

# Conditional Probability in Visual Search

by

Bryan Cort

A thesis  
presented to the University of Waterloo  
in fulfillment of the  
thesis requirement for the degree of  
Master of Arts  
in  
Psychology

Waterloo, Ontario, Canada, 2013

©Bryan Cort 2013

## **AUTHOR'S DECLARATION**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

I investigated the effects of probability on visual search. Previous work has shown that people can utilize spatial and sequential probability information to improve their performance on visual attention tasks. My task was a simple visual search in which the target was always present among a field of distractors, and could take one of two colors. The absolute probability of the target being either color was 0.5; however, the *conditional probability* – the likelihood of a particular color given a particular cue combination – varied from 0.1 to 0.9. I found that participants searched more efficiently for high conditional probability targets and less efficiently for low conditional probability targets. This modulation of efficiency was reduced or abolished when participants were not explicitly informed of the cue-target relationships. After establishing this effect, I investigated its mechanism using eye tracking methods. Early in trials, participants fixated preferentially, but not exclusively, on areas of the screen which contained predominantly stimuli of the color to which they had been cued. As the trial progressed, this color bias shifted to the target color. I conclude that search efficiency is modulated by the conditional probability of target features and that this is a top-down process that benefits from explicit knowledge of the probabilistic relationship between cues and targets, and that the modulation is a result of more efficient eye movements towards stimuli with a greater probability of being the target of search.

## **Acknowledgements**

I would like to thank my supervisor, Britt Anderson, for his guidance and teaching over the time I have worked in his laboratory. I would also like to thank Cameron Smith for his assistance with data collection. Lastly, I would like to thank all the people who participated in this research, without whom this work would not have been possible.

## Table of Contents

AUTHOR'S DECLARATION .....	ii
Abstract .....	iii
Acknowledgements .....	iv
Table of Contents .....	v
List of Figures .....	vii
List of Tables.....	ix
Chapter 1 Foreword.....	1
Chapter 2 Visual Search.....	2
2.1 History of Visual Search .....	2
2.2 Models and Theories of Visual Search.....	3
2.2.1 Serial Models of Search.....	3
2.2.2 Parallel Models of Search.....	4
2.2.3 Eye Tracking in Models of Visual Search.....	5
Chapter 3 Limitations of Current Paradigms, and Alternate Approaches .....	7
3.1 Probability, Attention, and Continuous vs. Binary Specifications .....	7
3.2 Statistical and Probability Learning .....	9
Chapter 4 Probability Cueing in an Inefficient Search Task .....	10
4.1 My Paradigm and Experiment 1 .....	10
4.1.1 Experimental Design .....	10
4.1.2 Apparatus.....	13
4.1.3 Experiment 1 Results and Discussion .....	14
4.2 Experiment 2 .....	16
4.2.1 Experimental Design .....	16
4.2.2 Results and Discussion.....	17
4.3 Experiment 3 .....	18
4.3.1 Experimental Design .....	18
4.3.2 Results and Discussion.....	19
4.4 Experiment 4 .....	22
4.4.1 Experimental Design .....	22
4.4.2 Results and Discussion.....	23
Chapter 5 Eye Tracking in my Task.....	27

5.1 Experiment 5.....	27
5.1.1 Experimental Design.....	27
5.1.2 Eye Tracking Data Collection, Pre-Processing, and Transformations.....	27
5.1.3 Reaction Time Results.....	29
5.1.4 Eye Tracking Results.....	30
Chapter 6 General Discussion.....	41
6.1 Recap of Results.....	41
6.2 Discussion.....	42
Appendix A : Eye Tracking Data Processing Details.....	47
Bibliography.....	50

## List of Figures

**Figure 4-1:** Experiment 1 task flow and magnified stimuli. Cue 1 appeared in the center of the screen for 1 second. After it disappeared, cue 2 appeared in a random location. After another 1-second interval, the search array appeared with the target and all distractors in random, non-overlapping locations. Possible stimuli colors were magenta and cyan. The size of every stimulus was 0.8 degrees of visual angle. Participants reported which of the four corners of the target was missing with the corresponding arrow key on the computer keyboard. .... 11

**Figure 4-2:** All possible cue combinations and their associated target color probabilities for each experiment. In experiments 1 and 4, when both cues were the same color (cyan or magenta) the target was 90% likely to be that color. When the cues were different colors, they were uninformative as to the target color. In experiment 2, all cue combinations were uninformative. In experiment 3, the presence or absence (counterbalanced across participants) of cue 2 indicated the validity of cue 1; a valid cue 1 predicted target color with 90% accuracy. .... 12

Figure 4-3: RTs by distractor number for experiment 1. The higher the conditional probability, the faster and more efficient the search. .... 14

Figure 4-4: Response times by distractor number for Experiment 2. When probability cuing is removed, search efficiency is equal regardless of color priming. .... 17

Figure 4-5: Response times by distractor number for Experiment 3. Search efficiency is modulated by conditional probability even when color priming is reduced. .... 19

Figure 4-6: Response times by distractor number for each of the three conditional probabilities in Experiment 3. Cue 1 never matched the target color when conditional probability was 0.1, always matched when conditional probability was 0.9, and matched half the time when conditional probability was 0.5. .... 21

Figure 4-7: Response times by distractor number for Experiment 4. Conditional probability modulates search efficiency only when participants have explicit knowledge. .... 23

Figure 5-1: Fixations, represented by yellow dots, in a typical trial. Duration is represented by the radius of the circle, and each fixation is numbered to show its order within the trial. .... 29

Figure 5-2: Reaction times in Experiment 5. As in Experiment 1, search efficiency increased with conditional probability. .... 30

Figure 5-3: Closest numbered stimulus for each fixation, categorized by number of stimuli on screen. Stimulus 0 is the target; stimulus 1 is the second cue (which stays on screen for the duration of the search). .... 31

Figure 5-4: Number of stimuli within 3 degrees of visual angle (top) and density score (bottom) by trial completion. The shaded area shows standard error. .... 33

Figure 5-5: Fixation duration by number of stimuli within 3 degrees (top) and density score (bottom). ..... 35

Figure 5-6: Match score by trial completion for each level of conditional probability. Searchers look preferentially towards the areas of the display containing stimuli of the cued color. The cued color matches the target color when conditional probability is higher, and does not match when conditional probability is low. .... 37

Figure 5-7: Fixation number (left) for all fixations from all trials, and fixations remaining (right) for fixations (from all trials) falling within 3 degrees of visual angle of the target..... 39



## List of Tables

Table 4-1: Intercepts and coefficients for the models generated by simple linear regression on the data from the <i>0.9 conditional probability</i> condition in experiment 1. ....	15
Table 4-2: Intercepts and coefficients for the models generated by simple linear regression on the data from the <i>0.9 conditional probability</i> condition in Experiment 3.....	20
Table 4-3: Intercepts and coefficients for the models generated by simple linear regression on the data from the <i>full information, 0.9 conditional probability</i> condition in Experiment 4.....	24
Table 5-1: Intercepts and coefficients for the <i>Stimuli Within 3 Degrees ~ Trial Completion</i> and <i>Density Score ~ Trial Completion</i> models. ....	34
Table 5-2: Intercept and coefficient for the <i>Duration ~ Density Score</i> model.....	35
Table 5-3: Intercept and coefficients of the Match Score ~ Trial Completion * Conditional probability linear model.....	38
Table 5-4: Logistic regression coefficients for the model Detection ~ Conditional Probability + Target Distance.....	40



# Chapter 1

## Foreword

This thesis is an investigation into the closely related areas of visual attention and visual search, with a focus on the effects of endogenous statistical and probabilistic cueing. Visual search and especially visual attention are quite broad categories, and so I would like to make clear at the outset exactly how I mean these terms. Visual search in this work refers to the seeking of a single target item amongst an array of highly similar distractors, distinct from the Posnerian target detection tasks that are often also categorized as search. Visual attention in this work refers to the deployment of visual resources to an area in space, and not to any type of sustained vigilance or differentiation between on and off task thoughts and behaviors. In addition, I should note that when using the term ‘attention,’ unless explicitly specified, I refer to overt attention (that is, accompanied by eye movements) and not covert attention. In the interest of brevity, throughout this document I refer to visual attention as attention, and to visual search as search.

My motivation for pursuing this work is a general dissatisfaction with commonly accepted characterizations of attention as an agency unto itself, and deficiencies of those characterizations regarding probabilistic information in both the lab and the outside world. In summary, I believe that more probabilistically driven definitions and models of attention yield more quantifiable, precise characterizations of attention that offer up better, more accurate predictions of how visual attention works.

Chapter 2 of this manuscript offers brief highlights of foundational search studies and their methodologies, followed by a more detailed look at current models and theories of search and the relatively recent addition of eye tracking methods to search experiments. In Chapter 3 I discuss the limitations of current theories of attention, ideas about how they might be improved by more quantitative, probabilistically driven definitions, and review existing work in this vein; I also provide some background on previous work exploring this topic. I present my own experiments in Chapters 4 (reaction time measures) and 5 (eye tracking) along with discussion of their results.

