{x,y} \mid s_{i,c})}{p(G_{x,y} \mid n)} \tag{1}$$

$LR_{i,c,x,y}$ is the likelihood ratio of signal *i* of contrast *c* at location *x* given current fixation position *k(T)*, $s_{i,c}$ is signal template *i* of contrast *c*, $G_{x,y}$ is the noisy data at location (*x,y*). Since the background noise is Gaussian and adjusted to be zero mean and unit standard deviation, and the strength of the additive signals are known to the model, the pdf of the signal (***s_{i,c}***) and noise (***n***) can be given by:

$$n \sim N(0,1) \tag{2a}$$

$$s_{i,c} \sim N(\mu_{i,c},1) \tag{2b}$$

Where $\mu_{i,c}$ is the strength of signal *i* of contrast *c*. Thus, the likelihood ratio is given by the expression:

$$LR_{k(T),i,c,x,y} = \frac{\frac{1}{\sqrt{2\pi\sigma^2{}_{k(T),x,y}}}\exp\left(-\frac{(r_{k(T),ic,x,y}-\mu_{i,c})^2}{2\sigma^2{}_{k(T),x,y}}\right)}{\frac{1}{\sqrt{2\pi\sigma^2{}_{k(T),x,y}}}\exp\left(-\frac{r^2_{k(t),i,c,x,y}}{2\sigma^2{}_{k(T),x,y}}\right)} \tag{3}$$

Where $r_{k(T),ic,x,y}$ is the linear template response for for signal *i* of contrast *c* at location (*x,y*) given fixation *k(T)*, and $\sigma^2{}_{k(T),x,y}$ is the overall noise variance at location (*x,y*) given current fixation *k(T)*, which is a combination of the external (i.e. pixel) noise and eccentricity dependent internal noise (see next section for further detail).

Each likelihood ratio is weighted by the prior probability of the target appearing at location *x*, yielding a sum of weighted likelihood ratios, *wLR*:

$$wLR_{k(T),i,c,x,y} = priori_{T,i,c,x,y}LR_{k(T),i,c,x,y} \tag{4}$$

Since the MAP model has perfect memory for previous visual input, the *priori* information is the product of the priors ($\pi_{x,y}$; note that this is the only aspect of the model which differs between tasks) given by the cue probabilities and all previous likelihoods. Note that when selecting the first eye movement, this term will only be the priors:

$$priori_{T,i,c,x,y} = \pi_{x,y} \prod_{t=1}^{T-1} LR_{k(t),i,c,x,y} \tag{5}$$

The location of the next eye movement is determined by selecting the location of maximum evidence for all letters combined, summing posteriors across letters and contrasts, but not spatial position:

$$k(T+1) = \underset{x,y}{\arg\max} \sum_{i} \sum_{c} wLR_{k(T),i,c,x,y} \tag{6}$$

The identification decision (*ID*) after multiple eye movements is made considering the total evidence for the presence of each letter, summing likelihoods across spatial position and contrast:

$$ID = \underset{i}{\arg\max} \sum_{x} \sum_{y} \sum_{c} wLR_{k(T),i,c,x,y} \tag{7}$$

A flowchart of the model can be seen in Figure 7.

2. Internal noise and foveation

There are multiple methods to incorporate the degradation in the quality of visual processing with increasing retinal eccentricity into models. These methods include spatially variant filtering of the image with different contrast sensitivity functions (Peterson & Eckstein, 2012, 2013), adjusting the strength of the signal (Najemnik & Geisler, 2005, 2008; Zhang & Eckstein, 2010), and varying the additive internal noise to the scalar response of linear models.

Here, the foveated properties of the visual system were implemented into the model by directly adding zero-mean eccentricity-dependent white internal noise ($\sigma^2{}_{I_{k(T),x,y}}$) to the stimulus image. This method was selected due to the non-linear form of the ideal Bayesian observer which makes the addition of internal noise at the decision level more complex. Therefore, the overall noise variance at any location in the image is given by:

$$\sigma^2{}_{k(T),x,y} = \sigma^2{}_E + \sigma^2{}_{I_{k(T),x,y}} \tag{8}$$

Where $\sigma^2{}_{k(T),x,y}$ is the total noise variance at position (*x,y*), which is the sum of external (pixel) noise ($\sigma^2{}_E$) and the internal noise variance ($\sigma^2{}_{I_{k(T),x,y}}$) at position (*x,y*) given current fixation position *k(T).* The variance of the internal noise is dependent the distance from the current fixation *k(T)*, with the lowest values at the point of gaze and then monotonically increasing with eccentricity. An illustration of the additive internal noise is shown in Figure 8. This eccentricity dependent noise increase serves to decrease the quality of peripheral visual information, simulating the drop in perceptual performance across the across the retina moving away from the fovea. Internal noise levels were selected by developing an ideal observer via an independent visibility map task (described below) and adjusting internal noise to match individual observer performance. A new internal noise sample was selected after each eye movement.

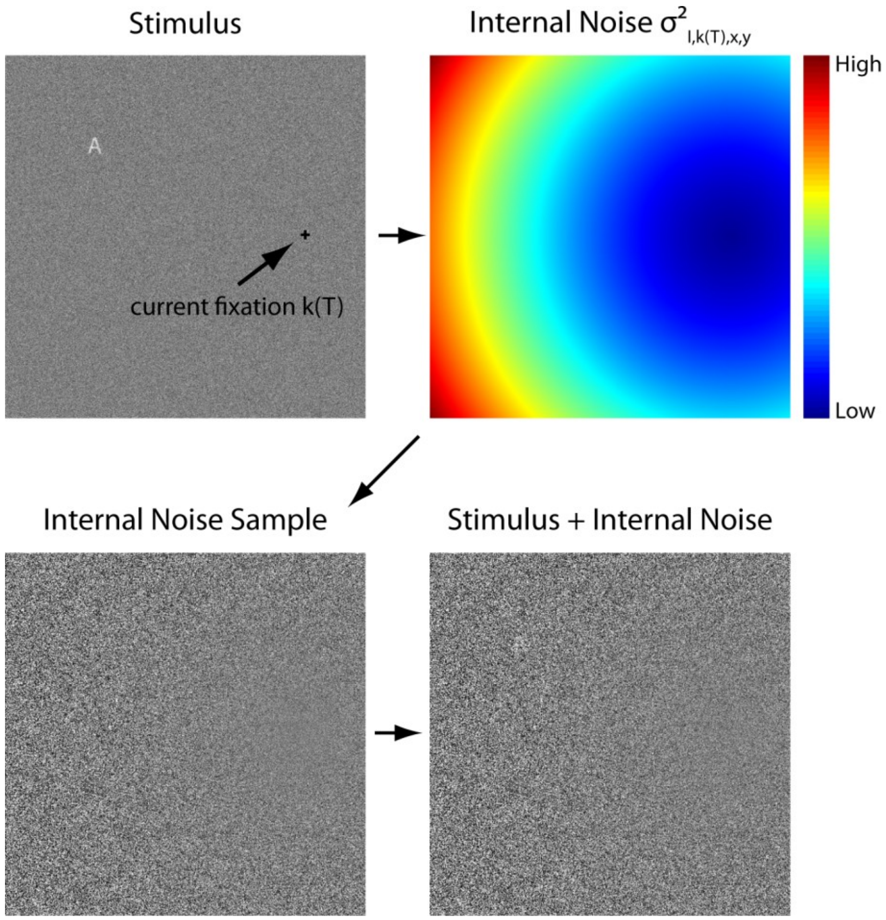

Figure 8. Visualization of the eccentricity dependent internal noise. The internal noise, whose variance increased monotonically with distance from the current fixation point, was added directly to the stimulus image to simulate the decline in visual acuity across in the peripheral retina.

3. Calibrating the model to human visual system properties

Measurement of visibility maps. An established way quantifying how the foveated nature of the visual system affects task-specific performance is through the measurement of visibility maps (Najemnik & Geisler, 2005, 2008; Peterson & Eckstein, 2013; Zhang & Eckstein, 2010). A visibility map measurement generally involves performing a simplified version of the original search task, measuring performance using the same (or similar) stimuli under conditions of forced fixation while systematically varying the retinal eccentricity of target elements. The end result is an estimate of sensitivity (e.g. d') or perceptual performance as a function of eccentricity, which is both task- and participant-specific.

To quantify the task-specific impact on performance of the foveated nature of the visual system, three participants (all who had participated in letter identification search task) completed a two part detect-then-identify task at a systematically varied range of eccentricities. Trials began with participants fixating on a precue image which contained a fixation cross as well as a small black precue dot above which a letter (A-E) would appear in the stimulus image on 50% of the trials. The precue could appear either at fixation (0º) or at 3, 6, 9, or 12º above fixation. Following a 300ms delay, an image of Gaussian noise which could contain a letter at the precued location was presented for a short duration (~100-150ms) which was dependent on the mean first saccade latency for each participant in the original search task. The probability of a target occurring was 50% (i.e. 50% target present). Participants were required to maintain fixation on a small cross throughout the duration of the trail. If a letter was present, its identity and contrast were chosen randomly. The parameters of the noise and letters were identical to those in the previously described tasks.

Participants were required to respond first whether they believed a letter was present. If they responded that a letter was present, they then chose which letter they believed was in the image via mouse click response. Binary feedback (correct/incorrect) was provided for both responses (see Figure 9 for task structure).

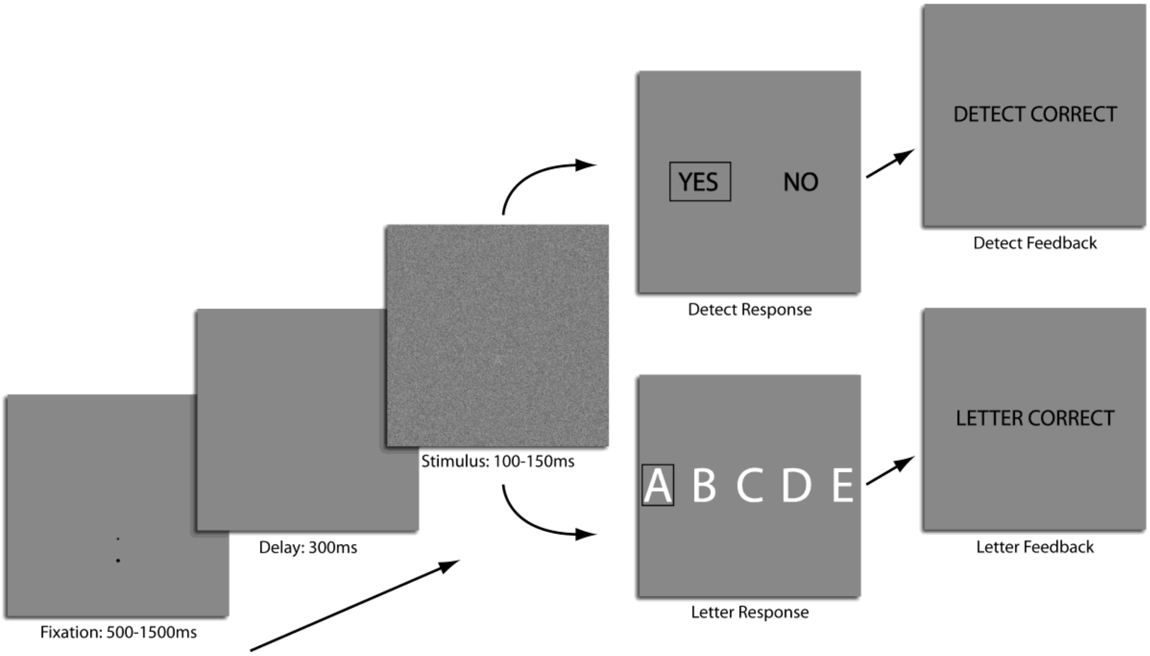

Figure 9. Structure of the visibility map task. The task consisted of a two part detect-then-identify task. If participants indicated that there was in fact a letter present ("yes" response), an identification decision was required.

Visibility map fitting. To estimate the profile of the eccentricity-dependent internal noise variance, visibility map data for each contrast and each participant were averaged across letter identity, yielding a performance profile across the five measured eccentricities. Hit rates, false alarm rates, and identification performance were calculated each contrast and eccentricity. Due to a small, but considerable, number of ceiling (100% hit rate) or floor (0% false alarm rate) values on individual contrast/eccentricity pairings in hit and false alarm data, fits were performed on the raw hit and false alarm data (as opposed to implementing an arbitrary correction on hit/false alarm rates) to avoid attempting to model infinite d'.

The relationship between retinal eccentricity and internal noise standard deviation was assumed to be linear, such that:

$$\sigma^2{}_{I_{k(T),x,y}} = (\sigma_{Fov} + s * ecc_{k(T),x,y})^2 \tag{9}$$

Where σ_{Fov} is the standard deviation of the internal noise at the point of fixation, s is the linear scaling factor, and $ecc_{k(T),x,y}$ is the retinal eccentricity of point (x, y), given fixation $k(T)$. Although more complex monotonically increasing functions could be implemented, a linear relationship was chosen to reduce computational complexity associated with optimizing visibility map fits.

Visibility map simulation. Visibility map task predictions were generated by implementing a simple foveated Bayesian Ideal Observer (fBIO). On each simulated trial, there was a 50% chance a letter of a random contrast and identity would be present at the predetermined target patch. All letter parameters, including external noise variance, were identical to the visibility map task. Additive internal noise, dictated by the linear visibility map described above, was added to the target patch. In the remaining 50% of target absent trials, no letter was present, leaving only the external and additive internal noise.

The fBIO calculated the likelihood ratio of each letter at each contrast appearing at the target location. Note that in the visibility map task there was no spatial uncertainty, as target location was explicitly cued before each trial. The likelihood ratio of a given letter at a given contrast appearing at the target location is given by:

$$LR_{i,c} = \frac{p(g \mid s_{i,c})}{p(g \mid n)} \tag{10}$$

$LR_{i,c}$ is the likelihood ratio of signal i of contrast c, $s_{i,c}$ is signal i of contrast c, g is the noisy data at the target location. Notice that there is no index for position (x,y), as the signal always appeared at a predefined location (i.e. no positional uncertainty).

The sum of likelihood ratios across all possible states (25 signal present and 1 signal absent) was calculated. Signals were present on 50% of the trials, and all letter/contrast signal combinations were equally likely, yielding:

$$sLR = \frac{1}{25}\sum_{i}\sum_{c} LR_{i,c} \tag{11}$$

This sum of likelihood ratios was then compared to a single criterion to determine Yes/No response. If the *sLR* exceeded the criterion (*crit*), the model responded yes. If not, the model selected no:

if *sLR* > *crit*, respond “yes”; else “no” (12)

If the model correctly concluded the target was present, an identification decision had to be made. The identification decision was made by choosing the letter with the maximum likelihood ratio, summed across contrast:

$$ID = \arg\max_{i} \sum_{c} LR_{i,c} \tag{13}$$

Visibility map parameter optimization. Due to the relatively small number of trials in the visibility map task and the resultant noisy data, visibility map fits were performed on the aggregate data of the three observers who performed the visibility map task. Separate visibility maps were fit for detection and identification performance. Two visibility maps were implemented to account for the fact that while the model uses optimal (i.e. matched) letter templates, the human observers likely had suboptimal letter templates, which would likely influence identification performance more than simple detection. Creating separate visibility maps for identification and detection thus allow us to take into account these differences in efficiency without taking on the complex modeling of specific suboptimal templates used by humans.

Internal noise parameters for the detection visibility map were estimated by minimizing squared error in the detection hit and false alarm rates between fBIO predictions and observer data via gradient descent optimization. Similarly, identification visibility maps were estimated by minimizing squared error in identification performance (PC) between the fBIO and observer data. Each iteration of the optimization consisted of 10000 simulated trials. Two free parameters were estimated for both the identification and detection visibility maps: foveal internal noise (σ_{Fov}) and the linear scaling factor (*s*). For the detection visibility map, the criterion (*crit*) was also a free parameter.

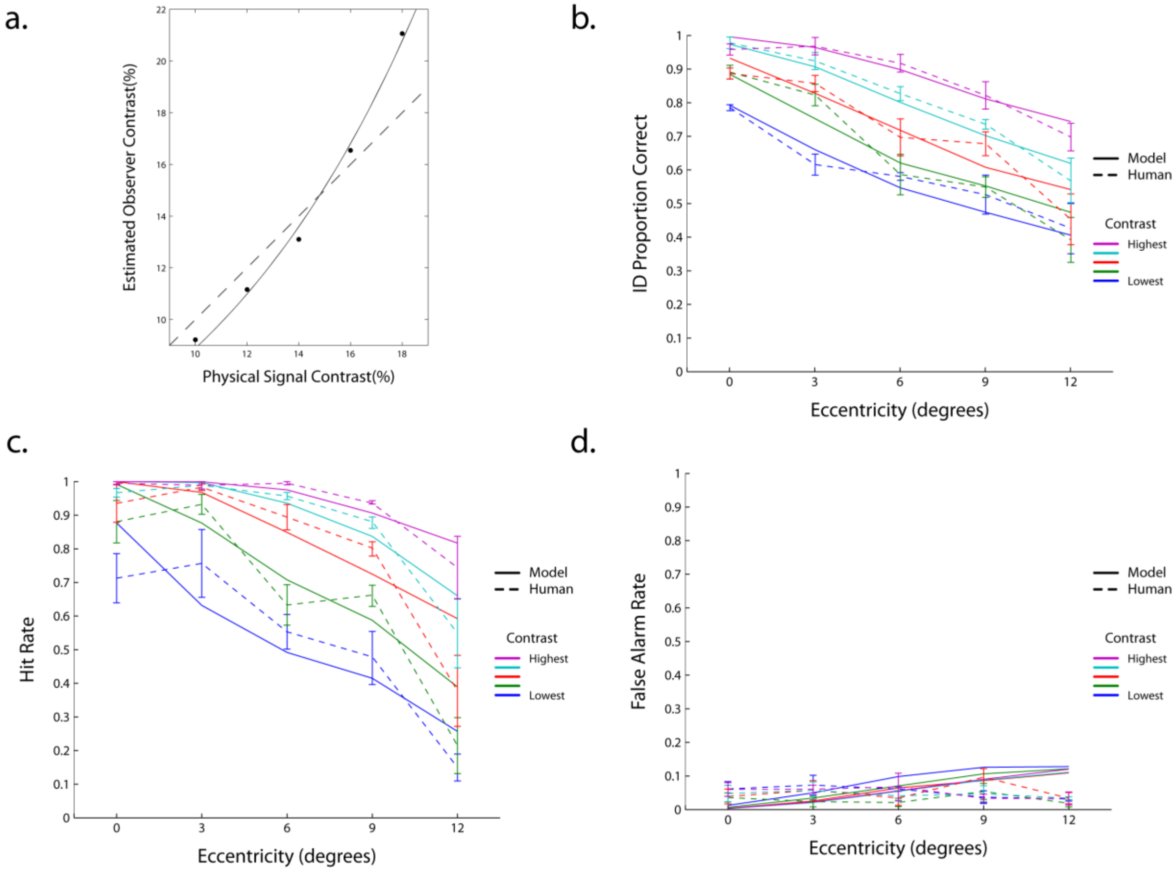

Figure 10. Contrast fitting and visibility map fits. (a) Estimated observer operating contrast as a function of actual signal contrast. (b) ID performance as a function of contrast and retinal eccentricity aggregate human data and ID visibility map fBIO fits. (c) and (d) show observer hit and false alarm rates as a function of signal contrast and retinal eccentricity along with the fBIO detection visibility map fits. Overall, the visibility maps correspond well to observer data in the visibility map task.

Signal contrast scaling. Driven by known nonlinearities between signal contrast and sensitivity (d'), in which sensitivity is much lower than would be predicted by a linear model at low signal contrasts (Beutter et al., 2003; Eckstein, Ahumada, & Watson, 1997), we modulated the contrast of the signal to better reflect observer sensitivity when doing the task. The concept of modulating signal contrast to account for this effect is well established (Dosher & Lu, 2000; Lu & Dosher, 2008) Gradient descent optimization was utilized to estimate the effective signal contrasts (i.e. the signal contrasts that humans appeared to be operating at) within the visibility map fBIO. Using the best fit detection and identification visibility maps, signal contrast was free to vary to minimize squared error between observer hit and false alarm rates and model predictions.

The relationship between actual signal contrast and estimated observer operating signal contrast is seen in Figure 10a. Individual points represent the best fit contrasts to match human performance with the visibility map fBIO. The solid line represents a best fit exponential expression, while the dotted line indicates the identity (i.e. estimated signal contrast = physical signal contrast). Note that the contrast of the signal is scaled down at lower contrasts (falls below the identity), and slightly up at higher signal contrasts (falls above the identity). Since the best fit visibility maps were done on average across all signal contrasts, it is reasonable to expect underprediction of internal noise at low signal contrasts (effectively increasing signal strength) and overprediction of internal noise at high signal contrasts. Figure 10b-c show aggregate human data for identification and detection (hit and false alarm rate) elements of the visibility map task. Best fit fBIO predictions for identification and detection (using the scaled signal contrast) show good correspondence for both, indicating they were a reasonable estimate of human visibility across the range of contrasts and eccentricities used in the current study.