## **Chapter 2**

### **Visual Search**

In this chapter, I give a brief summary of influential work in search and the methods and measures that such work pioneered. I then discuss in detail the more recent advances in methodology (chiefly eye tracking) and the modern theories and models of search. Such models can be roughly categorized as either serial or parallel processing models, with further distinctions concerning their capacity (limited or unlimited) and treatment of stochastic noise.

#### **2.1 History of Visual Search**

From cavemen foraging for food, to present day people looking for their car keys or favorite coffee mug, visual search is ubiquitous in everyday life. Search ranges from the trivially easy and inconsequential (e.g., the coffee mug) to the brutally difficult and vitally important (e.g., cavemen looking for tracks). Given the enormous range of tasks which involve or rely on visual search, and its value as a method for investigating the basic mechanisms of perception, it is no surprise that it is one of the most investigated topics in psychology.

It was during the 1960s that search began to be commonly used as a method for investigating attention. The foundational work of Neisser (1964) and Estes & Taylor (1964) formalized, respectively, the use of reaction time and accuracy measures in search experiments. In Neisser's work, participants scanned lists of alphanumeric sequences in order to locate a particular target letter, digit, or sequence, with reaction time collected as the primary measure. Assuming that each item was processed with approximately the same efficiency and with near-100% accuracy, reaction time then indexed both the overall speed and the per-item speed of search. The effects of set size (linear increase in reaction time with number of distractors) and of target-distractor similarity (parallel or 'pop out' vs. serial or 'effortful' search) both provided evidence that such assumptions were valid and reaction time was a valid measure.

Estes and Taylor were early pioneers of accuracy measures in search. In their 1964 work, they briefly (sub-100ms presentation times) showed a search array of alphanumeric characters on a screen to participants. In opposition to Neisser, they held search time constant and measured the accuracy with which participants could report displays of varying number of elements; this accuracy measure represented the upper limit of attention on visual perception. They were among the first to

refine accuracy measures based primarily on signal detection and utilizing the above-mentioned short presentation times as well as pre- and post-stimulus masks, controls now common in modern work in similar paradigms. They also pioneered paired stimulus, binary choice methods to control for memory in perception and attention experiments.

Search and attention research has benefitted immensely from advancing technology. For the most parts, these benefits have been incremental, such as more and more precise electronic displays and timing. However, one advancement has been enormously influential in the field. Eye tracking has provided a method to precisely measure the times and locations of ocular fixations and saccades, which before could only be inferred from behavioral data.

Eye tracking is a useful measure in search and other visual tasks due to the foveated structure of the human eye; visual acuity is much greater in the foveal and parafoveal regions in the center of the retina than in the rest of the visual field. This greater visual acuity stems from greater density of photoreceptors in these regions, as well as greater proportion of neurons dedicated to these photoreceptors in the fovea. Due to this asymmetry, during visual scanning and searching humans make many brief, rapid eye movements (saccades) interspersed with longer periods of gaze immobility (fixations).

## **2.2 Models and Theories of Visual Search**

Perhaps the most ubiquitous finding across decades of search experiments is that reaction time increases linearly as distractor set size increases when the target is similar to the distractors, and that set size has little effect when the target is sufficiently different from the distractors. In this section I review some prominent models of search, roughly divided by the serial/parallel distinction.

### **2.2.1 Serial Models of Search**

Feature Integration Theory (FIT) (Treisman & Gelade, 1980) was a widely accepted model of the mechanism of attention in search throughout the 1980s. It held that visual scenes were constructed bottom up by combining basic features (such as color, shape, and orientation) into what was perceived as a unified whole, and that serial attention was the mechanism by which features were combined. Treisman and Gelade conducted a series of experiments using a conjunction search paradigm; their primary finding was that attention was necessary for the correct perception of conjunctions between basic features. Their theory provided parsimonious explanations of many phenomena observed in previous search experiments, such as the distinction between pop-out and

serial search and the appearance of illusory conjunctions. For example, they explained pop-out search as occurring pre-attentively, since targets in this type of search were defined by a single feature which differed from the distractor set; because the target was not defined by multiple features, its identification did not require attention.

FIT inspired and guided a great deal of focused investigation into search and attention throughout the 1980s, and by the end of the decade a number of issues had been raised with the theory. The Guided Search (GS) model (Wolfe, Cave, & Franzel, 1989) and its later iterations GS2 (Wolfe, 1994) and GS4 (Wolfe, 2007) are similar to FIT in that they model attention as a serial processor; however, instead of feature detection being the only operation available without attention, it substitutes a parallel, pre-attentive low level visual mechanism that processes the display in parallel and subsequently guides the serial deployment of attention according to the activation or salience (subject to noise) of each item. This critical modification to the strict dissociation between the parallel feature detection and serial attention deployment asserted by FIT allowed GS to accurately predict search outcomes for many cases in which FIT fell short, such as triple conjunction searches and even special cases of the double conjunction searches on which FIT was built.

### **2.2.2 Parallel Models of Search**

The focus of the work presented here is on overt attention and eye movements, which are serial in nature. For this reason, theories and models which treat attention as a serial process are of most interest to us; however, models of attention as a parallel process deserve mention both because of the highly related nature of overt and covert attention, and because covert attention does play a lesser role even when eye movements are permitted in search, for example when making a fixation on a group or cluster of stimuli rather than a direct fixation on a single stimulus. I therefore provide a brief overview of several significant parallel models.

While serial models explain the relationship between increasing number of distractors and reduced search performance (in both RT and accuracy paradigms) as resulting from the increased time required to process each item in sequence, parallel models have no such constraint. Parallel models of search make the argument that all items in a search array or display can be processed concurrently; thus, there must be some other limiting factor. There are two main classes of parallel models. One (limited capacity models) holds that the global capacity to process items in the display is limited, implying that processing larger arrays will decrease the resources available to process each individual item. The other class of models (noisy unlimited capacity models) holds that capacity is

unlimited but performance is degraded by noise, with the consequence that increasing the set size will then increase the likelihood that any individual distractor will be confused with the target.

One of the earliest examples of a parallel model for search came from Townsend (Townsend, 1972), though by the end of the decade she was in favor of a serial self-terminating model of search (Snodgrass & Townsend, 1980). More recently, Bundesen (1990) has proposed a parallel model of attention, along with a proposal for how it might be implemented neurally (Bundenen, Habekost, & Kyllingsbæk, 2005).

Signal detection models, and their close cousins Bayesian models, are both rooted in the ideas of signal detection theory (Peterson, Birdsall, & Fox, 1954). According to such models, a search array can be considered as a collection of inputs to the searcher. Each item in the array consists of a signal plus some amount of noise. Assuming that the target is known to the searcher, its signal should be greater than the distractors; however, the noise component of each input can cause distractors to have target-like total inputs, and/or noise in the target input can cause it to resemble distractors.

Signal detection models and Bayesian models are differentiated by the rules (decision rules) they use to select a response from this collection of noisy inputs. Signal detection models set a threshold or criterion for the activation a stimulus must reach to be classified a target; depending on the decision rule, such models might choose the first to reach the threshold activation, the stimulus with the greatest activation over threshold, or a random stimulus from the set of all stimuli that exceeded the threshold.

Bayesian models use prior probabilities (such as rates of targets, distractors, particular features, and particular spatial locations) in combination with each input (signal + noise) to calculate for each stimulus that probability that that stimulus is the target. Decision rules for Bayesian models are functions which take those probabilities as inputs and make a response. The most intuitive (and most widely used) decision rule for Bayesian models is the Bayesian Ideal Observer (Geisler, 1989), which chooses the stimulus with the highest probability of being the target as its response. This decision rule usually exceeds human performance, but is useful to compare both other models and human performance to, as it represents the ideal performance on a given task.

### **2.2.3 Eye Tracking in Models of Visual Search**

Eye tracking is often used to investigate and expound on models such as those listed above. However, many models of search are predicated entirely on describing and predicting each individual eye movement within a search, and not just the outcome of the search itself. This kind of modeling and

analysis is beyond the scope of this work, which focuses on the effects of probability in particular on search and eye movements; nonetheless, I provide below a brief overview of prominent eye movement models.

Eye movement models are for the most part analogous to the models of search outlined above, with the exclusion of parallel models or the substitution of serial processes into such models (since eye movements are obviously serial in nature). Saliency models (Itti, Koch, & Niebur, 1998; Itti & Koch, 2000), as the name would suggest, compute a saliency score (from the entirety or a subset of the features of the stimuli in the visual field) for each stimulus in a manner similar to Guided Search's saliency ranking of stimuli. Eye movements then proceed serially through the array from high to low saliency until the target is located.

Saccadic targeting models bear more similarity to Bayesian models of search, with the probability of being the target computed for each item in the array, and saccades made based on those probabilities. Models differ primarily on what additional factors are included in the model, such as degraded sensory information with retinal eccentricity (Berkley, Kitterle, & Watkins, 1975), and distance required to make each saccade (Araujo, Kowler, & Pavel, 2001). In addition, such models predict different eye movement behavior based on the explicit, particular goal of the searcher; for example, models in which the goal is to directly fixate the target (Beutter, Eckstein, & Stone, 2003; Eckstein, Beutter, & Stone, 2001; Najemnik & Geisler, 2008) predict different saccade behavior than models in which the goal is to maximize the information gained about both (probable) target and distractors (Najemnik & Geisler, 2005).



## Chapter 3

### Limitations of Current Paradigms, and Alternate Approaches

In this chapter, I discuss the shortcomings of binary characterizations of attention, and review the body of work which supports a more graded and continuous account of attention. To introduce these ideas I begin with a brief recounting of the work which has demonstrated graded effects of visual cuing and attention. I then review data on spatial and sequential probability cues. Finally, I present some background on the closely related area of statistical learning.

#### 3.1 Probability, Attention, and Continuous vs. Binary Specifications

Manipulating and biasing visual attention, commonly referred to as cuing attention, is ubiquitous in the study of search and attention. The effects of cues in attentional tasks are well documented (Carrasco, 2011; Posner & Cohen, 1984; Wright & Ward, 2008). For the most part, cuing and attention are framed in terms of valid/invalid and present/absent, but such binary characterizations may obscure important distinctions (Anderson, 2011). Probability is an alternative characterization that quantifies the factors that direct perceptual processing resources; probability lends itself to a continuous account, and thus better describes many experimental manipulations.

As far back as 1980, it has been recognized that cuing is not an all or none process; the attentional value of a cue is graded by its validity (Eriksen & Yeh, 1985; Jonides, 1980; Madden, 1992; Riggio & Kirsner, 1997). Jonides (1980) employed a circular 8-item search display in which subjects were shown a neutral, valid, or invalid cue, and the predictive value of the valid cue varied between 30%, 50%, and 70%; the magnitudes of reaction time (RT) cost (for invalid cues) and benefit (for valid cues) increased in proportion to the validity of the cue. Eriksen and Yeh (1985) presented subjects with an identical display and varied the predictive value of both a primary spatial cue and a secondary spatial cue, and also found that RTs improved in proportion to the predictive value of the cues.

The last decade has seen an increase in interest on this topic. Vossel, Thiel, and Fink (2006) collected fMRI and RT data in a slight variation of Posner's (Posner, Snyder, & Davidson, 1980) seminal cuing task. They varied cue validity between a 60% condition and a 90% condition, confirming the finding that greater cue validity resulted in faster RTs and showing that cue validity modulates activation in a right-hemispheric fronto-parietal attentional network. In a similar design, Gould, Rushworth, and Nobre (2011) also demonstrate such a relationship between cue validity and

reaction time. Hahn, Ross and Stein (2006) demonstrated probability spatial cuing effects in a simple search task. Targets occurred at one of four peripheral locations, and a central symbolic cue indicated which quadrants were of greater probability for a particular trial. Any number (up to all four) of the quadrants could be cued on any given trial, and the cue validity was 80%, regardless of the number of quadrants cued. This yielded a graded cue validity ranging from 25% (all quadrants cued) to 80% (one quadrant cued). The primary behavioral result was a monotonic relationship between the number of primed positions and RT, with fewer primed locations (and thus greater cue validity) generating faster RTs.

The vast majority of work on graded cuing effects has involved spatial cues. To my knowledge, there is only one investigation of graded effects of feature cue validity. Egner, et al. (2008) explored spatial and feature cuing when the predictive value of the cues was parametrically varied. They utilized a simple search task in which a fixed grid of four locations contained diamonds that were either red or blue in color and left or right in spatial position. Central cues communicated independent information about location and color of the target. The validity of the spatial and color cues was 50, 70, or 90%, and the probability that any particular diamond was the target was the product of the individual cues. The task required the participants to locate the target diamond, which was distinguished by a missing corner. The principle result was a relationship for cue predictive value and RT. Trials with 90% valid cues were faster than 70% valid cue trials, and 90% invalid cue trials (or alternatively, 10% predictive value) were slower than 70% invalid cue trials (alternatively, 30% predictive value). There was no significant effect of cue dimension, meaning that both spatial and feature cues had equivalent effects. The relationship was non-linear because the magnitude of cuing effects between 70 and 90% was less than that for 50 and 70%. The magnitude of each cue was greatest when the other was non-informative (50% predictive value).

When tested, the relationship between attention and cue predictive value is found to be graded rather than all or none. What defines a cue's predictive value is its probabilistic relation to the target. This asserts an equivalence between cues and prior probability. There is a great deal of evidence that statistical relationships can be learned implicitly and on line (Chun, 2000; Chun & Jiang, 1998; Chun & Jiang, 1999; Druker & Anderson, 2010; Fiser & Aslin, 2001; Fiser & Aslin, 2002; Geng & Behrmann, 2002; Geng & Behrmann, 2005; Y. Jiang & Leung, 2005; Y. V. Jiang, Swallow, & Rosenbaum, 2012; Ono, Jiang, & Kawahara, 2005; Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Saffran, Johnson, Aslin, & Newport, 1999; Williams, Pollatsek, Cave, & Stroud, 2009).

### 3.2 Statistical and Probability Learning

As in the cuing literature, much of the work on implicit statistical learning involves the learning of spatial regularities. In Chun and Jiang's (1998) classic contextual cuing task, subjects searched for a rotated T among heterogeneously rotated L's. Chun and Jiang showed that when some search layouts were repeated across blocks, performance improved; this improvement was observed despite the fact that participants were neither aware of the manipulation nor able to identify the repeated search layouts. Williams et al. (2009) varied the probability of a target appearing in one of three possible clusters; their participants learned and acted upon the spatial probabilities associated with the clusters quickly and automatically.

I am aware of comparatively few studies that have investigated non-spatial aspects of probability learning. Chun and Jiang (1999) implemented a variation on their previous (Chun & Jiang, 1998) search task in which instead of searching for a T among L's, participants searched for a shape with a vertical axis of symmetry among shapes with non-vertical axes of symmetry, and they found that distractor identities were effective cues of unique target identity. However, Endo and Takeda (2004) employed a slightly modified version of the same Chun and Jiang (1998) task, in which participants searched for a closed contour among open contours. They found that although target position could be cued both by distractor position and distractor identity, target identity could not be effectively cued by either distractor position or identity. Although the findings of these two studies on statistical learning in search are contradictory, a growing body of work using non-search paradigms suggests that non-spatial probabilities, including feature probabilities similar to those of interest in the current work, can be learned (Baldwin, Andersson, Saffran, & Meyer, 2008; Brady & Oliva, 2008; Brady, Konkle, & Alvarez, 2009; Fiser & Aslin, 2001; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Turk-Browne, Jungé, & Scholl, 2005; Turk-Browne, Isola, Scholl, & Treat, 2008). To take an example from the listed works, Fiser and Aslin (Fiser & Aslin, 2001; Fiser & Aslin, 2002) demonstrated that participants can learn joint and conditional probabilities for sequences; this can be viewed as one item in the sequence probabilistically cuing the features of a subsequent item.

## **Chapter 4**

### **Probability Cueing in an Inefficient Search Task**

In this chapter, I describe four behavioral experiments conducted to investigate the effects of top-down, endogenous probability cues on visual search, the limitations of searchers' ability to benefit from such cues, and the extent of searchers' ability to infer the meaning of such cues from search arrays themselves when not informed of those meanings in advance. A brief discussion of the results of each experiment is provided directly following the relevant work, and more in-depth discussion of and conclusions from the entire series of experiments is provided at the end of the chapter.

#### **4.1 My Paradigm and Experiment 1**

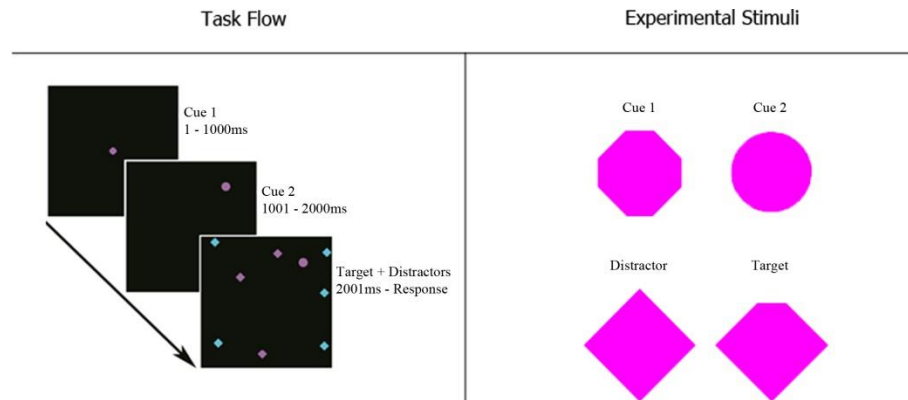
##### **4.1.1 Experimental Design**

To investigate how probabilistic cuing of features occurs, I measured participants' performance in a challenging visual search task where participants had to search for a diamond with one of its four corners missing and report which corner this was. This diamond was presented concurrently with varying numbers of distractor diamonds. The target diamond and distractor diamonds could take either of two colors (magenta or cyan), and preliminary cues provided statistical information about the likely color of the target diamond.