E. Human-Model Comparisons

Ten thousand simulated trials were run for the foveated MAP model at each signal contrast for both cues present and cues absent conditions. As mentioned in the previous section, the only aspect of the model which differed between cues absent and cues present simulations was the priors ($\pi_{x,y}$). That is, in cues present trials the target could only appear at the four locations indicated by the cue circles.

Note on human-model analyses and comparisons. For all human-model comparisons presented in this dissertation, the error bars for model predictions were small enough that they could not be displayed. As a result, statistical tests between human data and model predictions were run as single sample t-tests against the model prediction.

Identification performance as a function of signal contrast for the foveated MAP and human observers is shown in Figure 11a. Like human observers, cues present performance exceeded that of cues absent overall. Model predictions also displayed a pattern of performance where the largest benefit from cues is gained at the lowest signal contrasts, with this performance gain diminishing with increasing signal contrast. In comparison to human observers, when cues were absent, model predictions mirrored human performance and were statistically indistinguishable ($t(4)=-1.5, p=.12$). However, there were large differences between predicted cues present performance and human data ($t(4)=2.75$, $p=.02$), and thus, the predicted performance increase in the presence of spatial cues.

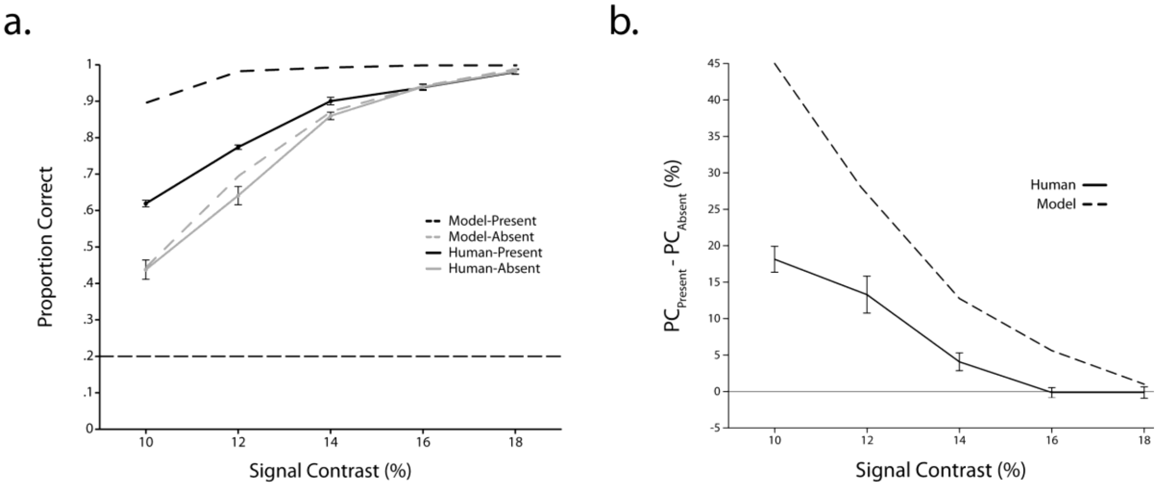

Figure 11. Comparison of identification performance for human and model observers. Panel (a) shows that, while the foveated MAP predicts human performance on cues absent trials well, model performance far exceeds that of human observers when cues are present. The magnitude of this difference is apparent in panel (b), which shows the cue benefit as a function of signal contrast. Over all signal contrasts, the model gains a much larger benefit from cue presence than human observers, indicating possible suboptimality.

Direct comparison of the facilitatory nature of the predictive cues on performance can be seen in assessment of what we deem the cue benefit, defined as $PC_{present}$-PC_{absent} (see Figure 11b), which provides a simple metric of the additional gain in performance afforded by the presence of predictive spatial cues. Although the general shape of the cue benefit as a function of contrast is similar between human observers and foveated MAP predictions, the model gained a substantially larger benefit from the spatial cues across all contrasts ($t(3)=-11.59$, $p<.001$), although this difference is particularly great at the lowest signal contrasts. Thus, while human performance qualitatively follows that of a near-optimal foveated MAP searcher, humans do not gain anywhere near the magnitude of benefit from predictive cues that would be expected from a MAP searcher.

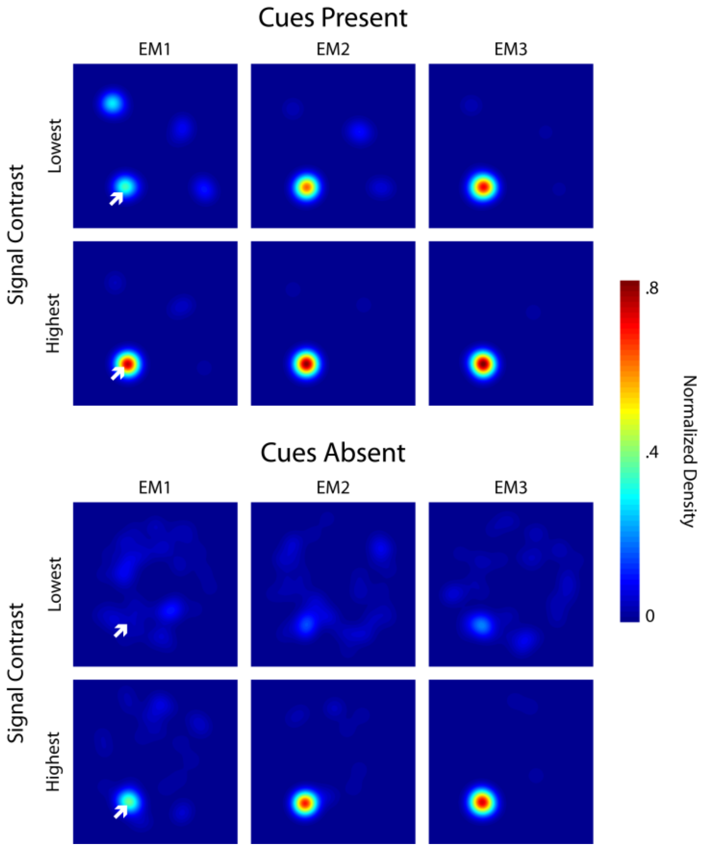

Figure 12. Normalized eye movement distributions for the model observer. Eye movement selection is remarkably similar to human observers on comparable trials (see Figure 3), with eye movements clustering around cued locations when cues are present, and relatively dispersed when not at the target location (indicated by the white arrow) when cues are absent. Note the scale difference between human and model observers, indicating higher densities at highly fixated location.

Eye movement analyses serve to strengthen this interpretation. Normalized distributions of the first three eye movements for the foveated MAP across signal contrasts and cueing condition are shown in Figure 12. As with observer data, eye movements when cues are present are heavily concentrated on both the target (indicated by a white arrow) as well as other cued locations. Since a MAP searcher can only move its eyes to potential target locations, eye movements are actually exclusively distributed at these locations when cues are present. When cues are absent, eye movements spread throughout the image when not at the target location. For both cues present and absent, as trials progress (i.e.

more information is accumulated) and signal strength increases, the density of fixations at the target location increases.

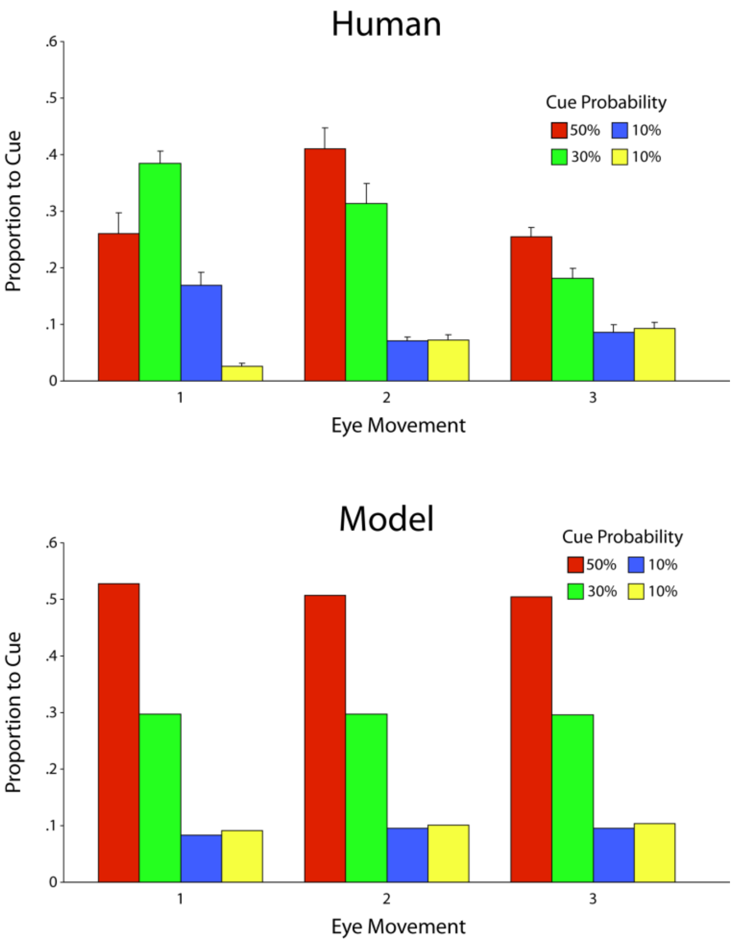

Figure 13. Proportion of eye movements to each cue for humans and model. All model eye movements are directed towards the cues, and the proportion of eye movements to each cue mirrors the underlying probability of the target appearing at that location. In contrast, human observers direct less eye movements towards cues as trials progress, and the proportion of eye movements to each cue varies across eye movement. Critically, the first human eye movement tends to target the 2nd highest probability cue, which falls the shortest mean distance from all fixation points.

We can also directly compare the nature of model cue-driven saccades to that of human observers. Figure 13 shows the proportion of model and human saccades to each cue (when present) as a function of eye movement. Two critical points emerge. First, the MAP model directs *all* of its eye movements towards the cued regions, as it can only move its eyes to potential target locations. While human observers directed the vast majority of their eye movements to cues when present (*M*=77.5%), they fell well short of foveated MAP predictions in that regard ($t(3)=-17.23$, $p<.001$). Second, the distribution of the eye movements that went towards the cues differed quite visibly between model and humans. For the model, eye movement selection to cues quite closely mirrors the underlying cue probabilities, with more likely target locations garnering more fixations. In contrast, human cue preference varies as a function of eye movement. As discussed above, human observers tend to direct their first saccade to the cue of 2nd highest likelihood of containing the target, which we suggested may be related to a preference for shorter saccades. The foveated MAP has no such preferences and incurs no cost from making long range eye movements, and considers a saccade of 1° to be just as viable as a saccade of 15°. This marked disparity in eye movement behavior highlights a potential source of suboptimality underlying the lower than predicted human performance when cues are present.

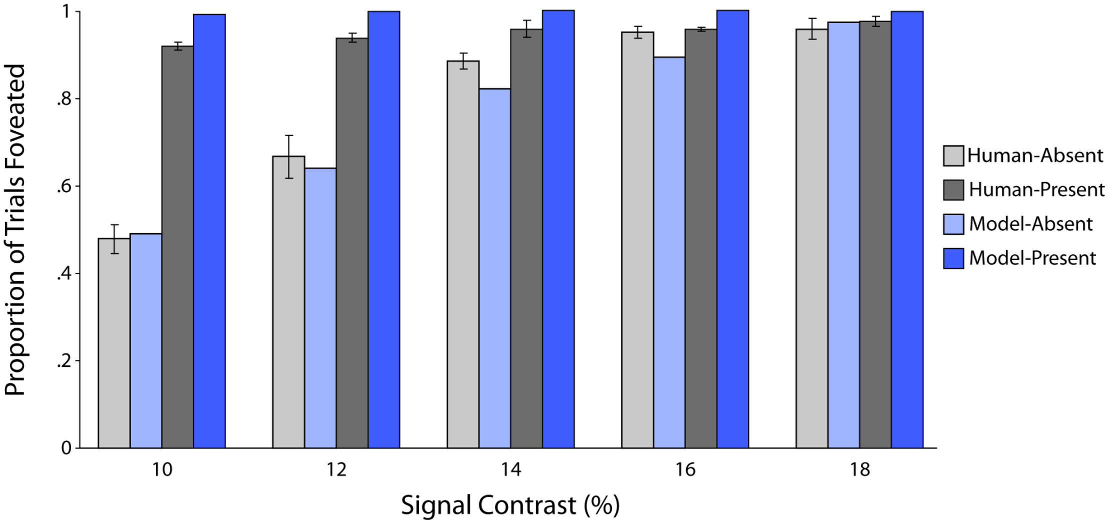

Figure 14. Proportion of trials with target foveated for model and human observers. The foveated MAP foveated the target slightly, but significantly, more often than human observers when cues are present.

The functional consequences of these eye movements are illustrated in Figure 14, which shows the proportion of trials in which the target was foveated (i.e. eye position within 2° of visual angle of the target) as a function of signal contrast for both cues present and absent conditions for the foveated MAP. Mirroring trends in observer data, the proportion of trials in which the target was foveated was overall considerably higher when cues were present than when absent. The size of this disparity was largest at low signal contrasts and steadily declined with increasing signal strength, reflecting the increased ease of detecting the target with increasing contrast. Interestingly, the differences between model predictions and human data in this case are quite small. The predicted proportion of trials foveated in cues absent sessions did not significantly differ from human data ($t(4)=-1.46$, $p=.22$), and while the model foveated the target significantly more often than human observers in cues present sessions ($t(4)=5.81$, $p=.004$), the size of the difference (94% vs. 99%) is rather small compared to the vast difference in perceptual performance between human and model on those same trials.

Alternative models. Given the considerable differences in human and model performance in cues present trials, as well as the disparity in the distribution of first eye movements to cues, one alternative explanation of human data were that human observers were mistakenly switching the predictive value of the top two most predictive cues (i.e. assuming the 2nd most predictive cue was actually the most predictive). The foveated MAP was altered so that the priors of the top two cued locations were switched (i.e. the 30% region was swapped to 50%, and vice versa). It is important to note that the priors underlying the actual *stimuli* present to the model did not change.

This switched prior model, however, produced negligible differences compared to the original foveated MAP (with optimal priors). In fact, at all contrast levels, predictions of the switched priors model were within ±0.5% in performance of the unaltered foveated

MAP, suggesting that a misattribution of prior values was not sufficient to describe the suboptimal human performance when cues were present.

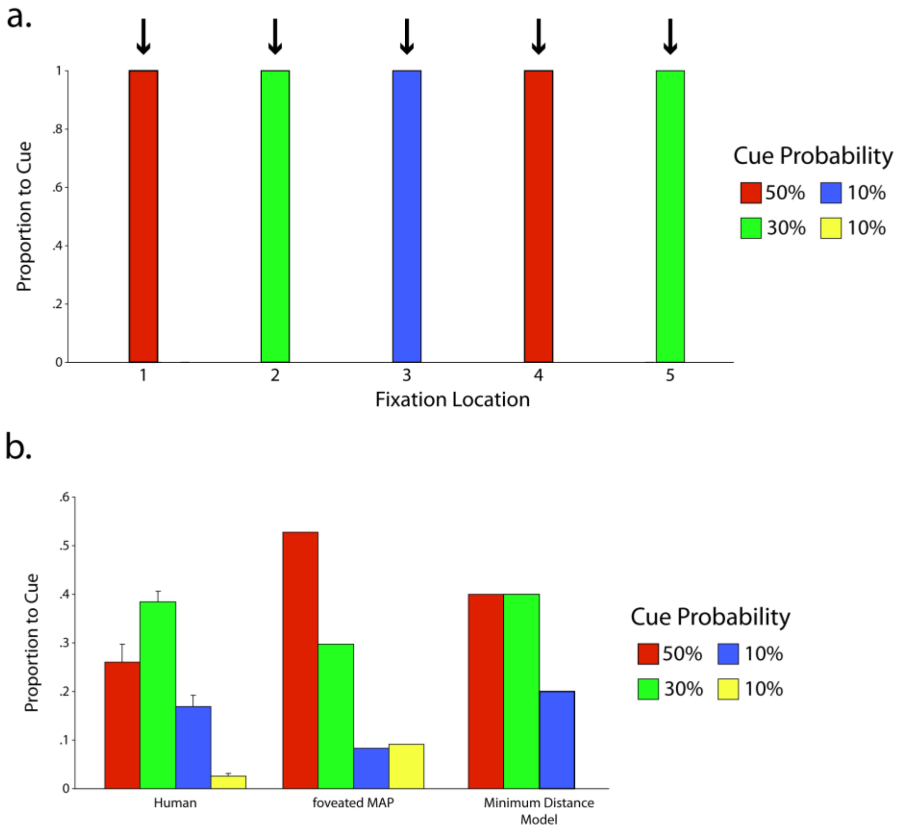

Figure 15. Comparison of eye movement selection for human observers and minimum distance eye movement model. (a) Shows the deterministic distribution of first saccade selection for the minimum distance model. The model always moves its eyes to the cued location closest to the initial starting fixation (shown by the arrow). Human eye movement distributions are also biased to the closest cue, but not to such an extreme. (b) Comparison of the overall distribution of first saccades to cued regions . The overall pattern of saccade distribution for humans is better predicted by a minimum distance model, although there are considerable differences, indicating other factors influence human eye movement selection.

We also suggested that the eye movement distributions to cues exhibited by human observers, particularly for the first eye movement, may be attributed to a strategy which is more likely to go to nearby cue locations, even if their predictive value isn't as high as further away alternatives (Araujo et al., 2001). To evaluate this interpretation, we implemented a variant of the foveated MAP in which the eyes always moved to the nearest cued location, regardless of the visual information in the image. The proportion of first eye movements to each cue as a function of fixation location can be seen in Figure 15a. The deterministic nature of the model is clear in that the fixation location completely determines where the first eye movement will go. There are some similarities between human first eye movement selection in the minimum distance model (e.g. the preferred location for the first eye movement is the same). However, the broader distribution in human eye movement selection patterns suggest that they are taking the visual information into account, and not choosing the first saccade solely on distance. Comparison of overall distribution of first saccades for human observers, the standard foveated MAP, and the minimum distance model can be seen in Figure 15b. Despite considerable differences, it is clear that the cost incurred by making a long distance eye movement is surely a part of how human observers selected their eye movements in the current task, as the distance model provides relatively good descriptor of human eye movement distributions, and certainly a better one than the standard foveated MAP.

F. Conclusions

Predictive spatial cues nearly invariably enhance perceptual performance on a variety of visual tasks (Cameron et al., 2002; Posner, 1980; Talgar et al., 2004), including single fixation visual search (Hawkins et al., 1990; Nakayama & Mackeben, 1989). However, whether these effects translate to multi-fixation visual search and targets of unknown and varying detectability, conditions more akin to natural visual search, is poorly understood.

In a letter identification task with an array of predictive cues, human observers were able to reliably direct eye movements to spatially predictive cue elements and displayed enhance search performance, particularly at low signal contrasts. This increase in perceptual performance was coupled with a pattern of eye movements which preferentially foveated predictive cues when present, leading to increased foveation of the target, ostensibly, enhanced perceptual performance. These results mirror the common finding in scene context literature that humans are willing and able to move their eyes to elements of the visual environment that predict target location to enhance perceptual performance (Castelhano & Heaven, 2011; Eckstein et al., 2006; Mack & Eckstein, 2011; Neider & Zelinsky, 2006; Spotorno et al., 2014; Torralba et al., 2006). Additionally, human observers seemed to bias their initial eye movement towards the closest cued locations when cues were present, regardless of cue predictiveness, indicating that additional strategies likely contributed to observer saccade distributions. However, what remains unresolved is how *well*, in an objective sense, human observers integrate this predictive information into both their eye movement behavior and perceptual decisions.

The implementation of a foveated MAP eye movement model, which takes into account the decreasing sensitivity of the human visual system with retinal eccentricity, provided an objective benchmark of eye movement selection and perceptual performance against which to compare human behavior. Although MAP searchers are not truly optimal models of visual search (Najemnik & Geisler, 2005), they often approximate ideal search behavior (Najemnik & Geisler, 2008; Zhang & Eckstein, 2010) and are good predictors of human eye movement distributions (Beutter et al., 2003; Verghese, 2012).

The predictions of the foveated MAP searcher qualitatively mirrored the search behavior and performance of human observers, as identification performance was enhanced overall in the presence of predictive cues, and most markedly so at low signal contrasts. This enhancement in performance was also reliably related to model eye movements which preferentially (and, in fact, exclusively) targeted cued locations when cues were present, leading to enhanced probability of foveating the target location.