Assessing probabilistic cuing of features in a search task allowed us to determine if cuing speeded detection and discrimination. It also allowed us to evaluate if probability cuing changes search efficiency, and (indirectly) whether items were weighted for search in proportion to their probability for being the target.

Participants in all experiments were University of Waterloo undergraduate students, and numbers of participants per experiment were as follows: experiment 1 had 10 participants (one male, nine female); experiment 2 had 10 participants (six male, four female); experiment 3 had 18 participants (four male, fourteen female); experiment 4 had 60 (20 per between subjects condition) participants (25 male, 34 female, one undeclared) of which one was dropped for low accuracy (over 5 SD below the mean). The University of Waterloo Office of Research Ethics approved the research and informed consent was obtained from all participants.

Participants completed five blocks of 100 trials of a visual search task. As shown in the first panel of Figure 4-1, participants viewed two cues in sequence, each for 1000ms, followed by a search array. Panel 2 gives a detailed view of the experimental stimuli.

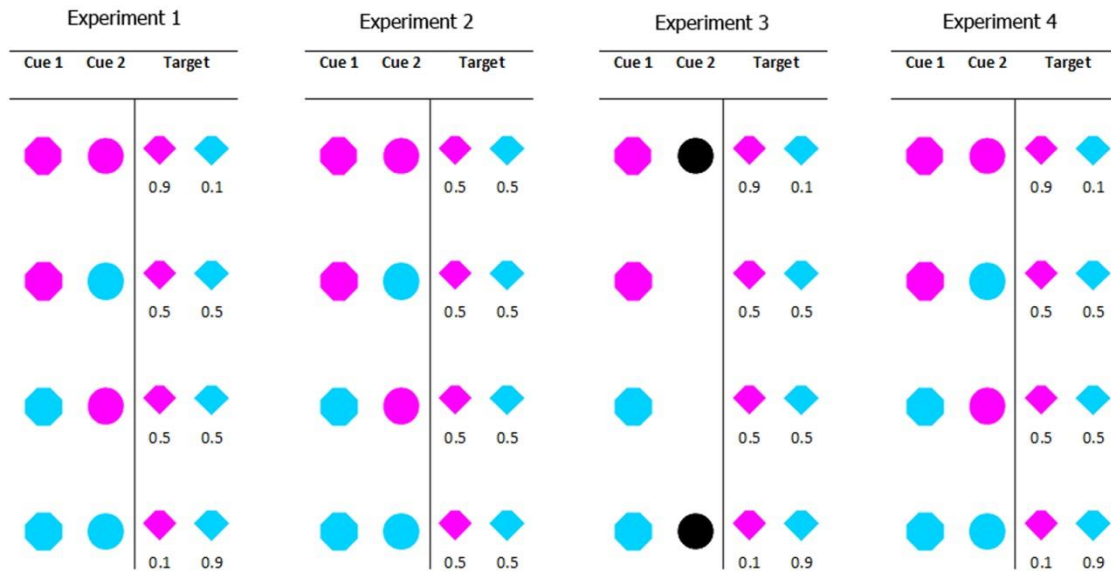


**Figure 4-1:** Experiment 1 task flow and magnified stimuli. Cue 1 appeared in the center of the screen for 1 second. After it disappeared, cue 2 appeared in a random location. After another 1-second interval, the search array appeared with the target and all distractors in random, non-overlapping locations. Possible stimuli colors were magenta and cyan. The size of every stimulus was 0.8 degrees of visual angle. Participants reported which of the four corners of the target was missing with the corresponding arrow key on the computer keyboard.

The first cue was always in the center of the screen, while the second cue appeared at a random location. During the search task, the second cue remained on the screen and  $N$  (8, 12, 16, or 20) items appeared on the screen, all in random locations. The items consisted of a single target, which was a diamond shape with one corner missing, and  $(N - 1)$  distractors, which were intact diamonds. These items remained until the participant indicated with a button press that they had located the target. After this button press, the search array disappeared and participants were prompted to indicate which corner of the target was missing by pressing the corresponding arrow key on a computer keyboard.

The absolute probability of a particular color target on any given trial was 0.5; however, the *conditional probability* – the likelihood of a particular target color given a particular cue combination – varied from 0.1 (when both cues predicted the non-target color) to 0.5 (when one cue predicted the target color and the other predicted the non-target color) to 0.9 (when both cues predicted the target color). For example, if both cues were magenta, the target was 90% likely to be magenta and 10% likely to be cyan; if one cue was cyan and the other cue was magenta, the target was 50% likely to be magenta and 50% likely to be cyan. The coloring of the target and distractors was selected on an item by item basis. This means that while the colors cyan and magenta were equally likely overall, on

individual trials the number and proportion of each color varied. Panel 1 in Figure 4-2 shows the full conditional probability distribution for the cues in experiment 1; the cues and their corresponding probability distributions for experiments 2 – 4 are displayed in panels 2 – 4.



**Figure 4-2:** All possible cue combinations and their associated target color probabilities for each experiment. In experiments 1 and 4, when both cues were the same color (cyan or magenta) the target was 90% likely to be that color. When the cues were different colors, they were uninformative as to the target color. In experiment 2, all cue combinations were uninformative. In experiment 3, the presence or absence (counterbalanced across participants) of cue 2 indicated the validity of cue 1; a valid cue 1 predicted target color with 90% accuracy.

Prior to beginning the task, the subjects were explicitly instructed that the color of the two cues was informative of the target. The exact wording used to communicate the relationship between cues and targets was, "...if both cues are red, the target is very likely to be red; if both cues are green, the target is very likely to be green; if the cues are each a different color, the target is equally likely to be red or green." Subjects were told that the second cue would appear in a random position on the screen, and that the location of the second cue was not predictive of any aspect of the search task.

In 50% of the total trials, the cues were different colors and thus uninformative (0.5 conditional probability). In the other 50% of trials, the cues were of the same color and were 90% predictive, resulting in the 0.9 conditional probability condition in 45% of the total trials, and the 0.1 conditional probability condition in 5% of the total trials.

Distractor color ratios were not fixed. Ratios were generated for each trial by assigning each distractor, one by one, a color. This color assignment was weighted by the proportion of colors

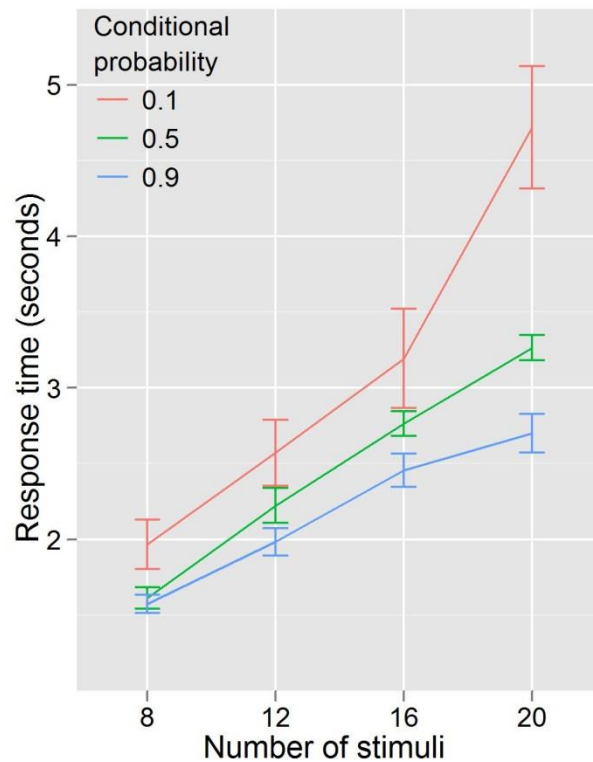
already assigned to the target and all distractors. The probability of a distractor receiving a certain color was equal to the proportion of stimuli (target + distractors) of the other color. So if there were 3 cyan stimuli and 4 magenta stimuli in an 8-item trial, the last stimulus had a probability of 4/7 to be cyan. This algorithm for generating distractor colors resulted in ratios that converged on 1:1 quite rapidly, and the ratio distribution across all trials was hyper-normal, centered on 1:1.

Accuracy in all experiments was near-perfect (98.7% overall; above 98% for each individual experiment) and did not differ between any conditions in any experiments; because of this uniformity, accuracy is analyzed only in Experiment 1, to illustrate the method used. RT Analyses used the RTs from correct trials only. As sub-300ms trials had chance levels (33%) of accuracy, I categorized these as accidental button presses and excluded them from RT analysis. Trials over 10 seconds were more than three standard deviations greater than the mean RT (in all experiments) and I excluded them from RT analysis as reflecting extended periods of off-task behavior.

#### **4.1.2 Apparatus**

Participants sat at a viewing distance of approximately 65 cm from a flat CRT monitor (36.5 cm × 27.5 cm viewable area, approximately 31° × 24° of visual angle computed at screen center) running at 85 Hz and at 800 × 600 resolution. All stimuli (cues, target, and distractors) subtended 0.8 degrees of visual angle and were presented on a black background.