However, the foveated MAP exceeded human observers substantially in the benefit gained by the predictive cues (i.e. the cue benefit), indicating considerable suboptimality in human search strategy and execution. Potential insight into the origin of this suboptimality may be seen in the eye movement selection by both human and model observers. Since the MAP can only move its eyes to potential target locations, *all* eye movements were directed towards the cues. In contrast, only 78% of human eye movements foveated the cued regions when present. Perhaps human observers could have benefited from a strategy which more tightly targeted cued locations when present. Additionally, observers departed substantially from model predictions in how they chose to distribute their saccades to cues, particularly during the first eye movement. This apparent preference for shorter, lower physical costs saccades has been documented in human observers (Araujo et al., 2001; Kowler, 2011) and has been shown to persist even at the cost of task performance. However, despite a small but significant difference, human observers managed to foveate the target nearly as often as the model in cues present trials. Thus, while there were considerable differences in observed and predicted eye movement behaviors, it seems likely that sources of suboptimality underlying the vast difference in model and human performance in cues present trials likely extend beyond saccadic selection.

A number of other factors also may have played a role in the discrepancies between human and foveated MAP perceptual performance and eye movement behaviors. First, the MAP model integrates information optimally and with perfect memory over time and space (Geisler, 2003, 2011; Watson, 1987). However, it is well known that human observers do not do this optimally (Irwin, 1991, 1996; Najemnik & Geisler, 2005), potentially leading to wide-reaching suboptimalities in perceptual performance and eye movement selection. A MAP model also makes its decisions based off of a Bayesian ideal decision rule, which consists of calculating and aggregating a set of complex likelihoods (Eckstein et al., 2002; Geisler, 2011; Shimozaki, Schoonveld, & Eckstein, 2012). Calculation and maintenance of such complex terms given limited neural machinery seems dubious at best. Additionally, factors which fall outside the scope of the model such as spatial masking (Polat & Sagi, 1993; Smith, 2000) or crowding (Balas, Nakano, & Rosenholtz, 2009; van den Berg, Roerdink, & Cornelissen, 2007; Whitney & Levi, 2011), in which visual elements adjacent to the target can perceptually degrade its appearance, via the cues could certainly have played a role in reducing perceptual performance when the target appeared at those locations.

In all, it is clear that observers are capable of leveraging predictive cue information to guide eye movements and enhance search performance. However, their proficiency in doing so falls short of a near-optimal eye movement model, indicating substantial suboptimality in eye movement selection, the mechanisms underlying perceptual decision making, or both.

III. Observer flexibility and peripheral predictive information (remote cues)

In Chapter II, we saw that human observers were able to utilize predictive spatial cues (albeit suboptimally) to guide visual search and enhance search performance. Our results provide a quantitative backing to complement a growing literature that humans often direct their eyes towards toward objects which often spatially co-occur with search targets (Castelhano & Heaven, 2011; Eckstein et al., 2006, 2006; Torralba et al., 2006).

But what about the cases when predictive information about target location does not appear right next to (or around) it? Here, we probe human observers' ability to flexibly adopt new and atypical search strategies when predictive visual information occurs spatially separated from potential target locations.

We report a scenario in which humans initially look away from the target in the presence of remote spatial cues to optimize performance. Observers performed a free eye movement search task in which they were to indicate the presence of a Gabor embedded in white luminance noise. The target appeared at one of ten locations arranged in two semi-circles. The orientation of three remote spatial cues (large C's), situated between the two groups of target locations indicated where the target would appear if it were present.

Observers performed task variants in which the cues differed in whether they provided target location information, how visible the cue information was, or whether the cues could be foveated. Performance when remote cues were predictive vastly exceeded that of when the cues did not predict location. Moreover, observers systematically directed early saccades towards the cues when they provided target location information. When cues were non-predictive, they were largely ignored and a sequence of eye movements which surveyed as many target locations as possible prevailed. Critically, when cues were predictive but observers were prevented from foveating them, performance suffered compared to when saccades were unrestricted, highlighting the importance of a cue-driven eye movement strategy.

Three variants of a foveated MAP eye movement model are developed, which differed in their saccadic planning and whether they used peripheral remote cue information. It is shown that when remote cue information is unavailable, a MAP eye movement strategy closely approximates human eye movement distributions and perceptual performance. However, when the remote contain information predictive of target location, incorporating cue-driven eye movements is necessary to match human proficiency.

Finally, a learning version of the remote cue task, in which participants were given no information about the predictive nature of the remote cues, was run to assess how willing observers would be to adopt atypical eye movement strategies when the benefits are not made explicitly clear. Only two of nine observers effectively learned and utilized the remote cues over the course of two thousand trials, indicating that while observers may be effective at adopting atypical eye movement strategies when the benefits are apparent, they are not necessarily actively looking to do so.

In all, while moving the eyes to likely target locations is a common and often effective strategy, we show that human observers can adopt atypical eye movement strategies when this behavior has clear benefits for the current search task.

A. Introduction

Due to inhomogeneities in the anatomy of the retina, visual search requires the orienting of the high resolution fovea to points of interest in the visual environment to gather information about potential target locations and identities (for review, see Eckstein, 2011). Since the time to complete search is rarely unlimited, selecting efficient and informative eye movements is of high priority (Najemnik & Geisler, 2005). As such, what types of information drive eye movement selection has been the focus of a wealth of research.

Not surprisingly, human observers have been shown to consistently direct their eyes towards image regions that resemble the target (Beutter et al., 2003; Findlay, 1997; Malcolm & Henderson, 2010; R. P. N. Rao et al., 2002), resulting in a number of computational models of search which utilize the known or expected appearance of the target to guide model saccades (Eckstein et al., 2006; R. P. N. Rao et al., 2002; Wolfe, 1994; Zelinsky, 2008). Additionally, a variety of studies have also shown that observers tend to bias saccades towards either scene regions (i.e. scene context; Brockmole & Henderson, 2006; Castelhano & Heaven, 2011; Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006) or individual objects which tend to spatially co-occur with the search target (Castelhano & Heaven, 2011; Eckstein et al., 2006; Mack & Eckstein, 2011). As a result, many modern models of search now explicitly incorporate scene context into the guidance of eye movements (Ehinger et al., 2009; Torralba et al., 2006).

Although the guidance of eye movements in search via target appearance and scene context reflect markedly different search strategies and underlying knowledge about the task at hand, their effects are largely similar: observers' eyes are directed towards locations likely to contain the target (often referred to as saccadic targeting). Not only has this strategy shown to be effective behaviorally (Eckstein et al., 2006; Mack & Eckstein, 2011; Torralba et al., 2006), but it also corresponds closely to a Bayesian maximum *a posteriori* (MAP) eye movement model (Beutter et al., 2003; Eckstein et al., 2006; Torralba et al., 2006). A MAP model optimally weights visual evidence for target appearance at each location in the visual display by prior knowledge (via context, object co-occurrence, or spatial cues) about how likely the target is to appear at that location and moves its eyes to the location of the highest posterior probability of containing the target. Although MAP models are not truly optimal in terms of maximizing perceptual performance, under many circumstances they approximate a computationally ideal searcher (as derived by Najemnik & Geisler, 2005). Given its strong computational underpinnings, it's not surprising that observers would adopt such an eye movement strategy.

Despite the benefits of saccadic targeting, there are often cases in which objects in distal regions of the visual environment contain information about the possible location of the target. In these cases, an observer which only moves its eyes to likely target locations may miss out on critical task-relevant information. For example, take the scenario depicted in Figure 16. When searching for the man in the image (labeled "target"), most observers would direct their eyes to street level since (which experience suggests is where pedestrians nearly invariably appear) and search through the individuals in the image (Ehinger et al., 2009; Torralba et al., 2006). However, if told that the man will appear underneath a red sign, a rational observer should direct their eyes to well above street level to locate the sign (even though the target cannot appear there) and vastly reduce potential target locations, increasing search efficiency.

Figure 16. Example of a remote cue in a natural scene. If told to search for the man with no further instructions, observers tend direct the eyes toward likely pedestrian locations at street level **(Torralba et al., 2006)**. However, if told the man will appear under a red sign, a rational observer should first move its eyes to a location where a target cannot appear (the sign) to subsequently reduce uncertainty about target location.

The idea of moving the eyes to a location which has no chance of containing the target to enhance search performance is absent in the current literature and clearly contradicts the well established saccadic targeting strategy utilized by observers and mirrored in modeling efforts. Although the ideal searcher and human observers have been shown to move their eyes to non-target locations, these so-called "center of mass" saccades target the centroid of a cluster of potential target locations to gather information about multiple nearby regions of interest (Najemnik & Geisler, 2005, 2008; Zelinsky, 2008; Zhang & Eckstein, 2010), and thus are largely target driven. As such, how well observers utilize predictive elements in the environment which are spatially separated from the target, which we call remote cues, and how effectively they can learn the spatial contingencies of remote cues in the absence of instruction are poorly understood.

Here, we present a scenario in which, for a simple search task in white noise, human observers initially move their eyes away from potential target locations the presence of spatially predictive remote cues to optimize performance. Moreover, we show that manipulating the availability and quality of remote cue information leads to drastic alterations in eye movement selection and search performance consistent with a rational observer. We develop three variants of a foveated maximum *a posteriori* (MAP) eye movement model for the task. The model reveals that cue-driven eye movement strategies are likely necessary to reach search proficiency displayed by human observers. Additionally, while we show that human observers are able to break from a common "saccadic targeting" strategy when given explicit information about remote cue contingencies, they have difficulty learning peripheral cue information in the absence of instruction. This difficulty in learning, however, is separable from ability to effectively use the remote cues when later given explicit instruction.

Thus, while moving the eyes to potential target locations is a valuable and well established search behavior, human observers appear capable of adopting atypical eye movement strategies when doing so enhances task performance. However, observers may be slow (or unable) to adopt such atypical patterns if the predictive nature of peripheral scene elements is not explicitly understood.

B. Remote cue study

To see whether human observers could break from a target-driven eye movement strategy to enhance search performance, participants performed a simple search task in which they were to determine the presence of a Gabor embedded in luminance noise in the presence of three spatially remote cues. The presence and visibility of the remote cue information was manipulated to determine the relationship between eye movement selection and the use of remote cue information.

1. Method

Observers. Observers were three undergraduate students (all female; ages 19-21) at the University of California-Santa Barbara. All observers had normal or corrected to normal vision.

Stimuli. Observers performed four variants of a simple search task which consisted of determining whether a single Gabor (12 cycles/°, SD of spatial envelope .5°)[1] embedded in white noise (25±4.9 cd/m²) was present at any of ten target locations indicated by thin black circles arranged in two evenly spaced semi-circles. The high spatial frequency and low contrast of the target were selected so that accurate detection would be extremely difficult if it were not foveated. Three remote cues were situated between the two clusters of target locations. The middle cue was in the center of the display with the peripheral cues at an eccentricity of 5° to either side. Target locations were an additional 5° eccentric from the respective peripheral cue (see Figure 17).

---

[1] Gabor contrast was determined on a per observer basis so that performance in the absence of predictive remote cues was approximately 60%.

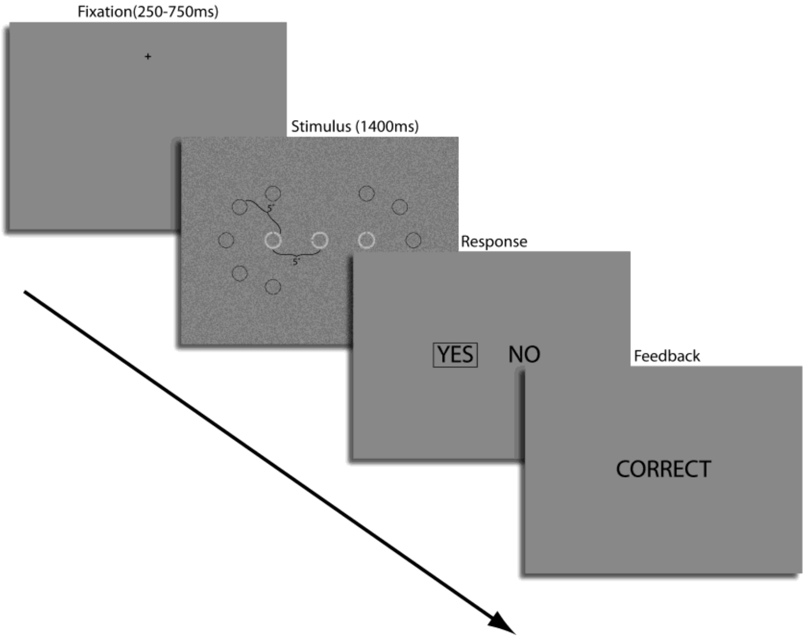

Figure 17. Temporal structure of a trial. The general structure and timing for all remote cue conditions was identical.

Task. Each trial began by the observer fixating on a randomly selected one of two fixation points located 7.4º directly above and below the center of the display (randomized 250-750 ms). A noisy stimulus image then appeared for 1400 ms (with a new noise field sampled every 100ms), easily enough time to make 3-5 saccades, during which observers were free to move their eyes to determine if a target was present in the image. A target was present in 50% of trials. Following stimulus presentation, observers indicated via mouse click whether they believed a target was present in the image and were provided binary (correct/incorrect) feedback about their response (see Figure 15 for trial structure). Observers performed 800 trials per remote cue condition.

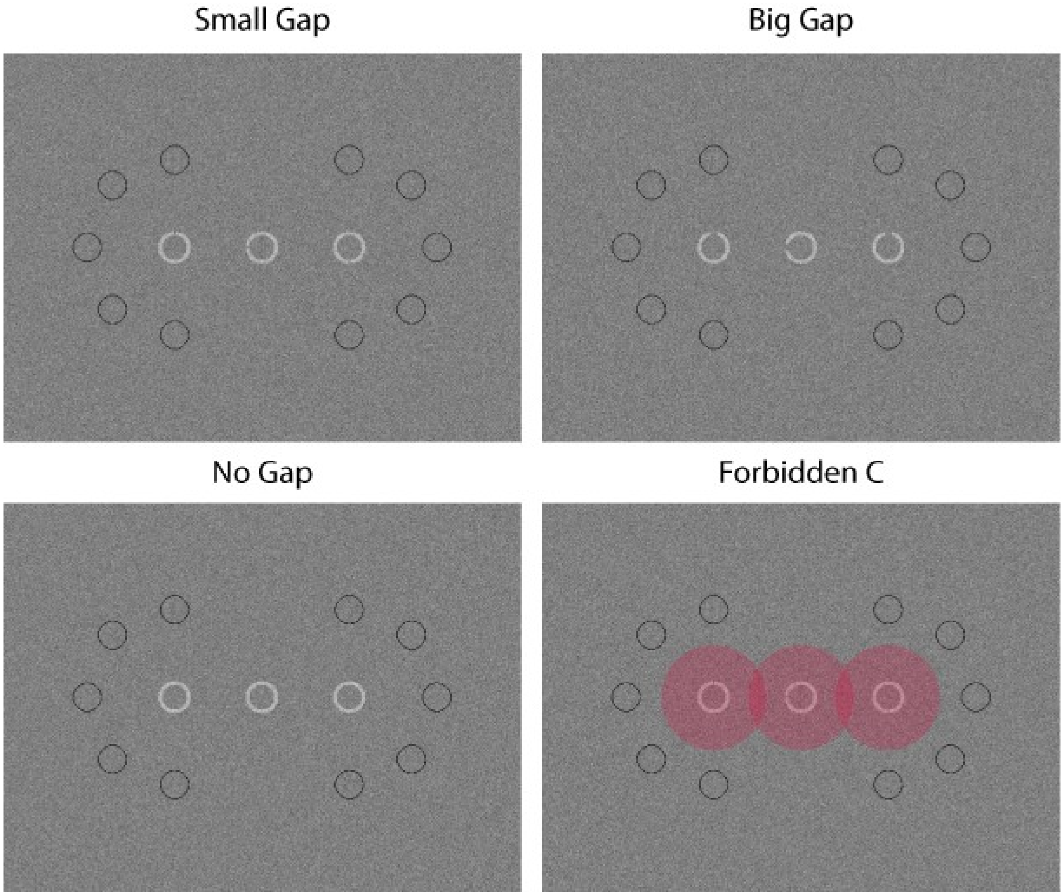

Figure 18. Illustration of remote cue conditions. In Small, Big Gap, and Forbidden C conditions, the orientation of the middle cue indicates which side the target will appear on if present while the cue on the target side points to the target location if present. The cue on the non-target side points to a random target location in that semi-circle. The remote cues in the No Gap condition are closed and provide no information about target location. Quality of remote cue information was manipulated via the size of the cue opening (Small vs. Big Gap) or whether the cues could be foveated (Small Gap vs. Forbidden C). Red circles indicate the area within which observers could not move their eyes in the Forbidden C condition and were not actually present in the stimulus.

Remote Cue Conditions. Stimuli and eye movement instructions were modified between conditions to manipulate both the availability and quality of predictive spatial information that the remote cues afforded. In the Small Gap condition, eye movements were unrestricted and three large C's (diameter 1.85°) served as remote cues and were fully predictive of target location (see Figure 18). The direction (left/right) of the opening in the central cue (.075°) indicated what side the target would appear on if present, while the orientation of the peripheral cue on that side indicated which of the five locations would contain the target if present. The peripheral cue on the non-target side was oriented randomly to one of the target locations in that semi-circle.

To determine how the visibility of remote cue information influenced eye movement selection and performance, two additional conditions were devised. In the Big Gap condition, the visibility of remote cue information was increased by widening the gap in the C’s (.37º) and eye movements were left unrestricted. For the Forbidden C condition, the appearance of the stimuli were exactly the same as in the Small Gap condition, but observers’ eye movements were restricted in that they were not allowed to move their eyes within 3° of any of the remote cues, degrading the quality of cue information that could be sampled. If observers moved their eyes into the restricted area, they received a message indicating the error and the trial was aborted. Finally, to provide a baseline against which to compare potential remote cue benefits, the No Gap condition employed remote cues that did not provide any predictive spatial information as the gaps in the C's were closed. There was no restriction of eye movements.

With the exception of remote cue appearance and restriction of eye movements in the Forbidden C condition, task structure and stimulus appearance was otherwise identical between conditions. Observers were informed verbally about the meaning of the remote cues before each condition.

Monitoring Gaze. Gaze was monitored using an Eyelink 1000 infrared eye tracker (SR Research) providing a monocular left eye track at a sampling rate of 250 Hz. Changes in eye position that exceeded a velocity of 35º/sec and an acceleration of 9500º/s/s were considered to be saccades. The first change in eye position which deviated 2° from initial fixation was considered the first saccade for purposes of analysis.

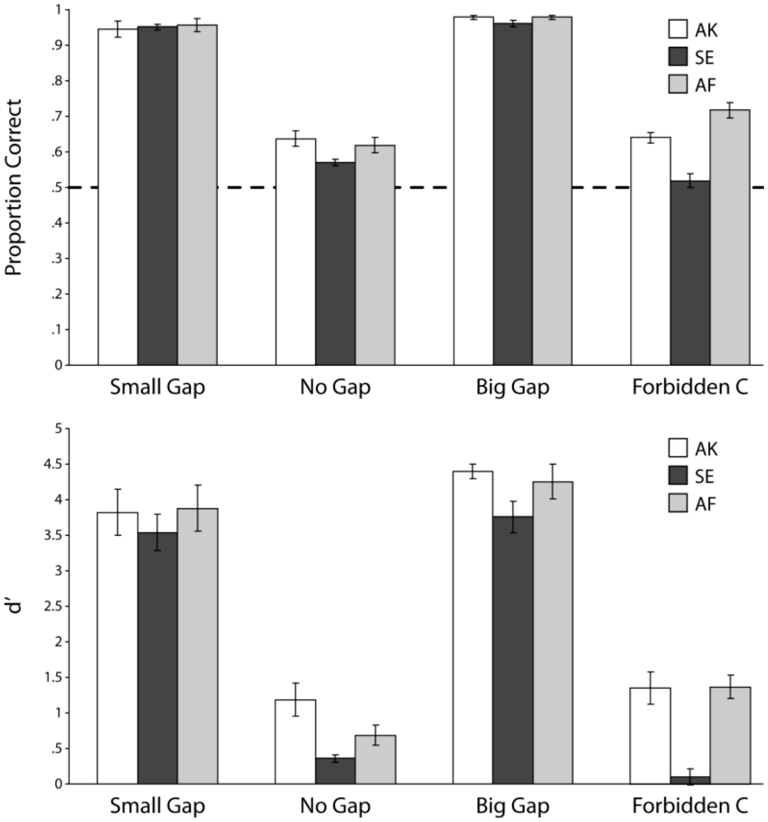

Figure 19. Performance in the remote cue task. Performance, in terms of both proportion correct and d', was markedly improved when remote cue information was present and able to be foveated (Small Gap and Big Gap) as compared to when it was unavailable (No Gap) or available but foveation of the cues was not allowed (Forbidden C). Dotted line indicates chance performance in terms of proportion correct. Chance for d' is a value of 0. Error bars represent ±SEM.

2. Results and discussion

Both the quality and the availability of remote cue information drastically impacted search performance. A one way repeated measures ANOVA (*F*(3,6)=55.42, *p*<.001) revealed that while observers performed (in terms of proportion correct) nearly at ceiling when they were able to freely use predictive remote cues (Small Gap and Big Gap conditions), a reduction in the quality of remote cue information by preventing foveation of cue (Forbidden C) or elimination of predictive cue information altogether (No Gap) reduced performance to barely above chance levels (paired t, *p*<.05; see Figure 19). In fact, when remote cues could not be foveated, performance was indistinguishable from when there was no predictive information in the images at all (*t*(2)=.36, *p*=.37). A nearly identical picture can be seen in analysis the of d', a measure of sensitivity which is invariant to individual variation in decision criterion (i.e. propensity to say "present" or "absent"; see Green & Swets, 1989 for a review). A one way repeated measures ANOVA revealed significant differences in sensitivity across remote cue conditions (*F*(3,6)=9.67, *p*<.001). The direction of all pairwise comparisons was identical to those described for proportion correct.

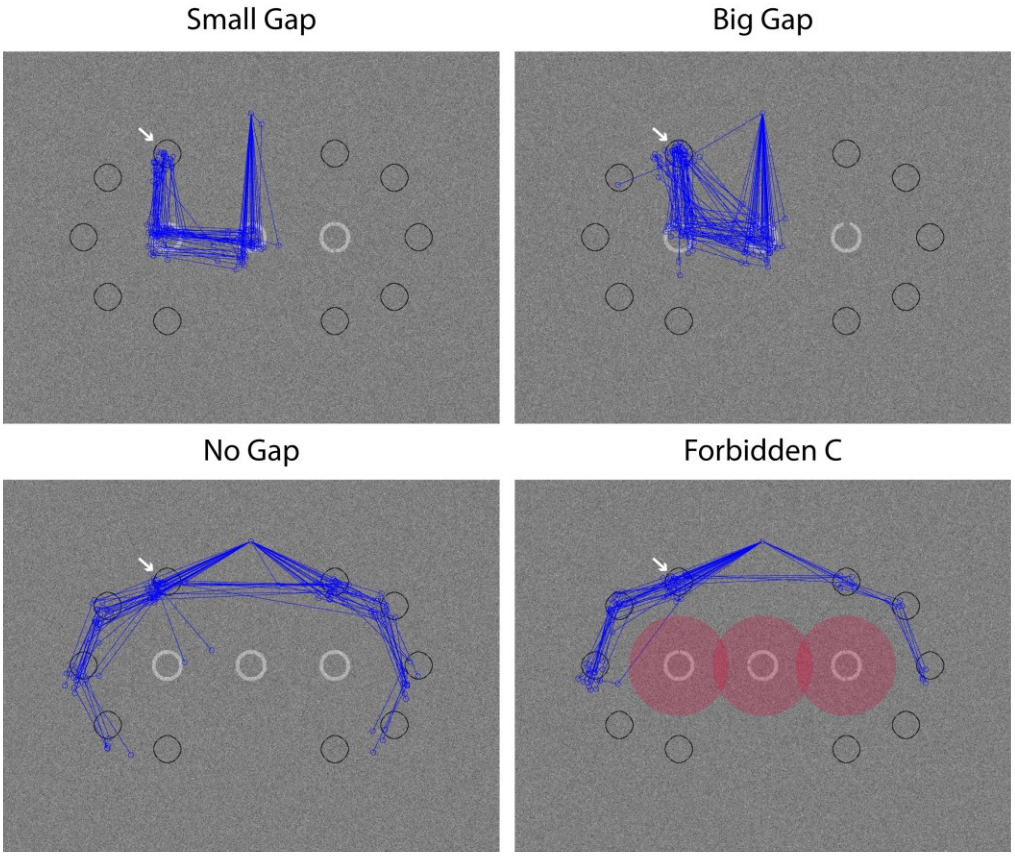

Figure 20. Eye movement selection in the remote cue task. Blue lines depict observer SE's scanpaths starting at the top fixation point for trials in which the target appeared at the location indicated by the white arrow. When the remote cues are informative and able to be foveated (Small and Big Gap), initial eye movements are directed towards them to help determine target location. When remote cue information is either unable to be foveated (Forbidden C) or unavailable altogether (No Gap), an eye movement strategy which simply surveys as many target locations as possible is adopted.

Insight into the origin of these performance differences can be gathered qualitatively by inspecting eye movement selection. Scanpaths for a single observer (SE) and single target location across all conditions can be seen in Figure 20. In both the Small Gap and Big Gap condition, early saccades were clearly directed away from potential target locations as eye movement selection favored the predictive remote cues to gather information about target location. Only after the remote cues were inspected were the eyes directed toward a target location to determine target presence. In stark contrast, eye movements in the No Gap condition almost completely ignored the remote cues since they provided no information about target location. Mirroring metrics of perceptual performance, eye movement selection in the Forbidden C condition closely mirrored that of the No Gap condition. Although the remote cues were predictive in this case, preventing foveation of the cues appeared to degrade their usefulness to the point that that eye movement selection devolved into the brute force "inspect as many locations as possible" strategy observed when no predictive information was available.

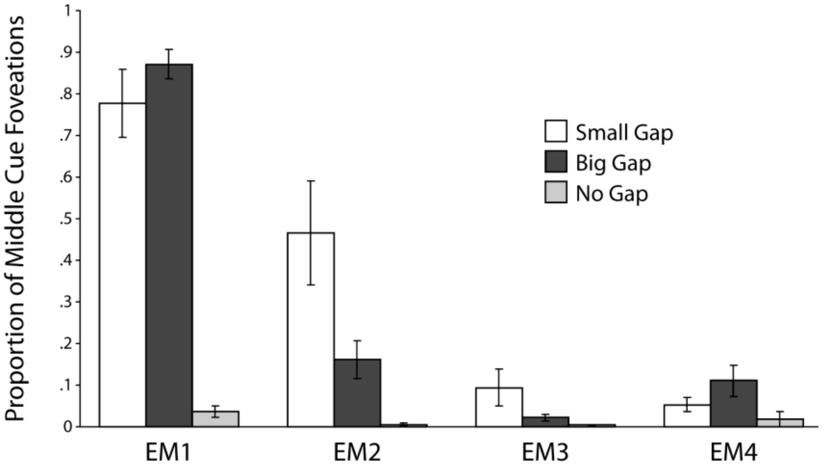

Figure 21. Proportion of eye movements to the central cue. When the central cue contains predictive information (Small and Big Gap), it is heavily fixated (eye position falls within 2° of cue) in the first and second eye movement. When it contains no information about target location, it is rarely, if ever, foveated (No Gap). Error bars represent ±SEM.

Quantitative measures of eye movement selection across participants supported the notion that observers sought out remote cue information via saccadic selection to determine target location and enhance perceptual performance. The mean proportion of saccades to the central remote cue (defined as a fixation landing within 2° of visual angle of the cue) as a function of eye movement is seen in Figure 21. A three (remote cue condition) by four (eye movement) repeated measures ANOVA ($F(2,4)=67.01$, $p=.001$) revealed that the central cue was fixated much more heavily in remote cue conditions where the cues contained predictive information (Small and Big gap) than when it did not (No Gap)[2]. Moreover, a significant interaction ($F(6,12)=26.64$, $p<.001$) showed that in the Small and Big Gap conditions, the central cue was heavily fixated on the first eye movement (80-90%), but this propensity to target the central cue dropped drastically in subsequent eye movements (down to 5-10% at the fourth eye movement), indicating that observers had moved to other elements of the display.

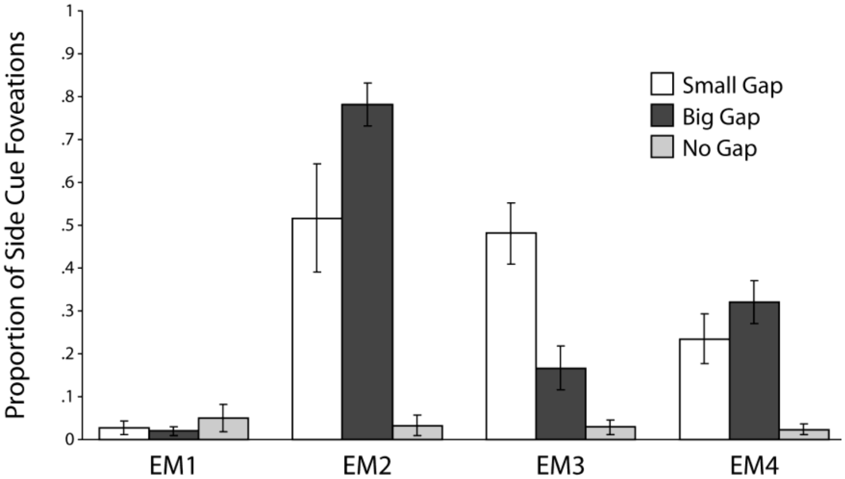

Figure 22. Proportion of eye movements to the correct side cue. When predictive information is available (Small and Big Gap), eye movements after the first are often directed towards the flanking cue on the side of the target. Note the discrepancy between the Small and Big Gap condition, as the peak of side cue-driven eye movements occurs later when the cue information is less visible (Small Gap). As with the central cue, the side cue is almost completely untargeted by eye movements when it contains no location information.

Analysis of the proportion of fixations directed towards the flanking remote cues (either side) provides further evidence compatible with a cue driven fixation selection

---

[2] Note that the Forbidden C condition was not included in this analysis as foveating any of the cues was not possible by design.

pattern when remote cues were predictive. The proportion of times the side remote cues were fixated as a function of eye movement is seen in Figure 22. The flanking remote cues were fixated far more often when they contained predictive information (Small and Big Gap) than when they did not (No Gap, *F*(2,4)=443.23, *p*<.001). A significant interaction (*F*(6,12)=16.39, *p*<.001) revealed that, while the side cues were almost never fixated in any task on the first eye movement, they were rather heavily fixated in the Small and Big Gap tasks on the 2nd and 3rd eye movements. Taken with analyses of the central cue fixations, it appears that observers used their initial eye movements to survey the central cue, then the corresponding side cue when remote cues provided predictive information.

The functional consequences of this early eye movement selection can be seen clearly when analyzing the cumulative probability that observers had foveated the target as a function of eye movement (Figure 23). While a significant main effect of eye movement (*F*(4,8)=237.5, *p*<.001) indicated that the probability of foveating the target increased markedly for all tasks with increasing number of eye movements, a significant cueing condition by eye movement interaction (*F*(12,24)=45.73, *p*<.001) revealed considerable cueing condition-dependent differences in the shape of this increase. When predictive information present (Small and Big Gap), the target location was almost never fixated within the first two eye movements. However, the proportion of target fixations then jumps markedly from the third eye movement on (reaching 85-95% by the fourth eye movement), indicating that observers utilized their early fixations to gather remote cue information, then move to the target location with considerable certainty. It is interesting to note that there were considerably less target foveations by the third eye movement of the Small Gap task compared to the Big Gap task (*t*(2)=-7.53,*p*=.008), potentially reflecting the differences in the visibility of the remote cue information (i.e. smaller gap in Small Gap condition).

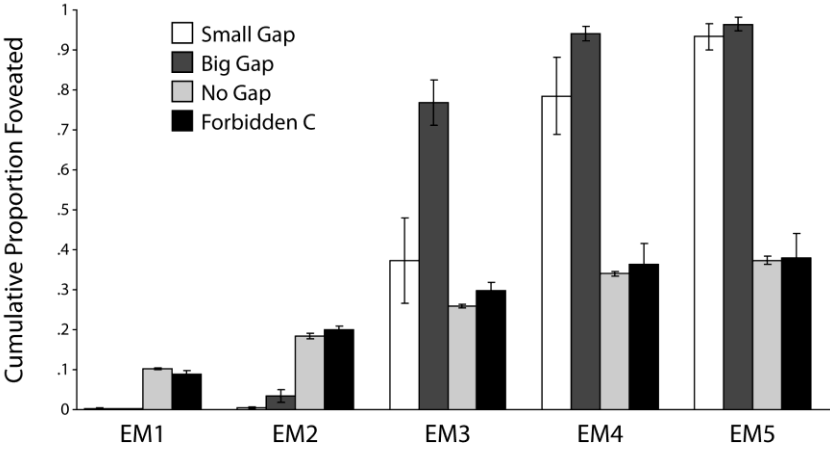

Figure 23. Cumulative probably of foveating the target by eye movement. When remote cue information is not available (No Gap) or unable to be foveated (Forbidden C), the probability of foveating the target on the first eye movement starts low (~10%) and slowly increases (~40% by EM5), indicating a lack of guidance in eye movement selection. In contrast, the probability of fixating the is nearly zero for the first two eye movements when remote cue information is available (Small and Big Gap), as early eye movements are directed to the cues. After the 2nd eye movement, the probability of fixating the target rapidly ostensibly as a result of exploiting previously gathered remote cue information. Error bars represent ±SEM.

In contrast, the cumulative proportion of target foveations when remote cue information was unavailable (No Gap) or unable to be fixated (Forbidden C) was well above zero (~10%) from the start, but increased rather slowly with increasing eye movements (to a maximum of roughly 40%). These results are consistent with an observer who is seemingly surveying conditions at random (i.e. probability of each eye movement landing on the target location is equal). Interestingly, the proportion of trials foveated for these two conditions exceeded that of the Small and Big Gap conditions for the first two eye movements (paired *t*, $p<.01$), reflecting the fact that early eye movements in conditions where remote cue information was available to be fixated were directed to the cues. As seen in analyses of perceptual performance and qualitative scanpath inspection, it appears that for the Forbidden C and No Gap conditions, observers essentially attempted to fixate as many locations as possible, but with very little guidance.

Taken together, our results indicate that observers are able to incorporate remote cues into their eye movement selection to enhance task performance. Clearly, however, the quality and availability of this predictive information has strong functional consequences on both oculomotor behavior and search performance.

C. Foveated MAP eye movement models

Here, we implement a foveated MAP eye movement model, which is nearly identical to that described in Chapter II. The foveated MAP provides a benchmark of expected performance for a model which can only move its eyes to potential target locations. In the case of the remote cue task, where targets cannot appear at cue locations, the MAP model cannot move its eyes to the cues.

As a result of this property, a MAP searcher may seem poorly suited for modeling the remote cue task. However, given the prevalence of "saccadic targeting" strategies in human data (Beutter et al., 2003; Eckstein et al., 2006; Findlay, 1997) and models of visual search (R. P. N. Rao et al., 2002; Zelinsky, 2008), a MAP model provides a.) a benchmark of performance and behavior for a searcher which adopts a saccadic targeting eye movement strategy, and b.) a point of comparison to quantitatively evaluate potential search benefits gained by human observers when they break from such a strategy. Moreover, MAP models have been shown to be nearly optimal when target visibility falls off rapidly with eccentricity (Zhang & Eckstein, 2010), which is the case with the high spatial frequency signal used in the remote cue task.

Three variants of a foveated MAP are outlined which differ in how much they use cue information and how eye movements are selected. Specifically, we describe 1.) a simple foveated MAP which ignores all cue information, 2.) a foveated MAP which peripherally gains information from the remote cues, and 3.) a foveated model in which the first 2 saccades are preprogrammed to go towards the remote cues. Evaluation of these three models provides insight into the origin of human performance benefits via remote cues as well as interesting comparisons with human eye movement selection.

1. Ignore cues (fMAP)

The basic form of the foveated MAP eye movement model is nearly identical to the model outlined Chapter II. The critical difference for the remote cue task, however, is the fact that the model needs to make a detection (i.e. present/absent) decision as opposed to an identification decision. As such, the decision rule for the foveated MAP of this detection task consists of taking the sum of likelihood ratios, weighted by the prior probability that the target will appear at each one of the target locations, and comparing that to a criterion. If the sum of weighted likelihood ratios exceeds this criterion, the model gives a target present decision; if not, the model gives a target absent decision. The weighted sum of likelihood ratios is given by:

$$wLR = \sum_{x}\sum_{y} priori_{T,x,y} LR_{k(T),x,y} \tag{14}$$

Where $priori_{T,x,y}$ is the product of the priors and all previous likelihoods. When selecting the first eye movement, this term will only be the priors:

$$priori_{T,x,y} = \pi_{x,y} \prod_{t=1}^{T-1} LR_{k(t),x,y} \tag{15}$$

It is also important to note that, since target location was randomly selected on each trial $\pi_{x,y}$=.1 for all target locations.

If the weighted sum of likelihood ratios exceeds a criterion *crit*, the model responds target present. Otherwise, it responds target absent:

if *wLR*>*crit*; "target present"; else, "target absent"

To select eye movements, the model moves its gaze to the target location of maximum weighted likelihood ratio (i.e. posterior probability) as described in Chapter II:

$$k(T+1) = \underset{x,y}{\arg\max}\, wLR_{k(T),x,y} \tag{16}$$

Note that these expressions are heavily simplified compared to the letter identification foveated MAP as the remote cue task only had one possible target of a set contrast as opposed to 25 possible targets in the letter identification task. Additionally, since human observers were presented with noise fields which refreshed every 100ms, the external (i.e. image) noise field updated with each eye movement, as well as the internal noise sample.

The implementation of eccentricity dependent internal noise to simulate the decrease in acuity in the peripheral retina was identical to that described in Chapter II, save for the fact that only a single visibility profile underlied the perceptual decision.

2. Peripheral cue use (Use Cues)

While a basic foveated MAP does not seek out non-target (in this case, remote cue) information, we augment the model to incorporate remote cue information, while maintaining a MAP eye movement strategy. In the remote cue tasks, the cues essentially serve as priors: if the observer knows which target location is indicated by the cues, they can base their perceptual decision solely off of information at that location. Here, we incorporate an additional foveated Bayesian Ideal Observer (fBIO) for the remote cues to modify the priors at each target location.

The fBIO consists of peripherally assessing the visual evidence at the remote cue locations and determining 1.) the likelihood that the central cue is pointing left or right, and 2.) the likelihood that the side cues are pointing in one of the five possible directions. The model handles both situations identically; the only difference being the number of possible remote cue orientations. The likelihood ($L_{k(T),i}$) of a single cue orientation (s_i) given the noisy evidence at the remote cue location (g) and current fixation position $k(T)$ is given by:

$$L_{k(T),i} = p(g \mid s_i) \qquad (17)$$

Likelihoods are calculated for both possible orientations for the central cue, as well as all five possible orientations for each side cue. The estimated prior probability that the target appears at any one location is then given by the product of the likelihood of each side cue orientation with the likelihood of the target appearing on that side (corresponding middle cue likelihood).

This modified prior is implemented into the foveated MAP in place of the uniform priors present when cues are ignored. The modified priors bias eye movement selection to locations likely to contain the target, and put more weight in the detection decision on those locations. However, the computational underpinnings of *how* eye movements and perceptual decisions are made does not change. The foveated MAP still cannot make eye movements to non-target locations. Instead, the foveated MAP now peripherally gathers information about remote cue information which can be implemented in target-driven eye movement selection.

3. Preplanned to cues (Preplan)

Both the basic foveated MAP and the version which peripherally gains remote cue information are unable to make eye movements toward remote cue locations. To provide a metric of performance in which remote cues are foveated, a third variant of the foveated MAP model was implemented in which the first two eye movements were essentially preselected.

The first eye movement of the preplanning model always went to the center of the central cue. After assessing the central cue, the model then moved to the side cue on the side which was deemed most likely based off the fBIO assessment (i.e. maximum likelihood) of the central cue. At the second fixation, the fBIO assessed the most likely orientation of the side cue and proceeded to move its eyes to the corresponding target location. When at the selected location, the model based its present/absent decision off of a single sample of information.

4. Calibrating the model to human visual system properties

Measurement of visibility maps

Visibility maps were measured for the target as well as the remote cues of both big and small gap size. For the target visibility map, all three observers from the original remote cue study performed a simplified detection task to measure performance as a function of retinal eccentricity. For the remote cue visibility maps, two of the observers from the original task (observer AF graduated and was not available to participate) performed a simplified alternative forced choice (AFC) experiment in which they were presented a single remote cue and asked to determine its orientation. There were variants for both the central remote cue (2AFC; left or right) and the side remote cue (5AFC), mirroring their possible orientations in the main experiment.

Observers performed 2600 trials of the target visibility map task. Trials in the target visibility map task began with observers fixating a fixation cross and pressing space to initialize the trial. After a randomized delay (500-1500ms), the stimulus image would appear for a duration matched to the mean first saccade latency for each observer. Within a single bounding circle at the center of a noisy image a target appeared on 50% of the trials (i.e. 50% target present). Following the stimulus image, observers made a yes/no response and were given feedback on their selection. The fixation cross could appear at one of thirteen locations ranging from 0-7.4º of visual angle. The location of the fixation cross was constant within each 100 trial session, and changed every two sessions.