### 4.1.3 Experiment 1 Results and Discussion



**Figure 4-3:** RTs by distractor number for experiment 1. The higher the conditional probability, the faster and more efficient the search.

Data in this experiment was analyzed by conducting a repeated measures ANOVA with RT as the dependent variable, and with conditional probability (3 levels: 0.1, 0.5, 0.9) and number of stimuli (4 levels: 8, 12, 16, 20) as factors. As shown in Figure 3, RT increased with number of stimuli,  $F(3, 27) = 96.29$ ,  $p < 0.001$ , and decreased with increases in conditional probability,  $F(2, 18) = 18.01$ ,  $p < 0.001$ . I also found a conditional probability by number of stimuli interaction,  $F(6, 54) = 7.451$ ,  $p < 0.001$ , indicating that as the number of stimuli increased, the effect of conditional probability became larger. To phrase this result more conventionally, the distractor number  $\times$  RT slope decreases as conditional probability increases. Distractor number  $\times$  RT slope is an index of visual search efficiency (Wolfe, Cave, & Franzel, 1989; Wolfe, 1994; Wolfe, 1998; Wolfe, 2007) and so this experiment provides evidence that not only are participants faster to locate and report targets with high conditional probability (and slower for targets with low conditional probability), but that this change in performance is accompanied by changes in the efficiency of search. That is, the greater the



number of stimuli to search, the greater the benefit of high conditional probability (or cost of low conditional probability).

Accuracy was analyzed as a function of the same factors (number of stimuli and conditional probability). There was no effect of number of stimuli,  $F(3, 27) = 0.893$ ,  $p > 0.05$ , no effect of conditional probability,  $F(2, 18) = 0.617$ ,  $p > 0.05$ , and no interaction between number of stimuli and conditional probability,  $F(6, 54) = 0.468$ ,  $p > 0.05$ .

This modulation of efficiency might result from subjects simply exhaustively searching the cued color first. I consider this unlikely, because no participants in this (or any subsequent) experiment reported using such a strategy in the post-questionnaire (in fact, no participants reported using *any* conscious search strategy). Nonetheless, I formally tested this possibility using linear models implemented in the R system for statistical computing (R Development Core Team, 2011) with the ‘lm’ function. I generated two linear models via simple linear regression on the data in the high conditional probability condition (comprised of all trials in which participants were predictively cued to a color *and that color actually was that color of the target on that trial*).

Model 1 takes into account only the number of cued-color stimuli on the screen (simulating the case in which participants search only the cued-color stimuli, which is the color of the target for this subset of trials), while model 2 takes into account both the number of cued-color stimuli and the number of uncued-color stimuli. If participants simply search the cued color first, then a model that takes into account the number of non-target colored distractors should not provide additional predictive value. Table 1 shows the intercepts and coefficients.

<b>Model 1: <math>rt = \alpha + \beta_1(\text{number cued-color stimuli})</math></b>					
	value	SE	df	t	significance
$\alpha$	1.042	0.086	2199	12.10	$p < 0.001$
$\beta_1$	0.163	0.012	2199	14.11	$p < 0.001$
<b>Model 2: <math>rt = \alpha + \beta_1(\text{number cued-color stimuli}) + \beta_2(\text{number uncued-color stimuli})</math></b>					
	value	SE	df	t	significance
$\alpha$	0.929	0.089	2198	10.448	$p < 0.001$
$\beta_1$	0.113	0.015	2198	7.310	$p < 0.001$
$\beta_2$	0.077	0.016	2198	4.773	$p < 0.001$

**Table 4-1:** Intercepts and coefficients for the models generated by simple linear regression on the data from the *0.9 conditional probability* condition in experiment 1.

Given that these models are constructed from a very specific subset of the data, the differences (or lack thereof) in coefficients, both within and between models, are not informative about the nature of my experimental results. It is the overall fit, or predictive value, of the models, measured by the  $R^2$  statistic, that is relevant to my question of whether the observed changes in search efficiency result from a shift to exhaustive search.

Model 1 has an  $R^2 = 0.083$ ,  $F(1, 2199) = 199.1$ ,  $p < 0.001$ . Model 2 has an  $R^2 = 0.092$ ,  $F(2, 2198) = 111.9$ ,  $p < 0.001$ . ANOVA verifies the improvement in fit from model 1 to model 2,  $F(1, 2198) = 22.781$ ,  $p < 0.001$ . Although the improvement in fit is objectively small, the strong correlation between the number of cued and uncued stimuli ( $r = .94$ ) leaves very little variance for the number of uncued-color distractors term to explain. That the addition of such a term, with such small explanatory potential, nonetheless results in an improvement in fit provides strong evidence that participants are not simply engaging in an exclusive, exhaustive search of cued-color items. If such a strategy were being employed, there should be no improvement in fit from the addition of the uncued-color term to the model, since in the high conditional probability condition participants *could have* located the target every time with an exhaustive search.

Although the second cue remains on screen during the search, the models described above do not consider it as a stimulus for two reasons. First, it is different in form from the true distractors; second, it is present on the screen for a full second before the search array is displayed, giving the participant time to recognize and categorize it as not part of the search array. Additionally, in the high conditional probability data, the second cue color always matches the target: so even if participants were responding to the second cue as a distractor during search, it would have no impact on the structure or fit of either model.

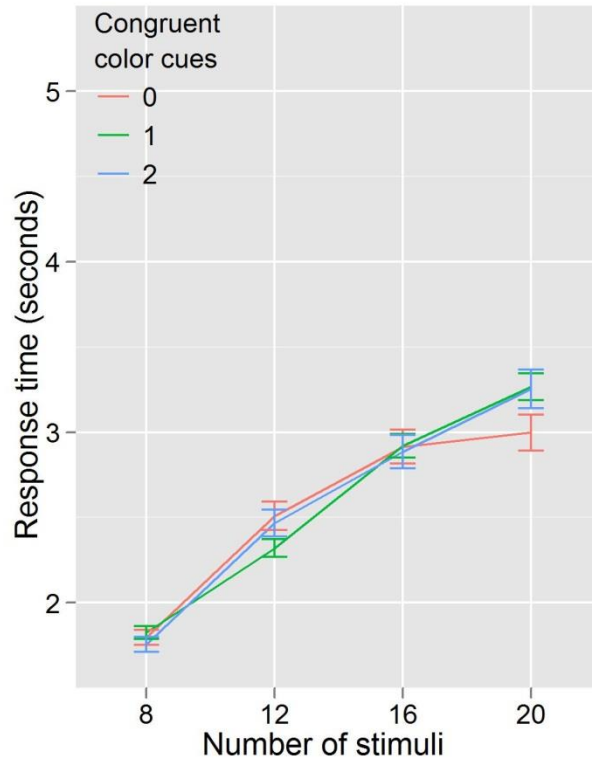
## **4.2 Experiment 2**

### **4.2.1 Experimental Design**

Experiment 1 does not separate conditional probability effects from the congruency or incongruency of the cued and target colors. When neither cue color was congruent with the target color, conditional probability was 0.1; when one cue color was congruent with the target color, conditional probability was 0.5; and when both cue colors were congruent with the target color, conditional probability was 0.9. Because of this, Experiment 1 by itself cannot confirm that the observed effect of cuing is related to conditional probability rather than color priming. Experiment 2 addresses this issue.

The method for Experiment 2 was identical to Experiment 1 with the exception that the cues were non-predictive. This preserved the color priming from Experiment 1, but dissociated it from conditional probability, which was 0.5 for every cue combination in this experiment. Panel 2 in Figure 2 illustrates the conditional probability distribution for this experiment.

#### 4.2.2 Results and Discussion



**Figure 4-4:** Response times by distractor number for Experiment 2. When probability cuing is removed, search efficiency is equal regardless of color priming.

Experiment 2 was analyzed by conducting a repeated measures ANOVA with RT as the dependent variable, and with congruent color cues (3 levels: 0, 1, 2) and number of stimuli (4 levels: 8, 12, 16, 20) as factors. As shown in Figure 4, there was an effect of number of stimuli,  $F(3, 27) = 115.587$ ,  $p < 0.001$ . However, there was no effect of number of congruent color cues  $F(2, 18) = 0.451$ ,  $p = 0.644$ , and no number of stimuli by congruent color cues interaction  $F(6, 54) = 1.433$ ,  $p = .219$ . This supports my hypothesis that it is conditional probability, not color priming, driving the effect on search efficiency. However, from the data in Experiments 1 and 2, I still cannot discount the possibility that the conditional probability information is necessary but not sufficient to drive the observed effects on

search efficiency; it could be that the combination of color priming and conditional probability is modulating search efficiency. Although this objection seems somewhat more unlikely than that raised with regard to color priming alone, I nonetheless seek to address it in Experiment 3.

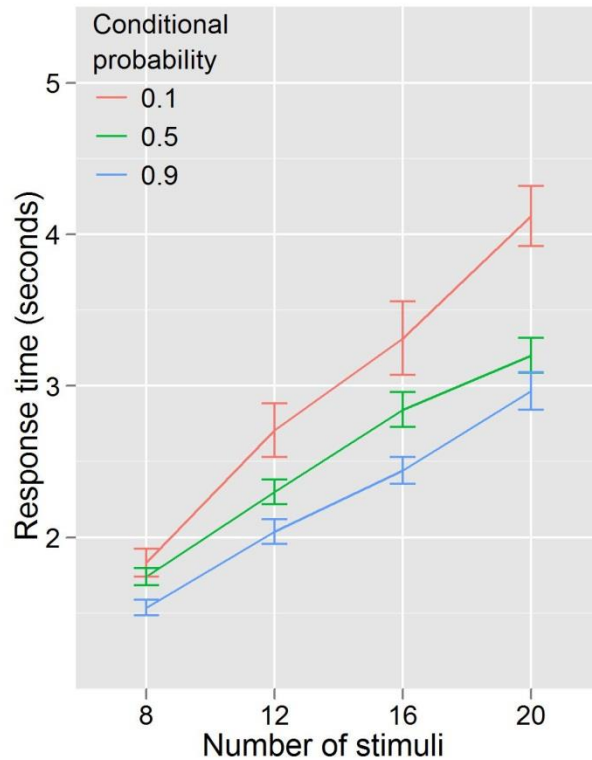
### **4.3 Experiment 3**

#### **4.3.1 Experimental Design**

Experiment 3 reinstated the manipulation of conditional probability, but the cues that delivered the conditional probability information to the participant were changed. In Experiments 1 and 2, the second cue was a circle with two possible states. State 1 (magenta) indicated an increased conditional probability of magenta, and state 2 (cyan) indicated an increased conditional probability of cyan. In Experiment 3, the second cue changed visually to a white circle that was either present (state 1) or absent (state 2). Panel 3 in Figure 2 shows the cues and conditional probability distribution for Experiment 3.

In this experiment, cue 2 no longer provided direct information about the probable target color; instead, it indicated whether the first cue was predictive or non-predictive. This change had two effects: it eliminated half of the color priming from the task, and it made the relationship between cues and target more complicated.

### 4.3.2 Results and Discussion



**Figure 4-5:** Response times by distractor number for Experiment 3. Search efficiency is modulated by conditional probability even when color priming is reduced.

Experiment 3 was analyzed by conducting a repeated measures ANOVA with RT as the dependent variable, and with conditional probability (3 levels: 0.1, 0.5, 0.9) and number of stimuli (4 levels: 8, 12, 16, 20) as factors. As seen in Figure 5, the pattern of effects was identical to Experiment 1; RT increased with number of stimuli,  $F(3, 51) = 158.34$ ,  $p < 0.001$ , and decreased with conditional probability,  $F(2, 34) = 18.239$ ,  $p < 0.001$ , and there was a conditional probability by number of stimuli interaction,  $F(6, 102) = 5.458$ ,  $p < 0.001$ , again showing that the effect of conditional probability was accompanied by changes in efficiency; however, unlike Experiment 1, changes in search efficiency were not observed between baseline conditional probability and high conditional probability. That is, when low conditional probability trials were excluded from the analysis, there was no interaction between number of stimuli and conditional probability,  $F(3, 51) = 1.160$ ,  $p = 0.330$ .

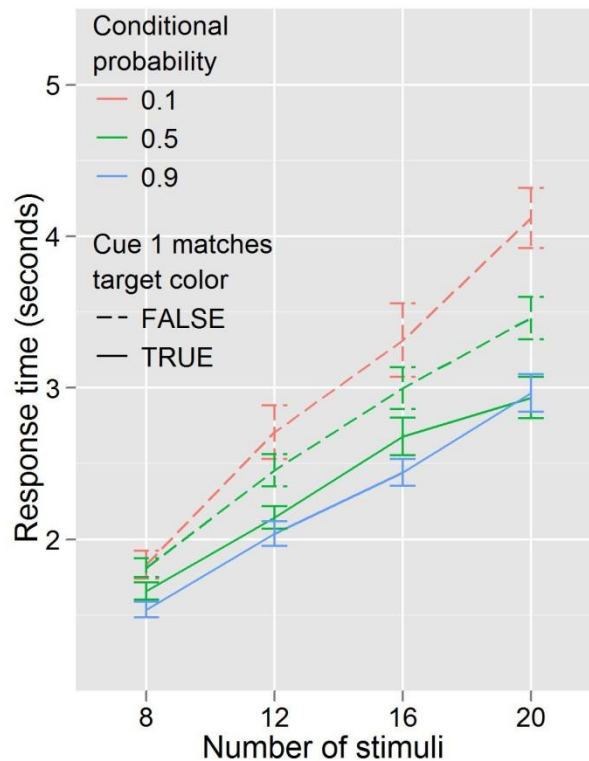
The perseverance of this effect when color priming is reduced, but not when conditional probability is removed, indicates that probability alone drives the observed effects on visual search efficiency.

I also replicate my linear modeling from Experiment 1 on the data from Experiment 3. The motivation and methods for this modeling are unchanged from Experiment 1; only the data used to construct the models differs. The findings from Experiment 1 are replicated in this analysis as well: model 1 (number of cued-color distractors as the only predictor) yields  $R^2 = 0.104$ ,  $F(1, 3910) = 454.5$ ,  $p < 0.001$ , while model 2 yields  $R^2 = 0.123$ ,  $F(2, 3909) = 274.2$ ,  $p < 0.001$ . Again, ANOVA confirms that the improvement in fit is significant,  $F(1, 3909) = 84.177$ ,  $p < 0.001$ . Table 2 shows the intercepts and coefficients for each model.

<b>Model 1: <math>rt = \alpha + \beta_1(\text{number cued-color stimuli})</math></b>					
	value	SE	df	t	significance
$\alpha$	0.878	0.068	3910	12.94	$p < 0.001$
$\beta_1$	0.195	0.009	3910	21.32	$p < 0.001$
<b>Model 2: <math>rt = \alpha + \beta_1(\text{number cued-color stimuli}) + \beta_2(\text{number uncued-color stimuli})</math></b>					
	value	SE	df	t	significance
$\alpha$	0.726	0.069	3909	10.492	$p < 0.001$
$\beta_1$	0.122	0.012	3909	10.105	$p < 0.001$
$\beta_2$	0.111	0.012	3909	9.175	$p < 0.001$

**Table 4-2:** Intercepts and coefficients for the models generated by simple linear regression on the data from the 0.9 conditional probability condition in Experiment 3.

In this experiment, the baseline condition is not a product of conflicting information from cue 1 and cue 2. Rather, one color or the other is cued, after which the searcher receives information about the validity of this information from the purely symbolic second cue. In contrast to Experiment 1, where it is unclear what the motivation or mechanism might be for selecting or acting based on one cue over the other, here it seems quite possible that searchers might disregard the actual validity of the second (symbolic) cue entirely, and search as if the first (color) cue was always valid. Instead, or in addition, there might be a degree of automaticity to the high conditional probability search behaviors which participants might be unable to completely suppress when predictive and non-predictive cues are interspersed. Figure 6 illustrates how performance differed based on this distinction.



**Figure 4-6:** Response times by distractor number for each of the three conditional probabilities in Experiment 3. Cue 1 never matched the target color when conditional probability was 0.1, always matched when conditional probability was 0.9, and matched half the time when conditional probability was 0.5.

Congruity of the first cue with the target color affects search behavior, even when that cue is indicated to be non-predictive. When conditional probability is held constant (at 0.5) participants are faster to respond when the target color matches the color of the (non-predictive) first cue,  $F(1, 17) = 24.834$ ,  $p < 0.001$ , and trend towards being more efficient,  $F(3, 51) = 2.630$ ,  $p = 0.060$ . When cue 1 color matches target color, search trends toward being faster for high conditional probability than for the baseline conditional probability,  $F(1, 17) = 3.495$ ,  $p = 0.079$ , but there is little evidence that it becomes more efficient,  $F(3, 51) = 1.301$ ,  $p = 0.284$ ; when cue 1 color does not match target color, search is slower,  $F(1, 17) = 5.558$ ,  $p < 0.05$ , and less efficient,  $F(3, 51) = 3.050$ ,  $p < 0.05$ , for low conditional probability compared to the baseline conditional probability.

Both possibilities outlined above would be expected to produce these effects: fast RTs for high conditional probability, slow RTs for low conditional probability, and a mix of fast and slow RTs in the baseline condition, depending on the validity of the (non-predictive) color cue. Within this paradigm, it is difficult to differentiate the perception-driven explanation (simplification of the cuing)

and the action-driven explanation (inability to suppress automatic search behaviors). However, this distinction, while intriguing, is of secondary interest in this experiment.

The goal of Experiment 3 was to explore the relationship between color priming and probability cuing, and to disambiguate the effects of each. There is an effect of cue 1 color validity; there is also an independent effect of the second, non-color cue. This effect of the pure probability cue, even when holding constant the congruity of the colored cue and the target color, provides strong evidence that the differences in search RTs observed here and in Experiment 1 are driven by probability and not reliant on color priming.