Observers performed ten 100 trial sessions of the middle remote cue visibility map task for both small and large gap sizes. The structure and timing of the task were nearly identical to that of the target task. Fixation could begin at one of five locations ranging from 0-7.4º from the central cue location. A single remote cue, oriented to the left or right (as in the main study) appeared at the center of the stimulus image. Observers were to respond whether they believed the opening in the remote cue faced the left or the right and were given feedback on their selection.

Finally, observers performed fourteen 100 trial sessions of the side remote cue visibility map task for both small and large gap sizes. Fixation was set at seven locations ranging from 0-8.9º from the target location. The stimulus image consisted of a single side remote cue at the same spatial position that the leftmost cue resided in the main experiment. The orientation of the cue was randomly selected from the five orientations possible in the main experiment on every trial. Following stimulus presentation, observers were shown a response screen with the five possible orientations and clicked on the one they believed to be presented and were given binary feedback (correct/incorrect) on their response.

Visibility map fitting

Internal noise parameters were estimated for each observer. The form of the eccentricity dependent internal noise as well as the visibility map fitting procedure was identical to that described in Chapter II except there was a single set of internal noise parameters per observer.

D. Human-Model Comparisons

Two hundred thousand simulated trials were run for each variant of the model. Each trial consisted of four simulated eye movements, matching the mean number of eye movements made by human observers in the remote cue task ($M=3.98$).

The most basic comparison to make is that of the basic foveated MAP and human performance in the remote cue task. The fMAP has no knowledge of the cues, so its predicted performance does not change for any of the remote cue conditions. Additionally, the fMAP serves as a baseline against which to assess human performance, as it performs the task without cues in a near optimal fashion. As seen in Figure 24, predicted performance of the fMAP is nearly indistinguishable from human performance in terms of PC in both the No Gap and Forbidden C conditions (single sample t, $p>.05$). A consistent set of results can be seen in the assessment of d', although predicted foveated MAP d' were slightly higher than human d' in the No Gap condition ($t(2)=-5.56, p=.03$). Thus, when remote cue information was not available (No Gap) or unable to be foveated

(Forbidden C), the fMAP, which has no knowledge of the remote cues, corresponds quite well to human performance. Critically, predicted foveated MAP performance, in terms of both PC and d', fell markedly short of human performance in both conditions where remote cue information was available and able to be foveated (Small Gap and Big Gap; single sample *t*, all *p*<.001). Clearly, observer performance when remote cue information was available cannot be explained by a model observer which ignores cue information.

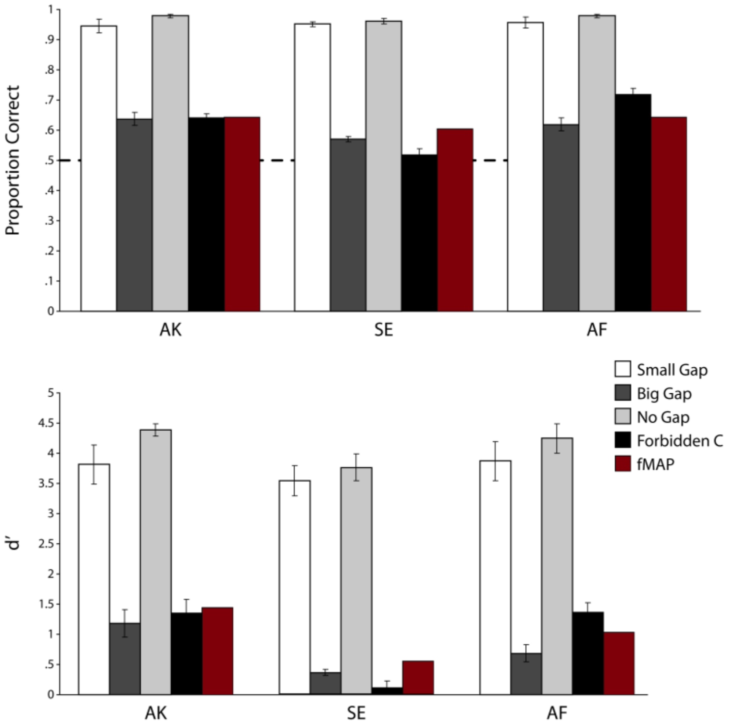

Figure 24. Comparison of the foveated MAP (fMAP) to human performance in the remote cue task. fMAP predictions correspond well with observer data when cue information is unavailable or unable to be foveated (No Gap and Forbidden C). But when cue information is readily accessible, fMAP performance falls far short of observer proficiency. Error bars represent ±SEM.

Assessment of the predicted eye movement patterns of the fMAP provides further insight into potential mechanisms underlying observer performance in the remote cue task. Figure 25 shows representative observer scanpaths for a target appearing in the location indicated by the white arrow in the No Gap task alongside the eye movement predictions of the foveated MAP eye movement model. Although there are salient differences (namely the eye movements which cross from one side to another) which reflect the fact that the MAP model doesn't incur a cost for making longer eye movements, these long range eye movements only occur on a small minority of trials (<10%). It is also worth noting that neither the fMAP (*M*=50.03%) or human observers in the No Gap (*M*=50.37%) or Forbidden C (*M*=54.91%) showed significantly above chance guidance towards the target (i.e. moved their eyes to the correct side), reinforcing the similarity of the predictions with human data when cue information is unavailable or unable to be foveated.

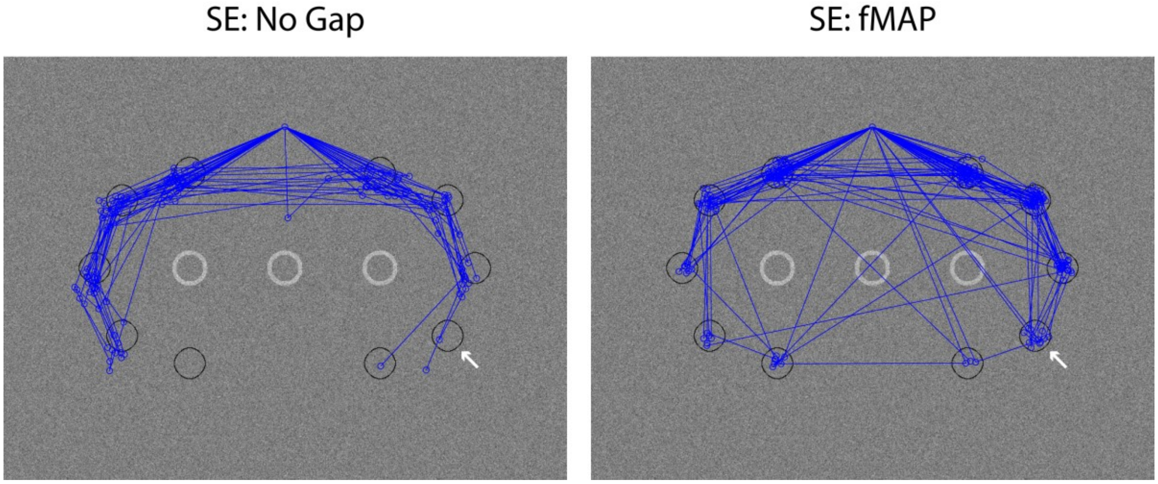

Figure 25. Comparison of human and foveated MAP (fMAP) eye movement patterns. When cue information is not available to human observers (No Gap condition), predicted fMAP (a model which has no knowledge of cue information) eye movements are remarkably similar to humans.

Further similarities between fMAP eye movement predictions and human saccadic selection when remote cue information is unavailable (No Gap) or unable to be foveated (Forbidden C) can be seen in Figure 26, which shows the cumulative proportion of trials in which the target was foveated as a function of eye movement. Like human observers, the proportion of trials where the target was foveated steadily and slowly increased across eye movement for the fMAP. Critically, there were no significant differences between human observers in the No Gap and Forbidden C tasks and the fMAP in terms of the proportion of trials foveated across eye movements (*F*(2,4)=.814, *p*=.505), reinforcing the idea that the fMAP may capture elements of human perceptual performance and eye movement behavior when remote cues were absent or unable to be foveated.

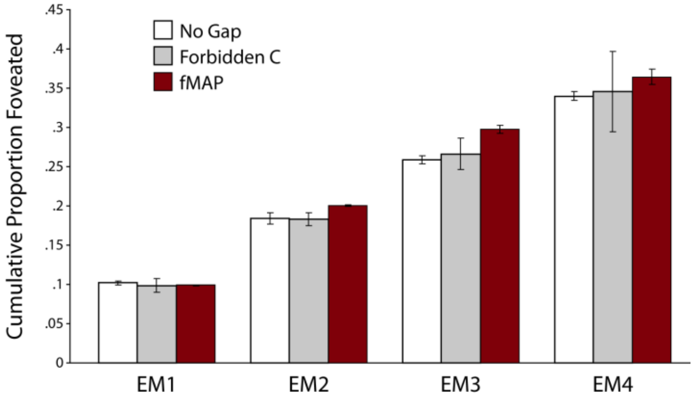

Figure 26. Cumulative proportion of trials foveated for human observers and the fMAP as a function of eye movement. When remote cue information is unavailable or unable to be foveated, human eye movement behavior closely mirrors that of a model which has no knowledge of the cues (fMAP). Error bars represent ±SEM.

It seems clear that the fMAP, which has no access to remote cue information, is a relatively good descriptor of human performance when remote cue information is absent or degraded. However, the question remains whether the enhanced human performance when remote cue information is readily available (Small and Big Gap conditions) is a result of the cue-driven eye movement strategy or is achievable simply through efficient use of remote cue information.

Figure 27 shows observer performance in the remote cue task as well as predictions for a foveated MAP model which uses peripheral remote cue information for both gap sizes (Use Cues-Big and Small) as well as a foveated model which makes two preplanned eye movements to the remote cues (Preplan-Big and Small). Models which accrue peripheral information about remote cues (Use Cues) show modest gains in performance in comparison to human No Gap and Forbidden C performance. However, these models fall well short of human performance in conditions where remote cue information is available and free to be fixated. In contrast, the set of models in which the first two saccades were preplanned to go towards remote cues (Preplan) nearly matched human levels of proficiency when cue information was available and able to be foveated, especially in terms of PC. Even though the Preplan models essentially made their decision off of one sample of visual information about the target locations (after surveying the remote cues to determine likely target location), they come the closest to human performance profiles. Taken together with results from the models which peripherally used remote cue information, it appears that the cue-driven human eye movement strategy when remote cue information was available was instrumental in contributing to the increased levels of performance in those conditions.

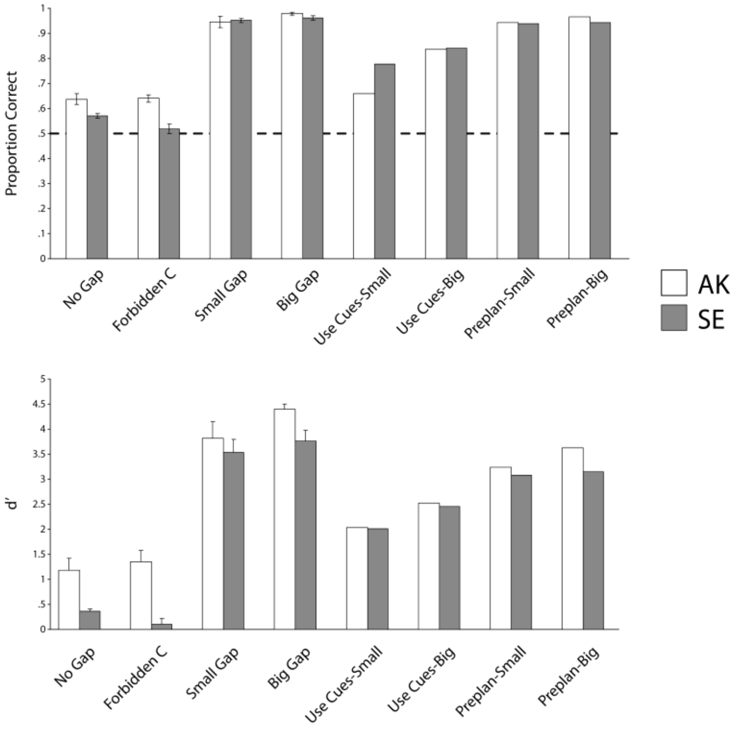

Figure 27. Predicted performance for all models. Models which peripherally use cues (Use Cues-Small and Big) enjoy modest gains in performance in comparison to human data. Models that preferentially target the remote cues with the first two saccades (Preplan-Small and Big) much better approximate human performance when cue information is available. Error bars represent ±SEM.

E. Learning remote cues

In the main remote cue experiment, we demonstrated that observers are able to adopt a remote cue-driven eye movement strategy when doing so enhanced task performance. Moreover, implementation of multiple variants of a foveated MAP eye movement model suggested that this eye movement strategy was instrumental to enhancing task performance. However, observers were given explicit information about the predictive nature of the cues, so it was simply on the observers to leverage this information. In real life search, predictive elements in the visual environment are often obviously labeled and must be learned through visual experience. To assess observer flexibility in learning predictive information, we ran a modified version of the Small Gap condition in which observers were given no information about the remote cues. The readiness with which observers adopt an atypical eye movement strategy in the absence of instruction or explicit information about potentially predictive cues lends insight into preferred eye movement strategies and willingness to explore alternative search solutions.

1. Method

Observers. Observers were nine undergraduate students (five female; ages 19-23) at the University of California-Santa Barbara, none of which had participated in Experiment 1. All observers had normal or corrected to normal vision.

Stimuli. Stimuli were identical to those in the Small Gap condition in the main experiment, except the gap in the remote cues was made slightly larger (.19°). This small increase in gap size was chosen so that the necessity to foveate the remote cues to gather target location information was maintained while also increasing the accessibility of this information if foveated.

Task. The timing and structure of the task were identical to that described for the Small Gap condition the main experiment. However, observers were not given any information about the nature of the remote cues. If an observer inquired about the cues, they were told that the cues were simply "an element of the display." Observers completed twenty 100 trial sessions of the task.

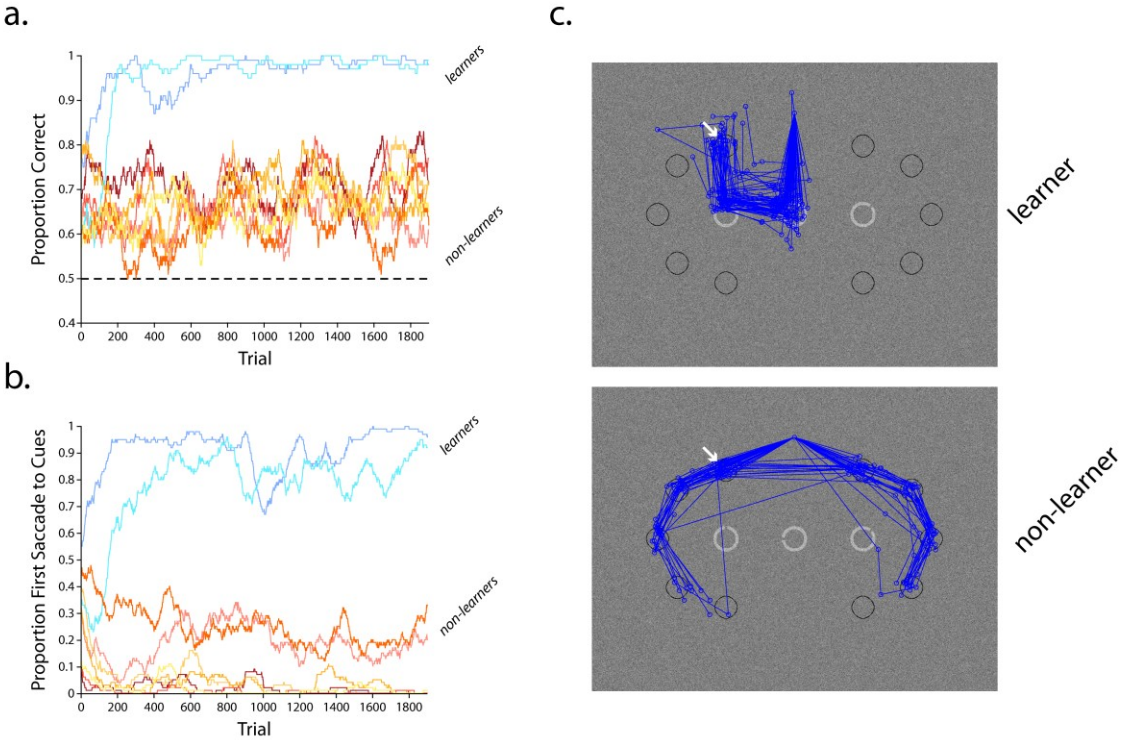

Figure 28. Perceptual performance and eye movement selection of learners and non-learners in the remote cue task. Panel (a) shows perceptual performance as a function of starting trial on a sliding 100 trial window. Learners rapidly learn remote cue contingencies and performance is markedly enhanced. Panel (b) shows the proportion of first saccades to the remote cues on an identical sliding window. Increases in performance from learners clearly coincides with a more cue-driven pattern initial eye movement selection. This interpretation is strengthened by assessment of example learner and non-learner scanpaths shown in panel (c), where initial eye movements for learners clearly target remote cues. White arrow indicates target location for those trials.

2. Results and discussion

To track performance across time, percent correct and sensitivity (d') were calculated on a 100 trial sliding window. While the majority of observers failed to measurably improve their performance over the course of the experiment (non-learners), two observers were able to markedly and rapidly learn the remote cues (learners; see Figure 28a). The learners learned the remote cue contingencies early and reached nearly ceiling performance within the first 200-300 trials, and this enhanced performance in comparison to non-learners remained throughout the experiment ($t(7)=11.40$, $p=4.4\text{e-}06$). In contrast, performance of non-learners stayed relatively constant (although significantly above chance performance; 60-70%) throughout the experiment.

Insight into the origin of the performance difference between learners and non-learners can be seen in their initial eye movement patterns. The proportion of first saccades within 2° of the remote cues was calculated on a 100 trial sliding window (Figure 28b). Clearly, the enhanced performance of learners strongly coincided with more remote cue-driven eye movements: the vast majority of learners' initial eye movements were directed to the remote cues, while very few of the non-learners' initial eye movements were cue driven. In fact, for the majority of eye movements, less than 10% of initial eye movements were directed to remote cue locations. This interpretation is strengthened by considering the example eye scanpaths from a learner and non-learner in Figure 28c. Initial learner eye movements are clearly directed towards remote cues to gather target location information, while those of the example non-learner almost completely ignore the remote cue region.

To assure that the differences between learners and non-learners was not due to non-learners inability to effectively utilize remote cues even with knowledge of their meaning, four of the non-learners were brought back to perform a brief follow-up experiment. The four non-learners were explicitly told the predictive nature of the remote cues and re-performed five 100 trial sessions from the original learning study.

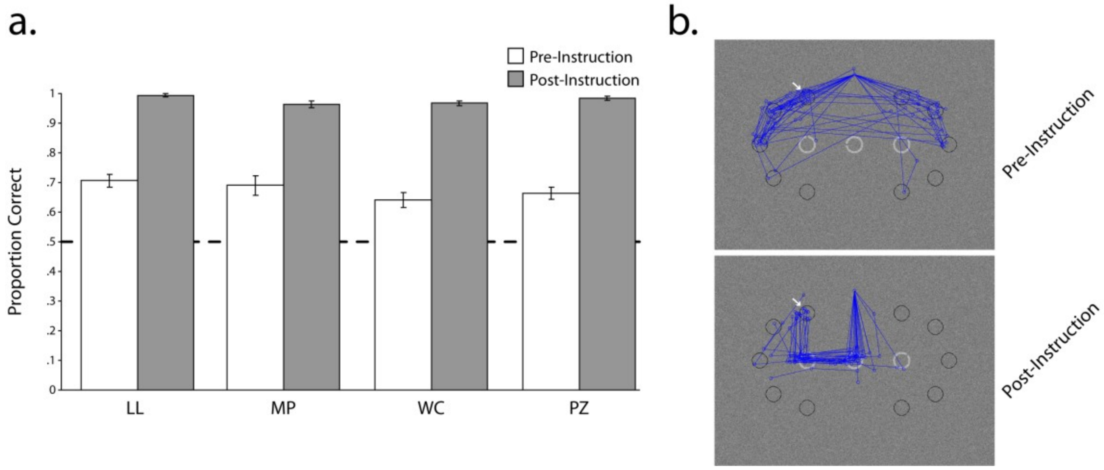

Figure 29. Performance and eye movements for recalled non-learners. Performance drastically improved for non-learners when informed of remote cue contingencies, as seen in panel (a). These post-instruction enhancements in performance were reliably associated with considerable changes in eye movement behavior (b), with initial post-instruction eye movements being directed towards the predictive remote cues.