## **4.4 Experiment 4**

### **4.4.1 Experimental Design**

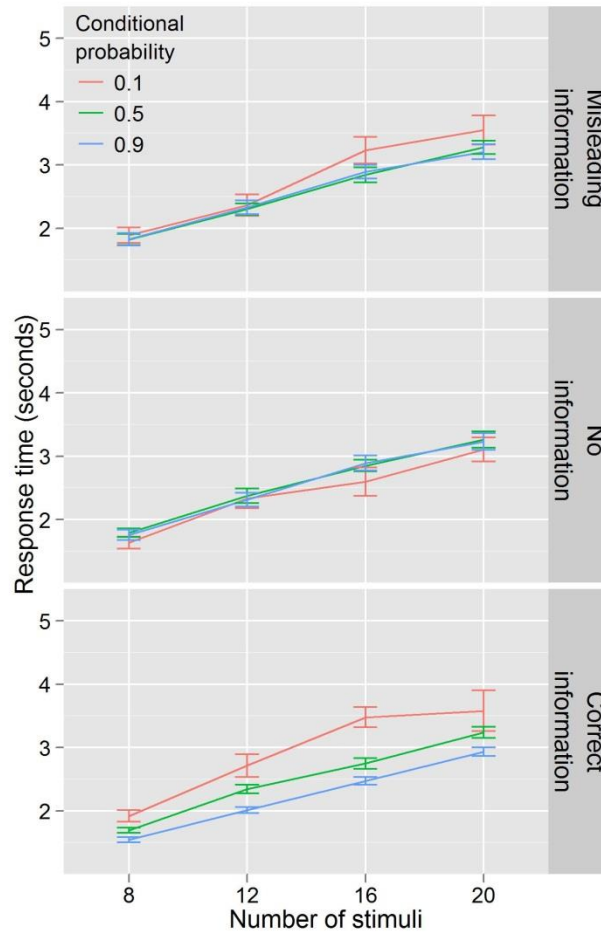
Having established that conditional probability modulates visual search efficiency and having excluded an exhaustive search of the cued color as the explanation for this change in search efficiency, there are several remaining questions. In Experiments 1-3, participants were always explicitly informed of the conditional probabilities. To what extent is explicit knowledge of conditional probability information required for it to be utilized in search? Conversely, how much knowledge of conditional probability can be learned or deduced through the act of searching? Experiment 4 investigates these questions by varying amount and quality of information provided to participants about the conditional probabilities in the task.

Experiment 4 expands the design from Experiments 1 – 3 to a 3 (conditional probability) x 4 (number of stimuli) x 3 (information quality) design. There were 3 levels (between subjects) of the information quality condition: full information, no information, and misleading information. All participants performed the same search task as in Experiment 1; panel 4 in Figure 2 shows the conditional probability distribution.

Although the task itself was the same for all three between-subjects conditions, participants in each condition received different instructions about the relationship between cues and target color. Participants in the full information condition received a full description of the relationship between cues and target color (this condition was an exact replication of Experiment 1, and participants received identical information and instructions). Participants in the no information condition were not informed of the relationship between cues and target color. Participants in the misleading information condition were told explicitly that there was no relationship between the cues and target color.



#### 4.4.2 Results and Discussion



**Figure 4-7:** Response times by distractor number for Experiment 4. Conditional probability modulates search efficiency only when participants have explicit knowledge.

Figure 7 illustrates the results of Experiment 4. Data in this experiment was analyzed by conducting an ANOVA with reaction time (RT) as the dependent variable, conditional probability (3 levels: 0.1, 0.5, 0.9) and number of stimuli (4 levels: 8, 12, 16, 20) as within-subjects factors, and information condition (3 levels: full information, no information, misleading information) as a between-subjects factor. Again, response time increased with number of stimuli,  $F(3, 167) = 398.159$ ,  $p < 0.001$ . There was a main effect of conditional probability  $F(2, 110) = 10.866$ ,  $p < 0.001$  and this effect differed across information conditions  $F(4, 110) = 9.236$ ,  $p < 0.001$ . There was no effect of conditional probability in the no information condition  $F(2, 34) = 1.751$ ,  $p = 0.188$ . In the misleading information condition, there was an effect of conditional probability  $F(2, 38) = 3.543$ ,  $p < 0.05$ , but

only for the low conditional probability targets, which participants were slower to locate. However, in the correct information condition, the pattern from Experiment 1 was repeated, with faster search RTs for high conditional probability and slower RTs for low conditional probability,  $F(2, 38) = 17.429$ ,  $p < 0.001$ . Evidence for an efficiency component to these differences was present in this condition as well, with a trending number of stimuli by conditional probability interaction,  $F(6, 114) = 2.019$ ,  $p = 0.069$ .

RT benefits from learning the probabilities in the task would be expected to take time to emerge in the uninformed and misinformed conditions. Although there was an effect of block in both conditions,  $F(4, 72) = 3.148$ ,  $p < 0.05$  for uninformed and  $F(4, 72) = 2.588$ ,  $p < 0.05$  for misinformed, there was no block by conditional probability interaction in either condition,  $F(8, 136) = 1.474$ ,  $p = 0.172$  (uninformed) and  $F(8, 144) = 1.320$ ,  $p = 0.238$  (misinformed). On this basis, I conclude that there was no learning of the probability information in either condition for this task.

Since the full information condition is an exact replication of Experiment 1, I repeat my linear modeling analysis from Experiments 1 and 3 on that data. As in those experiments, I generate a model 1 which includes only a number of cued-color stimuli term, and a model 2 which includes terms from both number of cued-color stimuli and number of uncued-color stimuli. Model 1 yields  $R^2 = 0.107$ ,  $F(1, 4430) = 527.9$ ,  $p < 0.001$ , while model 2 yields  $R^2 = 0.127$ ,  $F(2, 4429) = 322.8$ ,  $p < 0.001$ . ANOVA confirms that the improvement in fit is significant,  $F(1, 4429) = 105.36$ ,  $p < 0.001$ . Table 3 details the intercepts and coefficients for each model.

<b>Model 1: <math>rt = \alpha + \beta_1(\text{number cued-color stimuli})</math></b>					
	value	SE	df	t	significance
$\alpha$	0.891	0.062	4430	14.28	$p < 0.001$
$\beta_1$	0.193	0.008	4430	22.98	$p < 0.001$
<b>Model 2: <math>rt = \alpha + \beta_1(\text{number cued-color stimuli}) + \beta_2(\text{number uncued-color stimuli})</math></b>					
	value	SE	df	t	significance
$\alpha$	0.749	0.089	4429	11.84	$p < 0.001$
$\beta_1$	0.115	0.011	4429	10.30	$p < 0.001$
$\beta_2$	0.114	0.011	4429	10.26	$p < 0.001$

**Table 4-3:** Intercepts and coefficients for the models generated by simple linear regression on the data from the *full information, 0.9 conditional probability* condition in Experiment 4.

It is notable that, in the misleading information condition, participants are slower to locate and report low conditional probability targets, but no faster to report higher conditional probability targets (both compared to the baseline conditional probability condition). Not only did participants perform differently in response to differences in conditional probability that they had been explicitly told did not exist, but the reaction time cost for low conditional probability targets produced no corresponding benefit for high conditional probability targets.

I believe it likely that instead of providing evidence against the existence of conditional probabilities in the task (as they were meant to), my explicit instructions to participants that cue colors had no relation to target colors paradoxically primed them to explore or perceive exactly such a relationship. Two observations motivate this assertion. First, several participants were suspicious of the spontaneous instruction that there was no relationship between cues and target color (though no participants reported entertaining or acting upon such suspicions at debriefing). Such suspicions about the nature of the task may have led them to engage in exploratory behavior, which could result in the atypical ‘cost without benefit’ pattern of results in the condition as a whole.

Second, my (null) results in the no information condition suggest that participants are at floor performance for learning the statistics of the task when told nothing about those statistics. Because one could suspect that the implicit learning of probability relations might be subtle, I doubled the number of participants for the three conditions of this experiment over what I had used for Experiments 1 and 2. This gave us greater power to detect significant differences in Exp 4, and makes the negative results relatively more secure. Changes in conditional probability effects between the no information and misleading information conditions correspond to changes in learning and/or acting upon the statistics of the task. Since participants are at floor performance when given no information, such changes must logically be in the positive direction. On the basis of this evidence, I conclude that my misleading information condition actually functioned as a cue that there was a relationship between the cues and target color to be discovered. If a fourth condition were to be implemented with exactly this ‘partial’ information, I would expect to see results similar to those in the misleading information condition.

Overall, these results suggest that explicit knowledge of the probability relationships between cues and target color is necessary for conditional probability to facilitate search performance at the time scale (about 40 minutes) and number of trials used here. These data do not directly address whether this information would be learned implicitly if more time or trials were allowed, although I

speculate, based on results from the misleading information condition, that such implicit learning is possible.

## **Chapter 5**

### **Eye Tracking in my Task**

In this chapter, I describe Experiment 5, in which I employed eye tracking measures in addition to the reaction time measures of Experiments 1 - 4. Eye tracking was used to investigate specific, focused questions about the mechanisms of conditional probability in search demonstrated in chapter 4. I briefly report the reaction time data, after which I describe in detail the eye tracking data, and analysis. Finally, I provide a discussion of the broader implications of the results reported in this chapter.