Post-instruction performance for all four observers improved drastically (30-35%; $t(3)=23.55$, $p=8.38\text{e-}5$), reaching almost ceiling (see Figure 29a). Moreover, inspection of observer scanpaths pre- and post-instruction reveals a marked disparity (see Figure 29b), with initial eye movements in post-instruction trials clearly targeting remote cue information to gather task-relevant predictive information; a strategy which did not appear prior to instruction. Thus, it appears that non-learners simply were not able to learn the remote cue contingencies, not that they were incapable of utilizing remote cues in a more general sense.

E. Conclusions

Moving the eyes to likely target locations, a so-called saccadic targeting strategy, is a common and effective way to enhance visual search efficiency. However, there are occasionally cases when information predictive of search target location appears at regions of the visual environment considerably spatially removed from the search target.

Here, with a novel visual search task employing an array of predictive cues spatially separated from potential target locations (which we deem remote cues), we demonstrate that human observers are capable of adopting highly atypical eye movement strategies, which directly conflict with saccadic targeting patterns of oculomotor behavior, when doing so can lead to enhanced perceptual performance. A collection of foveated eye movement models, which differed in their saccadic planning and use of remote cues, provide converging evidence that human eye movement patterns in the remote cue task were reflective of attempts to maximize search accuracy. A learning version of the remote cue task provided valuable evidence that, although human observers may be able to adopt uncommon oculomotor strategies when the benefits of doing so are apparent, they do not necessarily look for opportunities to implement such behaviors when the subsequent benefits are not explicitly laid out.

In all, human observers demonstrate remarkable flexibility to optimize eye movement strategies to maximize search performance.

IV. Predictive cues can hinder multi-fixation visual search

Predictive cues nearly invariably improve perceptual performance on a variety of tasks, including visual search. In Chapter II, we showed that human observers were able to employ predictive cue information to guide search and enhance multi-fixation search performance (albeit less than predicted by a foveated MAP eye movement model). However, when cues were present in our letter identification task, they fully predicted target location (i.e. the target could not appear outside of the cues). The question remains, then, whether human observers are able to effectively utilize spatial cues during multi-fixation search when the cues only partially predict target location; conditions more akin to real-life and vocational search.

We return to the letter identification search task and present a modified version, in which the cues only partially (80%) predict target location when present. The modified image statistics are also applied to cues absent sessions. Surprisingly, with this increased spatial uncertainty, we show that cue presence does not lead to enhanced search performance overall. In fact, while cue presence does aid search at our lowest signal contrast, the presence of predictive cues actually hinders identification performance at high signal contrasts. Analysis of observer eye movements shows that this reversal in performance at high signal contrasts is associated with a pattern of eye movements which appeared to heavily favor the cues, impeding target foveated. A control task, in which the cues were not at all predictive of target location, provides evidence that our results are not merely the reflection of exogenous attentional capture via the cues.

The foveated MAP eye movement model, developed in Chapter II, is applied to the modified search task. Model predictions reveal that, even in the face of massive spatial uncertainty, a near-optimal observer would never predict this reversal in performance, indicating substantial suboptimalities in human cue usage and search performance. Moreover, human-model comparisons reveal that cue overuse likely was not the origin of the reversal effect, and the reversal in performance may be related to poor cue use in a more general sense.

It is clear that the mere presence of predictive spatial cues is not enough to improve search performance, particularly in more naturalistic situations where there is substantial uncertainty about signal strength and target location. With an increasing number of vocational and life-critical search tasks using synthetic cues, a deep understanding when cues help (and hurt) search is critical to optimizing and improving human-machine interactions.

A. Introduction

In natural visual search, scene regions or objects predictive of the location of the target generally only denote *likely* target locations, and not certain ones. This considerable uncertainty about target location even in the presence of predictive elements of the visual environment also extends to common vocational search situations in which synthetic cues are overlaid on task relevant images or visual displays (e.g. medical images: Gur & Sumkin, 2006; V. M. Rao et al., 2010; satellite and aerial images: Willhauck, Schneider, De Kok, & Ammer, 2000; air traffic control: Metzger & Parasuraman, 2006; Wickens, Conejo, & Gempler, 1999). In fact, nearly 75% of all breast mammograms are now read with computer aided diagnosis (CAD; V. M. Rao et al., 2010), in which cues are inserted directly onto the mammogram image to indicate likely pathologies. Critically, there is considerable uncertainty within the medical community about whether or not these synthetic cues even aid diagnostic performance (Baker et al., 2003; Chan et al., 1990; Ciatto et al., 2003; Drew, Cunningham, & Wolfe, 2012). In these important and often life-critical tasks, understanding the effects of partially predictive cues on human performance is necessary to optimize displays for human observers and maximize search accuracy and efficiency.

We return to the letter identification search task described in Chapter II and modify it so that the cues are only partially (80%) predictive of target location overall; conditions more akin to real-life and vocational search. Surprisingly, we find that cues present performance does not exceed that of cues absent overall, and that performance when cues are present is actually lower than when they are absent at high signal contrasts. It is hypothesized that this unexpected reversal is associated with patterns of eye movement selection which may overuse the cues when the target is outside of cued locations and of high visibility. Implementation of the foveated MAP model demonstrates that such an effect is *not* predicted by a rational observer, and provides evidence that human observers may be using cues suboptimally not just when the target appears outside the cues, but in a more general sense.

B. Main experiment

1. Method

Six observers (five female), four of which participated in the experiment described in Chapter II, participated in the modified letter identification search task. In the modified task, the four cued regions contained the target 80% of the time in all, with individual probabilities of 40, 20, 10, and 10% and a randomized color-probability pairing for each observer. In the 20% of trials in which the target letter did not appear in one of the high probability regions, it could appear anywhere in the image outside of those regions. The statistics of the images in the cues absent sessions were identical to those of the cues present sessions except for the absence of cue circles (see Figure 30a). The timing and structure of the task and stimuli were otherwise identical to the experiment described in Chapter II. Observers completed ten 100 trial sessions of the cues present condition and eight 100 trial sessions of the no cues condition.

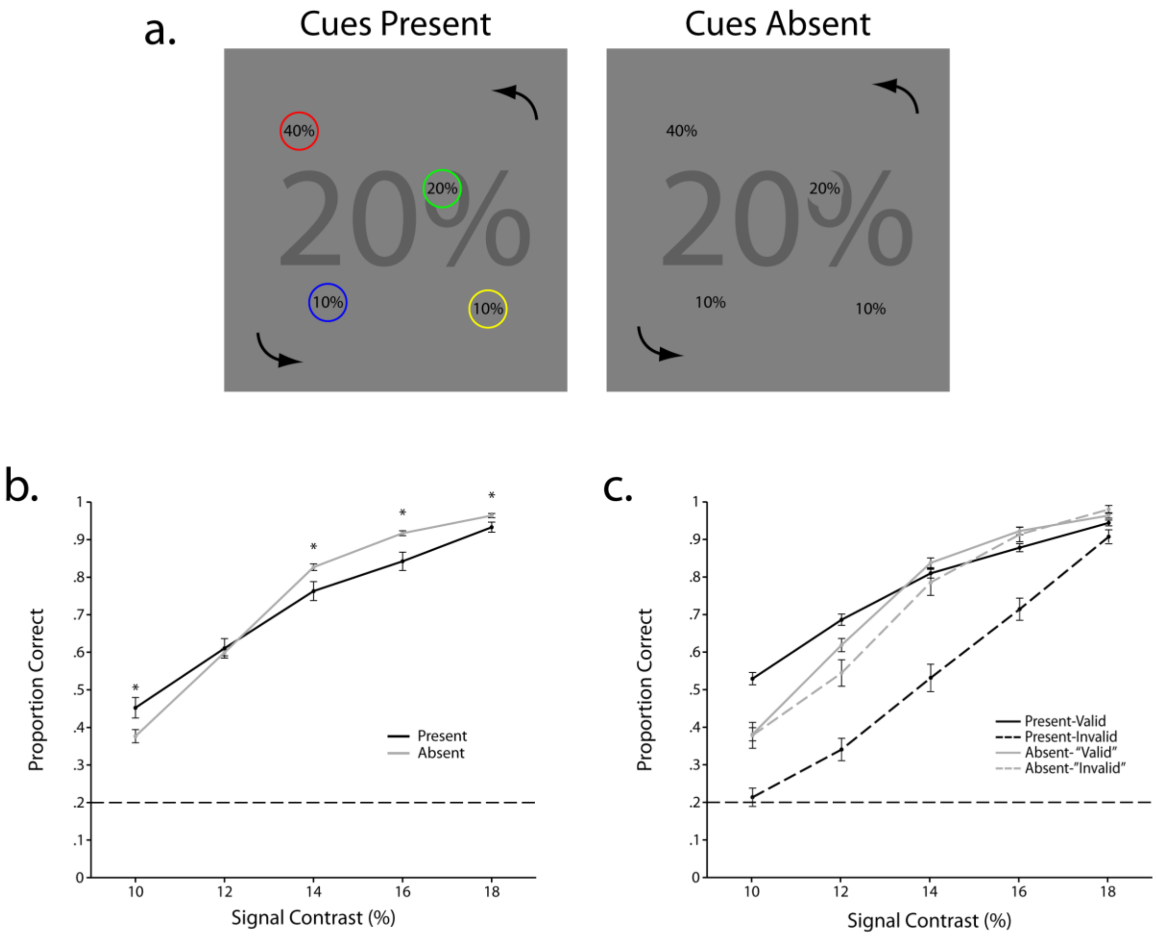

Figure 30. Modified letter identification search task and identification performance. Panel (a) shows the modified probabilistic structure of target location. Note that in 20% of the trials, the target appeared outside the cue circles when cues were present. Panel (b) shows identification performance as a function of signal contrast. At the three highest signal contrasts, task performance is actually superior when cues are absent, although performance is enhanced when the cues are present at the lowest signal contrast. As seen in panel (c), this reversal at high signal contrasts appears to be driven primarily by comparatively poor performance for trials in which cues were present and the target appeared outside the cued regions. Dotted line indicates chance performance. Error bars represent ±SEM. * denotes $p<.05$.

2. Results

Perceptual Performance. Figure 30b shows identification performance across all observers as a function of signal contrast for. A two (cueing condition) by five (signal contrast) repeated measures ANOVA revealed that, despite the fact that observers were given information that was strongly predictive of target location in the cues present condition, the availability of this information did not lead to enhanced overall task performance in comparison to the cues absent condition (*F(1,5)*=.75, *p*=.43). A significant interaction (*F(4,20)*=16.02, *p*<.001), however, indicated that there were performance differences between the two conditions across individual contrast levels. Paired samples t-tests revealed that while performance in cues present trials was superior at the lowest contrast tested (paired t, *p*<.05), at the three highest signal contrasts, observers performed the task better when the cue circles were absent from the image (paired t, *p*<.05). Cue presence actually impeded task performance as the target became more detectable.

To evaluate the origin of the reversal in performance, trials for each signal contrast condition were separated by whether the target appeared within one of the four high probability regions (valid cue; locations inside cue circles in cued sessions or corresponding regions in uncued sessions) or outside those regions (invalid cue; see Figure 31a). The statistics of the images were identical across conditions with the only difference being the presence of the cues. As seen in Figure 30c, a significant cueing condition by target location (valid/invalid) interaction (*F(1,5)*=38.15, *p*=.002) shows that that the inferior performance in cues present sessions at high signal contrasts appears to be mostly attributable to consistently poor performance in the cues present trials for which the target appeared outside of the cue circles (cues present invalid, dashed black line). That is, even when the target was easily detectable and identifiable in the periphery (as evidenced by the high performance in analogous cues absent invalid trials (dashed grey line) across the same contrasts), observers often failed to correctly identify the target. In contrast, performance for trials in which the target appeared within the spatially predictive cues (cues present valid, solid black line) was superior to that of analogous cues absent valid trials (solid grey line) at the two lowest signal contrasts (paired t, $p<.05$). Although performance for the inside trials converged at the three highest signal contrasts, there were no statistically significant reversals in performance over these contrasts.[3]

Note that at low signal contrasts, the detrimental effect of cues on performance for the 20% of trials in which the target appeared outside the cues is compensated by the performance benefits for the 80% of trials in which the target appeared inside the cues, resulting in a net positive effect of cues on performance across all trials. However, at higher contrasts, the cues provided little to no benefit to identification performance for the cued inside trials (as the target is foveated even in the absence of the cues), while the cues hindered performance in cued outside trials. Thus, there is a net detrimental effect of cues at the high signal contrasts.

Eye Movements. To assess whether the cue presence changed the observers' oculomotor strategies, we calculated the percentage of trials in which the target was foveated at any point during the stimulus presentation as well as the fixation durations during target foveations. Given the relatively small size of the target, foveating it was critical to achieving accurate identification performance. For our purposes, a foveation was defined as any eye movement which landed within 2º of the center of the target (although the general pattern of results which follows persisted across a range of radii of analysis from 1 to 3º).

---

[3] The data show a small possible advantage of the uncued condition vs. the cued condition at two highest SNR conditions which did not reach significance. If this effect reached significance, it could possibly be explained in terms of small spatial masking or crowding effects exerted by the cues on the letter targets.

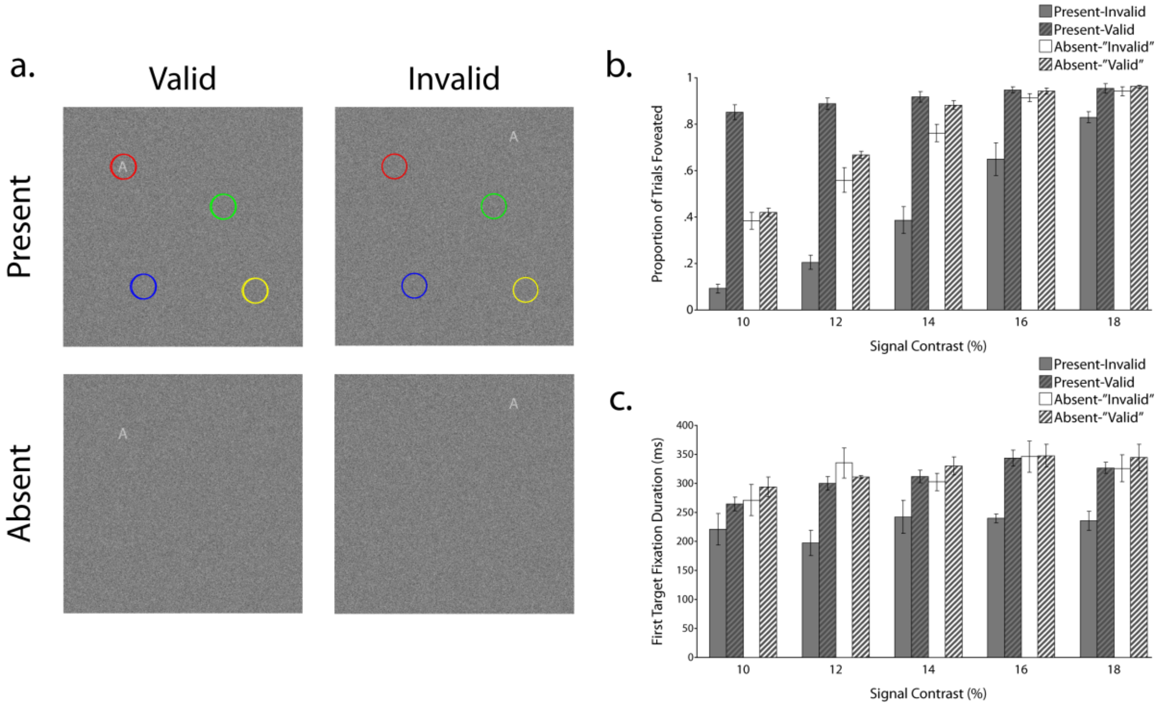

Figure 31. A visualization of analogous inside and outside trials for both cues present and absent sessions can be seen in (a). Panel (b) depicts the proportion of trials in which the target was foveated (fixation falling within 2º). Of particular note is the comparatively high proportion of trials in which the target was foveated for cues present-valid trials even at the lowest signal contrasts, indicating observers employed the cued regions to guide search. Conversely, the proportion of trials in which the target is foveated drops precipitously (and well below either uncued condition) when the target appears outside the circles, even at high signal contrasts when the target was easily detectable. Part (c) shows mean fixation durations for the first target foveation. No differences between cues present-valid or either cues absent condition are observed. Critically, however, target foveation durations are significantly shorter for cues present-invalid trials. Thus, not only are targets in cues present-invalid trials foveated less often, but when they are, less time is spent inspecting them. Error bars represent ±SEM.

Figure 31b shows that observers' saccadic selection in cues present sessions appeared to be dictated by the presence of the cues. A significant cueing by target location interaction (*F(1,5)*=228.35, *p*<.001) indicated that in cues present sessions in which the cues accurately predicted target location (cues present valid; see Figure 31a for visualization of conditions), observers nearly always fixated the target regardless of signal contrast (nearly 85% of trials even at the lowest signal contrast), indicating that the cues provided efficient landmarks to guide visual search. In contrast, in analogous cues absent valid sessions, the probability of fixating the target was quite low at low signal contrasts (~

40%), although the proportion of trials in which the target was fixated increased with increasing signal contrast until it eventually converged with the cued inside data at the highest signal contrast. Also, there was not a substantial difference between the proportion of trials in which the target was foveated in cues absent valid and invalid trials (paired t, *p*>.05), suggesting that the slight increase in average starting target eccentricity in outside trials (.7º) did not significantly impact performance. Critically, the trials in which the target was foveated the least across all signal contrasts were those in the cues present invalid condition (paired t, *p*<.05), where the cues did not accurately predict the target location. In this condition, the target was rarely foveated at the lowest signal contrast (~10%) and continued to be foveated less than the other three conditions, even at the highest signal contrast (83% vs ~95%), mirroring the consistently lower behavioral performance seen in cues present invalid trials.

Inspection of eye movement distributions provides a potential explanation for these foveation patterns. Figure 32a shows eye movement distributions for cues present-invalid trials as well as analogous cues absent-invalid trials (which is possible due to the matched statistics of the images. In invalid trials, the target can appear anywhere outside the high probability regions, so eye movements should be dispersed throughout the image if they are directed towards the target. We see that this is the case in cues absent-invalid trials. However, when cues are present, we see a large proportion of eye movements being directed towards the cues, particularly when signal contrast is low. Critically, this pattern perists (although not as strongly) for the highest signal contrast in cues present trials. The difference between cues present- and absent-invalid distributions is seen in Figure 32b.

When cues are present, more eye movements are still being directed to cued regions by the third eye movement, even though the target is quite visible and not at a cued location.

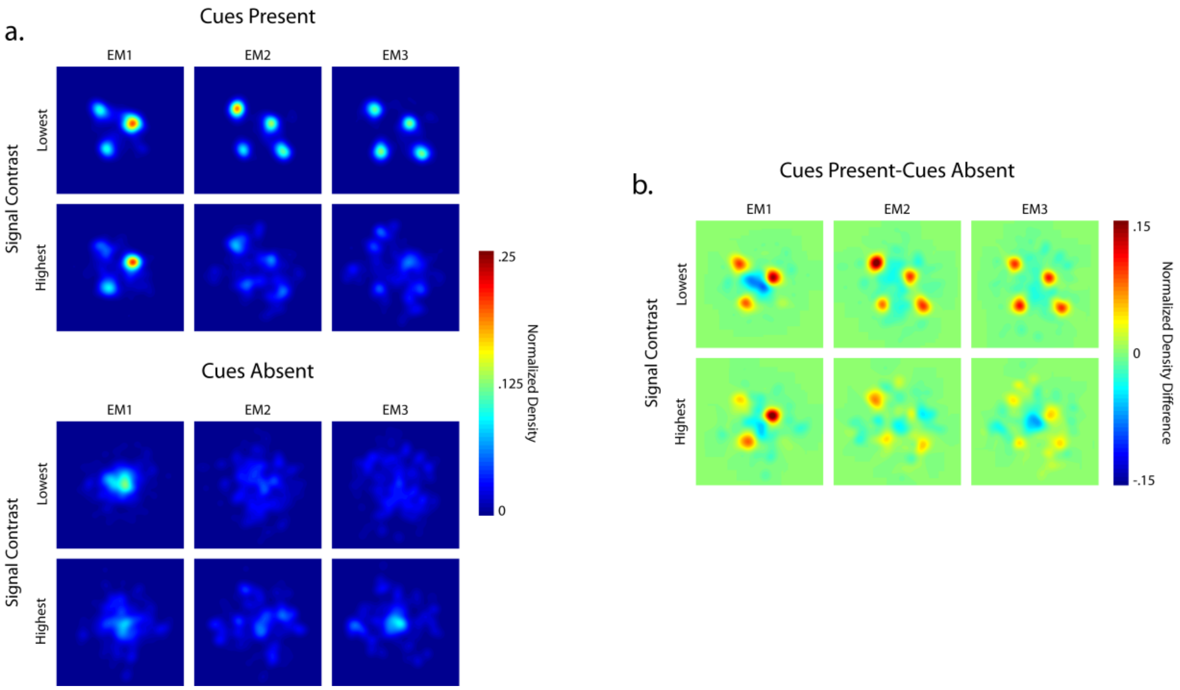

Figure 32. Normalized eye movement distributions for cues present-invalid trials and analogous cues absent-invalid trials. When cues are present, observers direct their eyes towards cued regions, even when the target is highly visible and appears outside these high probability locations. Panel (b) shows the difference maps between cues present and cues absent, highlighting this cue-driven strategy in invalidly cued trials when cues are present. Note that the target can appear anywhere outside the high probability regions, so there is no set target location.

A complementary picture can be seen in the analysis of fixation durations, which have historically been used as a measure of the quality of processing of visual information at the point of fixation (Guo, Mahmoodi, Robertson, & Young, 2006; Moffitt, 1980). Figure 31c shows the mean fixation durations for the first target foveations. A significant cueing condition by target location interaction ($F(1,5)=20.51$, $p<.006$) revealed that while fixation durations for cues absent valid, cues absent invalid, and cues present valid trials did not differ significantly from each other (paired t, $p>.05$), first target foveation durations were significantly (15-30%) shorter across all contrasts in cues present invalid trials than any other condition (paired t; $p<.05$). Thus, in cues present valid trials, where consistently poor performance led to the unexpected reversal in performance at high signal contrasts, not only did observers foveate the target far less often, but when they did, they spent considerably less time assessing it.

3. Discussion

Here, we demonstrated a case in which the presence of an array of predictive synthetic cues actually hindered search for targets with large spatial uncertainty at high signal contrasts. Moreover, this result appeared to be attributable to an overuse of cue information in eye movement selection, reducing observers' ability to foveate the target and decreasing the amount of information gleaned from target locations when they appeared outside the spatially predictive cues. Despite our previous demonstration that human observers were capable of improving search performance by utilizing spatial cues, it is clear that their mere presence is not enough to guarantee this enhancement.

Observers' cue-driven eye movement selection in the presence of cues, however, raises some critical issues about why such eye movements would be made. It was proposed that observers were simply overusing the cues: a suboptimal application of a rational solution. However, perhaps cue-driven eye movements were simply the reflection of exogenous capture from the peripherally presented cues. If this were the case, eyes should be driven heavily to the cues even when they hold no meaning, and performance should suffer in their presence. To test this alternative explanation to our data, a control study was run in which the cues were not at all predictive of the target location.

C. Control experiment

Non-predictive cues have been shown to hijack attentional focus and impede task performance (Farah, Wong, Monheit, & Morrow, 1989; Fecteau & Munoz, 2006; Hietanen, Nummenmaa, Nyman, Parkkola, & Hämäläinen, 2006; Hommel, Pratt, Colzato, & Godijn, 2001). Although our cues were partially (80%) predictive of target location, the decrease in performance when cues were present may be related to exogenous attentional capture by the cues. To test this, a third variant of the letter identification task, in which the cues did not provide any predictive information about target location, was evaluated.

1. Method

Four observers (2 female), none of which had participated in either of the previous letter identification search experiments, completed the control experiment. In the control experiment, the likelihood of the target appearing at any location across the image was equal (as long as the target did not overlap with the edge of the image) in both cues present and cues absent sessions. In cues present sessions, cue circles were present in the same spatial configuration utilized in the previously described letter identification tasks, but they did not hold any predictive value (i.e. targets were no more likely to appear within the circles than anywhere else in the image). Observers were informed explicitly that the circles did not hold any predictive value. The timing and structure of the trials, as well as all other stimulus properties, remained identical to those described above. Observers performed ten 100 trial sessions of each the cues present and cues absent conditions.

2. Results and discussion

Identification performance as a function of signal contrast in shown in Figure 33a. Over all signal contrasts, there was a small (~2-3%) but significant decrease in performance when cues were present (*t*(3)=5.31,*p*=.013). However, despite this slight overall impairment, there were not any significant differences in performance at individual contrast levels (paired t, all *p*>.25). However, examination of the cue benefit as a function of signal contrast for the two current variants of the letter identification task (80% predictive and cues not predictive) makes it clear that the reversal in performance seen in the 80% predictive cue condition cannot be explained solely by exogenous capture alone (see Figure 33b). In the 80% predictive condition, there was a substantial cue benefit at the lowest signal contrast, which gave way to a deficit in performance when cues are present at high signal strengths. In contrast, the slight decrement in performance seen when non-predictive cues are present was not modulated by signal contrast, as it stayed rather small and relatively stable across the range of signal strengths examined. These vastly different performance profiles suggest rather strongly that although cues which do not accurate predict target location may slightly decrease performance, this small effect cannot explain the magnitude or shape of the reversal observed in the main experiment.

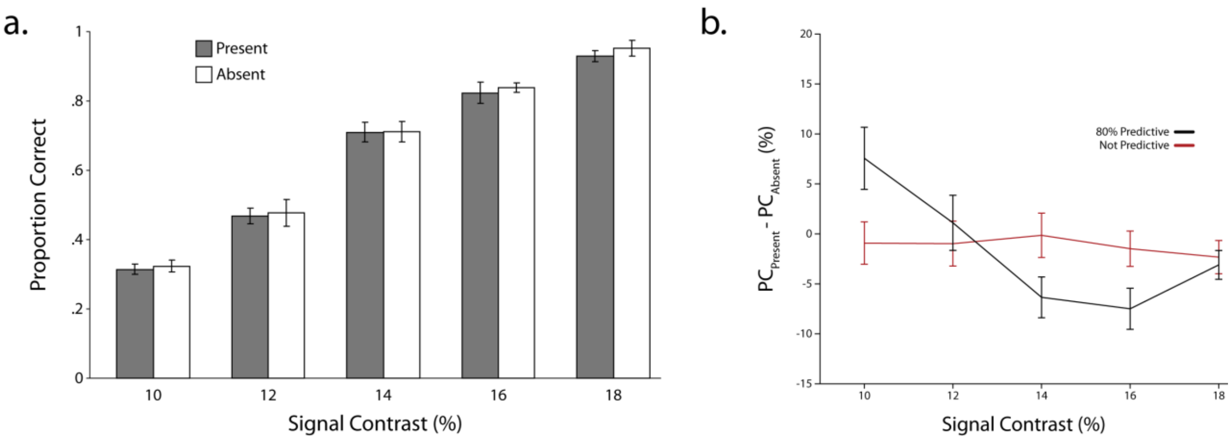

Figure 33. Identification performance as a function of signal contrast for the control task. When cues had no predictive value, their presence has a minimal impact on performance (Panel (a)). Inspection of the cue benefit for the letter identification task in which cues were 80% predictive and that of when cues were not predictive (b) reveals that the marked reversal in performance in the 80% predictive condition cannot be explained simply by the presence of non-informative cues.

Although the impact of non-predictive cues on performance seems to be minimal, another critical point is assessing to what extent the spurious cues may attract eye movements. Figure 34 shows the proportion of eye movements within 2º of any cue for cues present sessions, and the proportion of eye movements towards analogous image regions in cued absent sessions (i.e. regions that would contain the cues if present), as a function of eye movement. Although more saccades are directed towards cued regions overall when cues are present (*F*(1,3)=32.78, *p*=.001), this effect is strongest with initial eye movements, but rapidly decreases (*F*(2,6)=24.24, *p*=.001). In fact, by the third saccade, there is no difference in the proportion of eye movements (*t*(3)=2.03, *p*=.13) directed towards the region which contains the cues (or would contain the cues, in the case of cues absent). These results are consistent with recent work showing that early eye movements may first be driven to salient image regions, with top-down influences holding progressively more weight in eye movement selection as search progresses (Schütz, Trommershäuser, & Gegenfurtner, 2012)

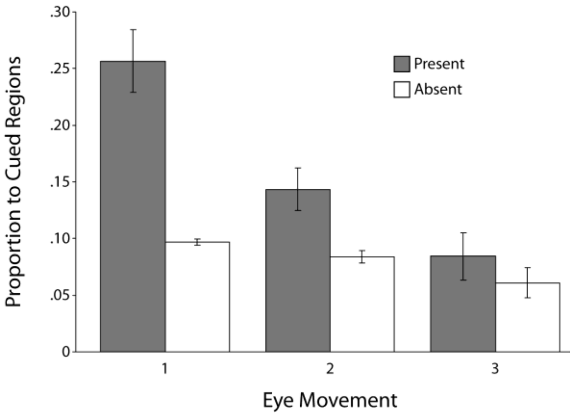

Figure 34. Proportion of eye movements within 2º of any cued location in the not predictive cue task. For Cues absent data, this is the proportion of eye movements to the region which would contain the cue if present. Although more early eye movements are more often directed towards cued locations when they are present, this tendency to foveate cued regions falls off dramatically as search progresses.

In all, there seem to be at least modest effects of non-predictive cues on human observer eye movement selection and search performance, indicating a potential source of suboptimality in human performance. However, direct comparison of the shape and magnitude of the effects in terms of perceptual performance between the control study and the main experiment make it clear that the strategy to fixate cues is primarily driven in a top down fashion, particularly later in search. These results suggest that the detrimental effects of cues seen in the main study may be a result of poor strategic use of cue information.

D. Human-model comparisons

With our finding of a surprising reversal in performance at high signal contrasts in the main experiment, we are left with two critical questions: 1) Would such a reversal ever be predicted by a rational model of search (in this case, the foveated MAP)? Given that the only difference between cues present and cues absent trials is that observers had *more* information about potential target location when cues were present, it seems highly unlikely that a foveated MAP would ever produce similar results. If this is in fact the case, this leads us to question 2): If it would not be predicted by a rational model, in what ways to human observers differ from the foveated MAP? Identifying differences in human/model patterns of behavior and performance provide valuable insight into the origins of our unexpected effect and provide a means of potentially identifying observer and task characteristics which could lead to such a situation.

To assess whether such an effect would be predicted by a rational eye movement model, the foveated MAP model described in Chapter II was implemented with the modified statistical structure of the main experiment. The model was identical to that described in Chapter II (including described visibility map and contrast fitting) with the exception of the modified priors. These modified priors corresponded to the probabilistic structure described for the current experiment (see Figure 30a). Ten thousand simulated trials were run for the foveated MAP model at each signal contrast for both cues present and cues absent conditions.

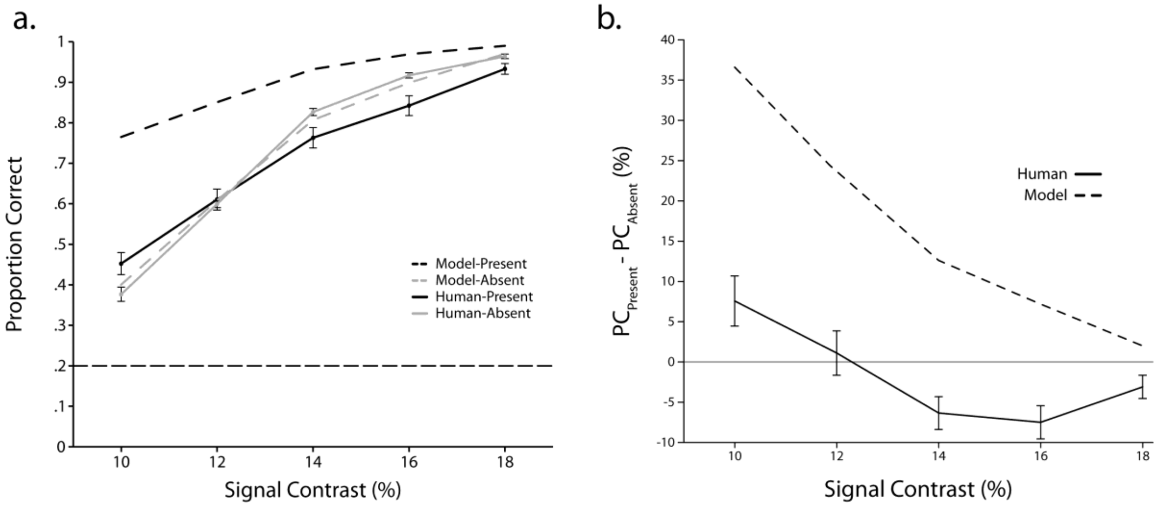

Figure 35. Comparison of human and model performance for the modified letter identification search task. As seen in (a), model performance tracks well with human data when cues are absent. However, predicted model performance far exceeds that of human observers when cues are present. Critically, the model does not predict the reversal in performance seen in human data. This is apparent in (b), where the predicted cue benefit for the model is far above that of humans, and critically is always above zero, indicating no reversal in performance.

Identification performance as a function of signal contrast for both the foveated MAP and human observers is shown in Figure 35a. Although model predictions tracked well with observer performance when cues are absent (*t*(5)=.617, *p*=.717), the foveated MAP far exceeded human observers when cues were present across all signal contrasts (*t*(5)=-3.55, *p*=.016). Critically, there were no traces of the human performance reversal in model predictions. Since the model has more information about potential target locations when cues are present, cues present performance consistently exceeded that of cues absent. A clear demonstration of this can be seen in Figure 35b, which shows the cue benefit ($PC_{present}$-PC_{absent}) for human and model observers. The predicted cue benefit from the foveated MAP was markedly larger than that seen in human data across signal contrasts (*t*(5)=4.98, *p*=.005) and always stayed well above zero, indicating cues facilitated identification performance across the range of contrasts in the experiment.

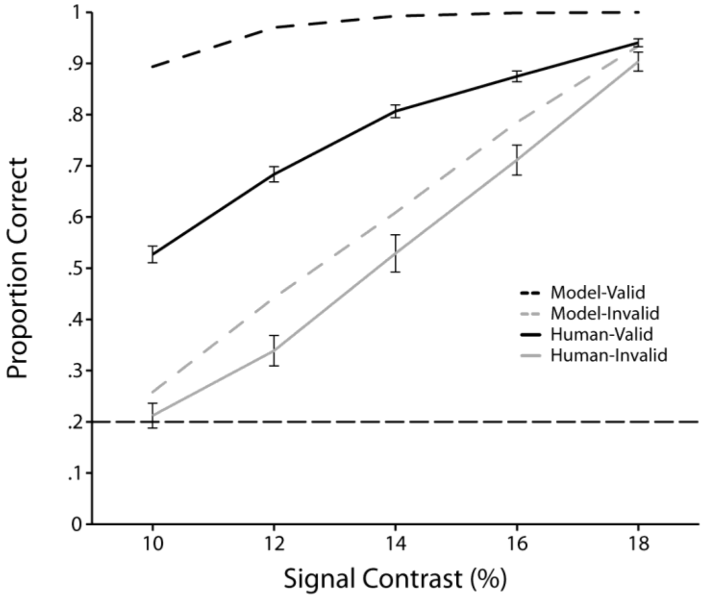

Figure 36. Model and human performance as a function of signal contrast for cues present trials. Although human performance is exceeded by model predictions when the target appears outside of a cued location (invalid), there is an even more pronounced difference when the target appears at a cued location (valid).

Since the performance differences between human and model observers are most pronounced when cues are present, it is worth delving further into this subset of trials. Figure 36 shows performance of a function of signal contrast for humans and the foveated MAP for cues present sessions when the target appeared at a cued location (valid trials) and when it appeared outside of a cued location (invalid trials). We proposed that the reversal in performance in human data may be driven by poor performance in these cues present invalid trials. Indeed, human performance is poorer than model predictions over all contrasts in invalid trials ($t(5)=-3.71$, $p=.01$). However, there was an even larger disparity in human and model performance when the target appeared at a cued location. Thus, it appears that the low observer performance in the cues present trials (which led to the reversal in performance at high signal contrasts) may not have only been related to poor performance when cues didn't contain the target, but also when they did.

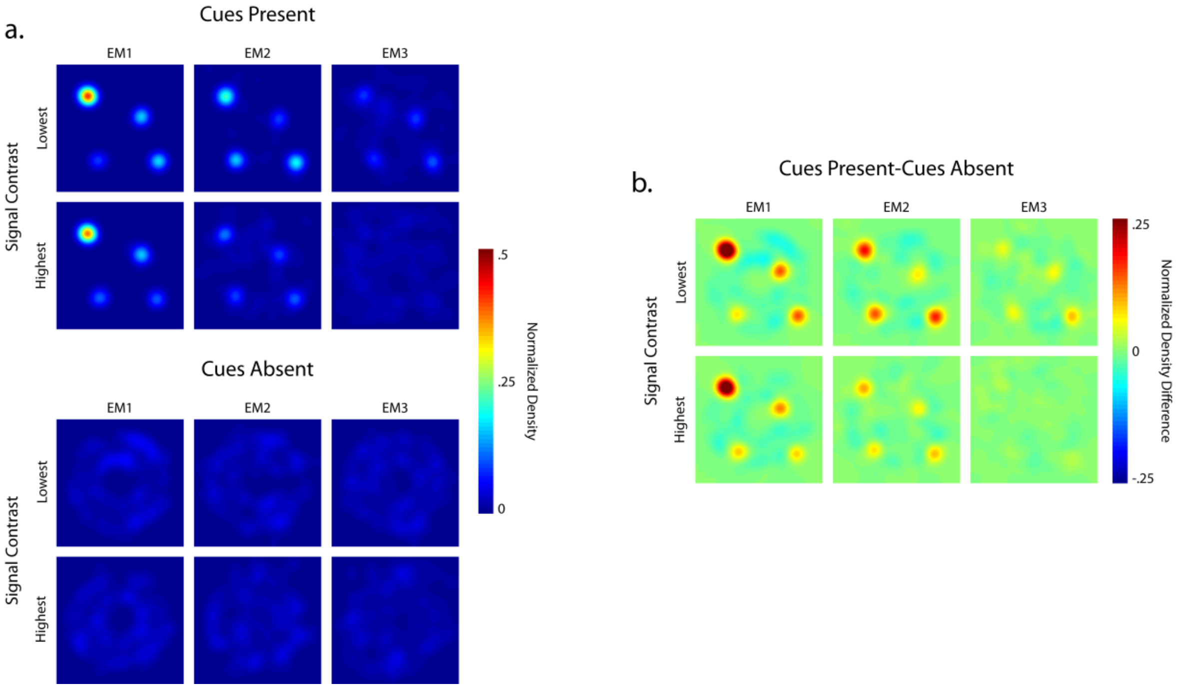

Figure 37. Normalized eye movement distributions for invalid trials for the foveated MAP. Model eye movements cluster around cued locations when cues are present, and are dispersed rather evenly through the image when cues are absent (a). Critically, the persistence of cue-driven eye movements into later eye movements is seemingly decreased for the model observer (Panel (b)).

Further insight into the potential origin of the reversal in human performance can be seen in analyses of eye movement selection. One of our primary hypotheses was that human observers may have been overusing cue information when cues were present and the target appeared outside the cued locations (cues present-invalid). Figure 37a shows normalized eye distributions for the model in these invalid trials for both cues present and absent. Overall, the patterns are remarkably similar to what is seen in human data. Eye movements cluster around cued locations when cues are present, even though the target appears outside these locations in invalid trials. Additionally, predicted eye movements to cues decrease across eye movements as the model accrues information. This decrease is most pronounced when signal contrast is high. When cues are absent, eye movements are dispersed thoughout the image. One critical difference between model and human predictions is seen in Figure 37b, which shows the difference between cues present and absent in normalized eye movements distributions in invalid trials. The human tendency to keep fixating cues into the third eye movement even when the signal is of high contrast appears to be diminished in the model.

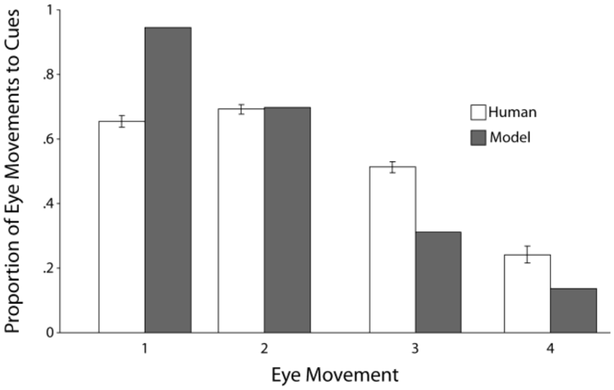

Figure 38. Eye movements to cued regions in cues present invalid trials. Although the model makes more first saccades to the cued regions, human observers make more third and fourth saccades to the cues, indicating that a cue-based eye movement strategy may be persisting longer than it should.

A quantitative assessment of these eye movement patterns is seen in Figure 38, which shows the proportion of eye movements within 2° of any cued location in cues present-invalid trials for both human and model observers. Although the model made more cue driven first saccades (*t*(5)=-16.03, *p*=8.59e-6), by the third (*t*(5)=11.86, *p*=3.76e-5) and fourth (*t*(5)=4.13, *p*=.005) eye movements, human observers were consistently directing more eye movements towards the cued regions, consistent with inspection of eye movement distributions.. This persistent cue-driven eye movement strategy late into trials provides at least some evidence for idea that human observers were overweighting the cues in their eye movement selection when cues were present and did not contain the target.

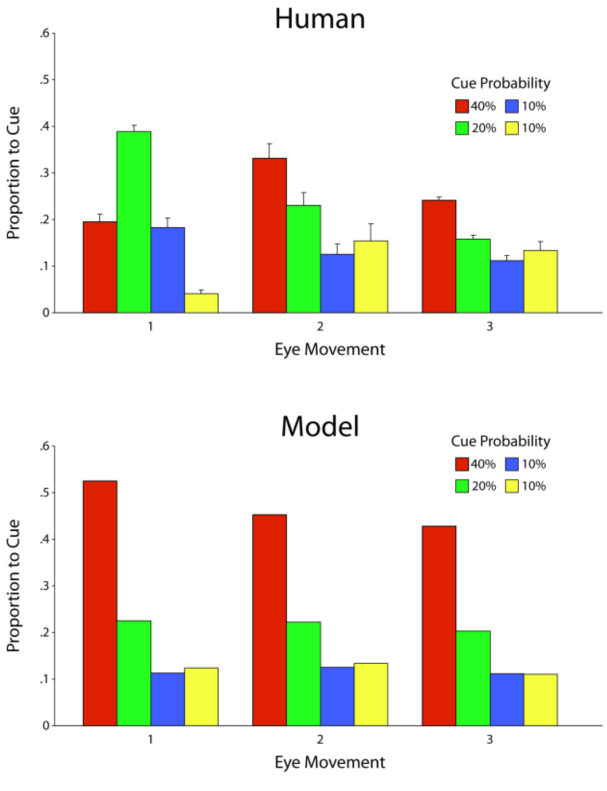

Figure 39. Proportion of eye movements to each cue as a function of eye movement when cues are present. Observers make less eye movements to all cues overall in comparison to the foveated map. Moreover, the distribution of the eye movements which do go to the cues differs considerably in the first eye movement. Observers direct a large proportion of their initial eye movements to the 2nd highest probability cue, while the foveated MAP directs far more to the highest probability cue.

However, our comparisons of human and model performance also showed substantial suboptimality in human performance when cues were present and targets appeared at cued locations (cues present-valid). Is there any insight to be gleaned by inspecting eye movement selection for humans and models in those trials? Figure 39 shows the proportion of eye movements to each cue (labeled by its probability of containing the target) as a function of eye movement. Critically, observers direct less eye movements overall to the cued regions when cues are present than the foveated MAP ($t(5)=-16.76$, $p=6.9e-6$). Along with the tendency to direct less eye movements to cues when present overall, there are marked differences in which cues received the cue-driven eye movements that were made. Regardless of eye movement, the foveated MAP prioritized cue-driven in a manner which reflects the cue probabilities (higher probability cues garnered more fixations). In contrast, what cue human observers directed the most eye movements to changed as a function of eye movement ($F(26,30)=24.81$, $p<.001$). Although the highest probability cue was foveated the most in the second and third eye movement, observers actually directed the largest proportion of eye movements to the cue of the second highest probability of containing the target in the first eye movement (paired t, all $p<.01$), similar to the pattern of eye movements seen in the original task in Chapter II. Clearly, human and model eye movement selection diverged considerably when cues were present.

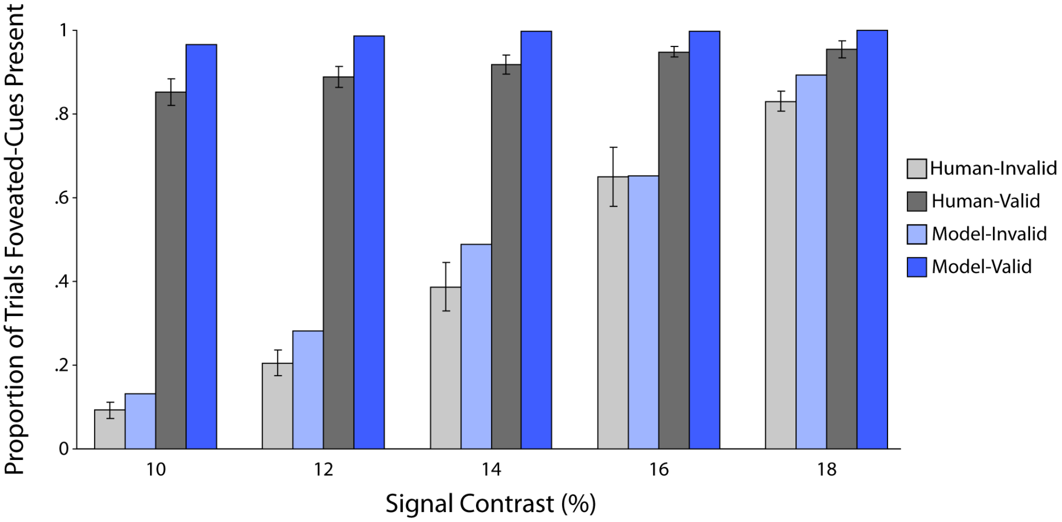

Figure 40. Proportion of trials in which the target was foveated in cues present trials for human and model observers. The model is able to foveate the target more often when cues are present, both when the target appears at a cued location (valid) and when it does not (invalid).

But what are the consequences of these eye movement patterns on observer ability to foveate the target when cues are present? Figure 40 shows the proportion of trials in which the target was foveated for human and model observers in cues present trials. The foveated MAP was able to direct its gaze towards the target significantly more often both when the target appeared at a cued location (valid; *t*(4)=-6.42, *p*=.002) and when it did not (invalid; *t*(4)=-3.70, *p*=.013). We had originally hypothesized that human observers may be overusing the cues when the target appeared outside of them, which prevented foveation of the target and led to decreased performance. The current analysis suggests that, while this may be the case, human observers also failed to fixate the target as well as a foveated MAP when the target appeared at cued locations. This provides further converging evidence that the reversal in human performance may be related to more general suboptimalities in utilizing cue information, regardless of if the target appears at a cued location or not.

E. Conclusions

Surprisingly, with a modified version of the letter identification task introduced in Chapter II, we found that the presence of partially predictive spatial cues did not improve performance overall, and actually hindered search accuracy at high signal contrasts. We originally hypothesized that this unexpected effect was due to overuse of the cues when they were present, but the target appeared at an uncued location. However, the implementation of a foveated MAP eye movement model revealed sizable differences between human and model performance an eye movements in a more general sense. Our results thus indicate that the reversal in human performance may be attributable to suboptimal cue usage both when the target appears at cued locations as well as when it does not.

Despite a plethora of research supporting the idea that predictive cues enhance perceptual performance, our results demonstrate that in visual search with the additional complexities of the real world (i.e. including eye movements and large uncertainty about the target's location and visibility), predictive cues can surprisingly hinder human performance. As such, the conception of predictive spatial cues being purely facilitatory deserves further inspection. Whether cues improve performance does not depend simply on their presence, but also on how effectively observers can utilize them during task performance. In simple tasks with small search areas, targets of known detectability, and no eye movements, predictive spatial cues consistently lead to improved performance (e.g. Cameron et al., 2002; Eriksen & Yeh, 1985; Posner, 1980). However, when search displays become more complex and eye movements are required, we show that inefficiencies in how observers integrate cue information into eye movement selection and perceptual decisions can markedly reduce and even reverse cue benefits. To the best of our knowledge, the current research is the first to demonstrate such a situation and critically calls into question the ubiquity of predictive cue benefits.

In addition to enhancing our understanding of predictive cues, what are the implications of the current findings for specialized real world search tasks? Predictive synthetic cues have become an increasingly common tool in potentially aiding human search performance in a variety of important and sometimes life-critical vocations. Perhaps the most prevalent example is the use of computer aided detection (CAD) systems, which overlay synthetic cues on medical image regions likely to contain masses, to aid interpretation of x-ray mammograms (Fenton et al., 2011; Gur & Sumkin, 2006). Although an estimated 75% of x-ray mammograms are now read with CAD (V. M. Rao et al., 2010), how much CAD enhances performance and when it is most (if at all) useful are still topics of debate (Drew et al., 2012; Fenton et al., 2007; Gur & Sumkin, 2006; Philpotts, 2009). The use of synthetic cues to aid performance has also emerged in the analysis of satellite and aerial images (Willhauck et al., 2000) as well as in air traffic control and pilot navigation and collision avoidance (Metzger & Parasuraman, 2006; Wickens et al., 1999), with mixed results (Fadden, Ververs, & Wickens, 1998). In these life-critical search situations, the current work suggests that even though cues provided by complex image analysis algorithms may provide information that could enhance search proficiency, a detailed understanding of how human observers utilize these cues is necessary to understand their impact on task performance.

In all, the current research provides an opportunity to reconceptualize putative benefits of predictive cues and highlights the necessity of understanding what factors lead to performance enhancement in their presence.

V. General conclusions

We set out to investigate not only if, but how well, human observers can use predictive spatial cues to guide eye movements and enhance visual search accuracy. There plentiful evidence that humans exploit predictive spatial structure natural images and environments to guide search (Eckstein et al., 2006; Mack & Eckstein, 2011; Spotorno et al., 2014). However, the underlying statistics of natural scenes are poorly understood, thwarting objective assessment of these strategies. The use of synthetic spatial cues allows the experimenter control over image statistics, providing the information necessary to develop computationally tractable models of rational search. However, previous investigations of search with synthetic cues have generally focused on single fixation search (Hawkins et al., 1990; Nakayama & Mackeben, 1989), leaving objective assessment of cue use during naturalistic, multi-fixation search poorly understood.

In a novel letter identification search task, we find evidence that human observers are capable and willing to integrate spatially predictive cues into oculomotor strategies to enhance search performance. This finding falls in line with scene context literature, which has similarly shown that observers move their eyes to regions of the visual environment that are likely to contain the target. Critically, the implementation of a Bayesian foveated MAP eye movement model revealed that the performance gain experienced by observers in the presence of predictive cues falls far short of that of a near-optimal observer. While there is evidence for a potential host of suboptimalities (e.g. eye movement selection, spatial masking/crowding, inefficiencies in perceptual decision mechanisms), what is clear is that human observers are, at least in the case of synthetic cues, suboptimal in their usage of this predictive information during search.

Despite this suboptimality, we also see evidence that observers are able to flexibly adapt their eye movement strategies to maximize performance in the face of changing task demands. In a task which required eye movements be directed to locations of the visual display which had no chance of containing the target, observers were able to quickly and efficiently adopt an atypical pattern of saccadic selection to enhance search. Although a learning study for the remote cue task revealed that human observers may not readily explore or adopt such strategies when the benefits of doing so are not explicitly presented, the human observers still display remarkable flexibility in their ability to tailor their search strategy to novel tasks.

Finally, we show a situation in which the presence of predictive cues actually hinders human performance at high signal contrasts. Critically, this effect was tied to a search situation in which cues did not fully predict target location. The large amount of spatial uncertainty linked to partially predictive cues is a hallmark of natural and vocational search, generating a clear message the simple presence of predictive cues is not sufficient to enhance search performance with human observers. While a foveated MAP searcher reveals modest suboptimality in human eye movement patterns, the precise host of factors which lead to the surprising performance reversal in human data warrant further investigation. With the growing prominence of computer aid in vocational search, there is a clear impetus to understand the dynamics of human-cue interactions to maximize searcher performance in these often life-critical search scenarios.

Human observers are able to integrate predictive cues into their search behaviors to enhance accuracy and efficiency. However, they do not do so without substantial suboptimality, with predictive cues even hindering performance in extreme cases. Given the growing importance of computer aided search in life-critical settings, the work presented here provides a valuable starting point from which to objectively assess and begin to optimize human-cue interactions.

References

1. 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.

2. Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R., & Soo, M. S. (2003). Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion. *American Journal of Roentgenology*, *181*(4), 1083–1088.

3. Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), 13.
4. Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

5. Bashinski, H., & Bacharach, V. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Attention, Perception, & Psychophysics*, *28*(3), 241–248. doi:10.3758/BF03204380

6. Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1341–1355.

7. Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. doi:10.1080/13506280500165188

8. Burgess, A. E., & Ghandeharian, H. (1984). Visual signal detection. II. Signal-location identification. *JOSA A*, *1*(8), 906–910.

9. Cameron, E. L., Tai, J. C., & Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vision Research*, *42*(8), 949–967.
10. Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. doi:10.1016/j.visres.2011.04.012

11. Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-011-0107-8

12. Chan, H. P., Doi, K., Vyborny, C. J., Schmidt, R. A., Metz, C. E., Lam, K. L., … MacMahon, H. (1990). Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis. *Investigative Radiology*, *25*(10), 1102.

13. Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178.

14. Ciatto, S., Del Turco, M. R., Risso, G., Catarzi, S., Bonardi, R., Viterbo, V., … Indovina, P. L. (2003). Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography. *European Journal of Radiology*, *45*(2), 135–138. doi:10.1016/S0720-048X(02)00011-6

15. Dosher, B. A., & Lu, Z. L. (2000). Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*(10), 1269–1292.

16. Drew, T., Cunningham, C., & Wolfe, J. M. (2012). When and Why Might a Computer-aided Detection (CAD) System Interfere with Visual Search? An Eye-tracking Study. *Academic Radiology*, *19*(10), 1260–1267.

17. Droll, J. A., Abbey, C. K., & Eckstein, M. P. (2009). Learning cue validity through performance feedback. *Journal of Vision*, *9*(2).

18. Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5).

19. Eckstein, M. P., Ahumada, A. J., Jr, & Watson, A. B. (1997). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *14*(9), 2406–2419.

20. Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–980.

21. Eckstein, M. P., Mack, S. C., Liston, D. B., Bogush, L., Menzel, R., & Krauzlis, R. J. (2013). Rethinking human visual attention: Spatial cueing effects and optimality of decisions by honeybees, monkeys and humans. *Vision Research*. doi:10.1016/j.visres.2012.12.011

22. Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, *2*(1), 25–45.

23. Ehinger, K., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling visual search in a thousand scenes: The roles of saliency, target features, and scene context. *Journal of Vision*, *9*(8), 1199. doi:10.1167/9.8.1199

24. Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance;*

Journal of Experimental Psychology: Human Perception and Performance, *11*(5), 583.

25. Fadden, S., Ververs, P. M., & Wickens, C. D. (1998). Costs and Benefits of Head-Up Display Use: A Meta-Analytic Approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(1), 16–20. doi:10.1177/154193129804200105

26. Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia*, *27*(4), 461–470. doi:10.1016/0028-3932(89)90051-1

27. Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390.

28. Fenton, J. J., Abraham, L., Taplin, S. H., Geller, B. M., Carney, P. A., D'Orsi, C., … Barlow, W. E. (2011). Effectiveness of computer-aided detection in community mammography practice. *Journal of the National Cancer Institute*, *103*(15), 1152–1161.

29. Fenton, J. J., Taplin, S. H., Carney, P. A., Abraham, L., Sickles, E. A., D'Orsi, C., Elmore, J. G. (2007). Influence of Computer-Aided Detection on Performance of Screening Mammography. *New England Journal of Medicine*, *356*(14), 1399–1409. doi:10.1056/NEJMoa066099

30. Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, *37*(5), 617–631.

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

32. Geisler, W. S. (2003). *Ideal Observer analysis*. Boston: MIT Press.

33. Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.*, *59*, 167–192.

34. Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781. doi:10.1016/j.visres.2010.09.027

35. Giesbrecht, B., Sy, J. L., & Guerin, S. A. (2013). Both memory and attention systems contribute to visual search for targets cued by implicitly learned context. *Vision Research*, *85*, 80–89.

36. Green, D. M., & Swets, J. A. (1989). *Signal Detection Theory and Psychophysics*. Peninsula Pub.

37. Guo, K., Mahmoodi, S., Robertson, R., & Young, M. (2006). Longer fixation duration while viewing face images. *Experimental Brain Research*, *171*(1), 91–98. doi:10.1007/s00221-005-0248-y

38. Gur, D., & Sumkin, J. H. (2006). CAD in screening mammography. *American Journal of Roentgenology*, *187*(6), 1474–1474.

39. Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 802–811. doi:10.1037/0096-1523.16.4.802

40. Henderson, J. M. (1991). Stimulus discrimination following covert attentional orienting to an exogenous cue. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 91.

41. Henderson, J. M. (1996). Spatial precues affect target discrimination in the absence of visual noise. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 780.

42. Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

43. Hietanen, J. K., Nummenmaa, L., Nyman, M. J., Parkkola, R., & Hämäläinen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMRI study. *Neuroimage*, *33*(1), 406–413.

44. Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, *17*, 273–291. doi:10.1080/13506280802193367

45. Hollingworth, A. (2012). Guidance of visual search by memory and knowledge. In *The Influence of Attention, Learning, and Motivation on Visual Search* (pp. 63–89). Springer.

46. Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic Control of Visual Attention. *Psychological Science*, *12*(5), 360–365. doi:10.1111/1467-9280.00367

47. Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456.

48. Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science*.

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

50. Lee, D. K., Koch, C., & Braun, J. (1997). Spatial vision thresholds in the near absence of attention. *Vision Research*, *37*(17), 2409–2418.

51. Lu, Z. L., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review*, *115*(1), 44–82.

52. Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 887–904. doi:10.1037/0096-1523.20.4.887

53. Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, *11*(9).

54. Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, *10*(2).

55. Metzger, U., & Parasuraman, R. (2006). Effects of Automated Conflict Cuing and Traffic Density on Air Traffic Controller Performance and Visual Attention in a Datalink Environment. *The International Journal of Aviation Psychology*, *16*(4), 343–362. doi:10.1207/s15327108ijap1604_1

56. Moffitt, K. (1980). Evaluation of the fixation duration in visual search. *Attention, Perception, & Psychophysics*, *27*(4), 370–372.

57. Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391. doi:10.1038/nature03390

58. Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, *8*(3). doi:10.1167/8.3.4

59. Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*(11), 1631–1647.

60. Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621. doi:16/j.visres.2005.08.025

61. Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research*, *45*(14), 1867–1875.

62. Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, *109*(48), E3314–E3323.

63. Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science*, *24*(7), 1216–1225.

64. Philpotts, L. E. (2009). Can Computer-aided Detection Be Detrimental to Mammographic Interpretation? 1. *Radiology*, *253*(1), 17–22.

65. Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, *33*(7), 993–999.

66. Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

67. Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. doi:10.1037/0096-3445.109.2.160

68. Preston, T. J., Guo, F., Das, K., Giesbrecht, B., & Eckstein, M. P. (2013). Neural Representations of Contextual Guidance in Visual Search of Real-World Scenes. *The Journal of Neuroscience*, *33*(18), 7846–7855.

69. Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*(11), 1447–1463.

70. Rao, V. M., Levin, D. C., Parker, L., Cavanaugh, B., Frangos, A. J., & Sunshine, J. H. (2010). How widely is computer-aided detection used in screening and diagnostic mammography? *Journal of the American College of Radiology*, *7*(10), 802–805.

71. Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences*, *109*(19), 7547–7552.

72. Shimozaki, S. S., Schoonveld, W. A., & Eckstein, M. P. (2012). A unified Bayesian observer analysis for set size and cueing effects on perceptual decisions and saccades. *Journal of Vision*, *12*(6), 27.

73. Smith, P. L. (2000). Attention and luminance detection: effects of cues, masks, and pedestals. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(4), 1401.

74. Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, *14*(2), 7.

75. Swensson, R. G., & Judy, P. F. (1981). Detection of noisy visual targets: Models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics*, *29*(6), 521–534. doi:10.3758/BF03207369

76. Talgar, C. P., Pelli, D. G., & Carrasco, M. (2004). Covert attention enhances letter identification without affecting channel tuning. *Journal of Vision*, *4*(1).

77. Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391–412.

78. Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, *113*(4), 766.

79. Van den Berg, R., Roerdink, J. B., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision*, *7*(2), 14.

80. Verghese, P. (2012). Active search for multiple targets is inefficient. *Vision Research*. doi:10.1016/j.visres.2012.08.008

81. Vincent, B. (2011). Covert visual search: Prior beliefs are optimally combined with sensory evidence. *Journal of Vision*, *11*(13). doi:10.1167/11.13.25

82. Watson, A. B. (1987). The ideal observer concept as a modeling tool. *Frontiers of Visual Science*, 32–37.

83. Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168.

84. Wickens, C. D., Conejo, R., & Gempler, K. (1999). Unreliable Automated Attention Cueing for Air-Ground Targeting and Traffic Maneuvering. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(1), 21–25. doi:10.1177/154193129904300105

85. Willhauck, G., Schneider, T., De Kok, R., & Ammer, U. (2000). Comparison of object oriented classification techniques and standard image analysis for the use of change detection between SPOT multispectral satellite images and aerial photos. In *Proceedings of XIX ISPRS Congress* (Vol. 33, pp. 35–42).

86. Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. doi:10.3758/BF03200774

87. Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835. doi:10.1037/a0013118

88. Zhang, S., & Eckstein, M. P. (2010). Evolution and Optimality of Similar Neural Mechanisms for Perception and Action during Search. *PLoS Computational Biology*, *6*(9). doi:10.1371/journal.pcbi.1000930