#### **5.1 Experiment 5**

##### **5.1.1 Experimental Design**

Experiment 5 repeats the procedure from Experiment 1 with the addition of eye tracking data collection and intermittent calibration checks. Eye tracking data was collected with a Mirametrix S2 eye tracker operating at 60hz and connected to a laptop running Windows XP, which hosted the software necessary to run the S2. This laptop was connected to the Linux machine on which the experimental protocol was executed via crossover cable. Head position was controlled using a chin rest. A short test of the sustained quality of the initial calibration was conducted after each block of trials, in which participants repeated the initial calibration procedure for the S2 in a random order. Apart from these additions, the method and apparatus used was identical to the apparatus for Experiment 1.

25 University of Waterloo undergraduate students participated in this experiment, 5 male and 20 female. One participant was dropped from the analysis because the eye tracker was unable to maintain calibration on the eyes. The University of Waterloo Office of Research Ethics approved the research and informed consent was obtained from all participants.

##### **5.1.2 Eye Tracking Data Collection, Pre-Processing, and Transformations**

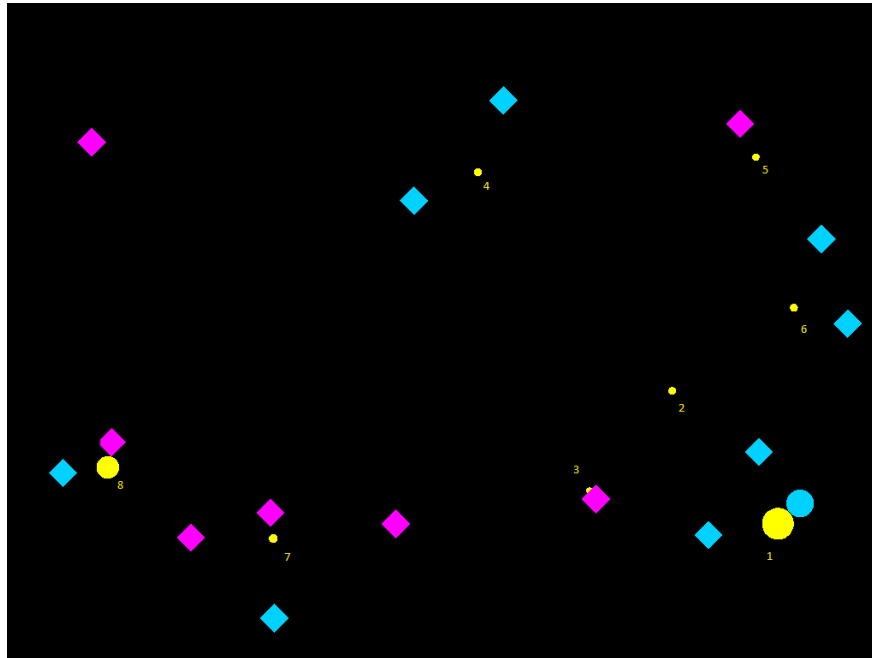
Before beginning data collection, it was necessary to create a communication protocol for the Windows machine hosting the S2, which as of this writing has drivers only for Windows operating systems, and the experimental machines, which run Linux. The communication protocol was written in Python, and functions by first establishing a connection to the S2 host machine over a TCP/IP

socket, then then querying that connection repeatedly (at greater than 60hz) to request and process each data packet (consisting of paired XML tags and string values) that the S2 outputs during operation. Data packet loss using this communication protocol is minimal (approximately one packet lost per million packets received).

The S2 uses a three dimensional model of the eye to track eye and gaze position. For each subject, this model of the eye is constructed by the S2 during a brief calibration procedure, after which gaze position is accurate within one degree of visual angle for almost all participants. Output from the S2 consists of time stamped (by its own internal clock, with 0.00s representing the most recent calibration) x-y coordinates for each eye; the tracker also groups these data into saccades and fixations using its own algorithm.

The fixation and saccade classification of the S2 data suite were often inconsistent with visual inspection of the raw (x,y) data. To address this deficiency, I used the Python programming language to implement an alternate method of classifying saccades and fixations. First, to correct for noise in the measurement of eye position, the data was transformed by replacing the x- and y-coordinates of each observation with the median value of that observation and the two preceding and the two following observations. Second, velocity was calculated for each observation and observations for which velocity exceeded 100.00 degrees of visual angle per second were classified as saccades; the remaining values were classified as fixations. Fixation coordinates were calculated by taking the averages of the x and y coordinates of each raw data point; fixation duration was a trivial calculation of the difference between the timestamp of the last raw data point in that fixation, minus the timestamp of the last raw data point preceding that fixation (ie., the last data point in the previous saccade). Third, fixations which were very close (less than 1 degree of visual angle) in position were combined, and fixations of very short duration (less than 60ms) were eliminated. Several other post-processing measures were taken on the data to correct minor issues, including an algorithm to correct for the infrequent packet loss, as well as a transformation from the coordinate system used by the S2 (normalized Cartesian coordinates ranging from 0 to 1 on both the x and y axes) to the non-normalized (to reflect the difference in the height/width dimensions of the monitor) Cartesian coordinates ranging from  $-W/2$  to  $W/2$  on the x-axis, and  $-H/2$  to  $H/2$  on the y-axis, with W and H representing the width and height, respectively, of the monitor in degrees of visual angle. More details on these data processing methods, along with sample code, can be found in Appendix A.

I used several preliminary methods to verify these saccade/fixation classifications. Figure 5-1 shows a typical trial visualized using a fixation visualization program that I designed in Python.

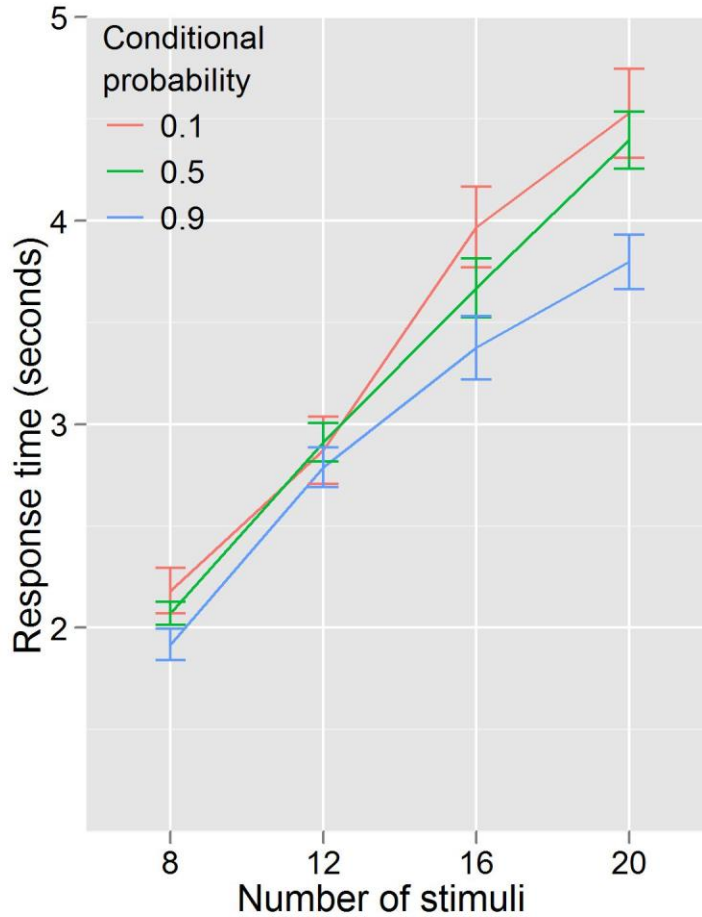


**Figure 5-1:** Fixations, represented by yellow dots, in a typical trial. Duration is represented by the radius of the circle, and each fixation is numbered to show its order within the trial.

This visualization program was used to verify that the coordinates of fixations classified by my post-processing algorithms were consistent with the coordinates for each component of the raw data, as well as roughly consistent with what might be expected in a search task (fixations reasonably near stimuli, few fixations to empty areas of the screen). The visualization program was also valuable in verifying fixation classifications on some of the calibration checks conducted after each block, since participants were instructed precisely what to look at during these checks and thus had more concrete values for comparison of calculated fixations.

### 5.1.3 Reaction Time Results

Reaction time results from Experiment 1 were replicated in this experiment, as shown in Figure 5-2. RT increased with number of stimuli,  $F(3, 69) = 171.63$ ,  $p < 0.001$ , and decreased with conditional probability,  $F(2, 46) = 11.09$ ,  $p < 0.001$ . The conditional probability by number of stimuli interaction,  $F(6, 54) = 7.451$ ,  $p < 0.001$ , was present as well, replicating the modulation of search efficiency found in Experiment 1.



**Figure 5-2:** Reaction times in Experiment 5. As in Experiment 1, search efficiency increased with conditional probability.

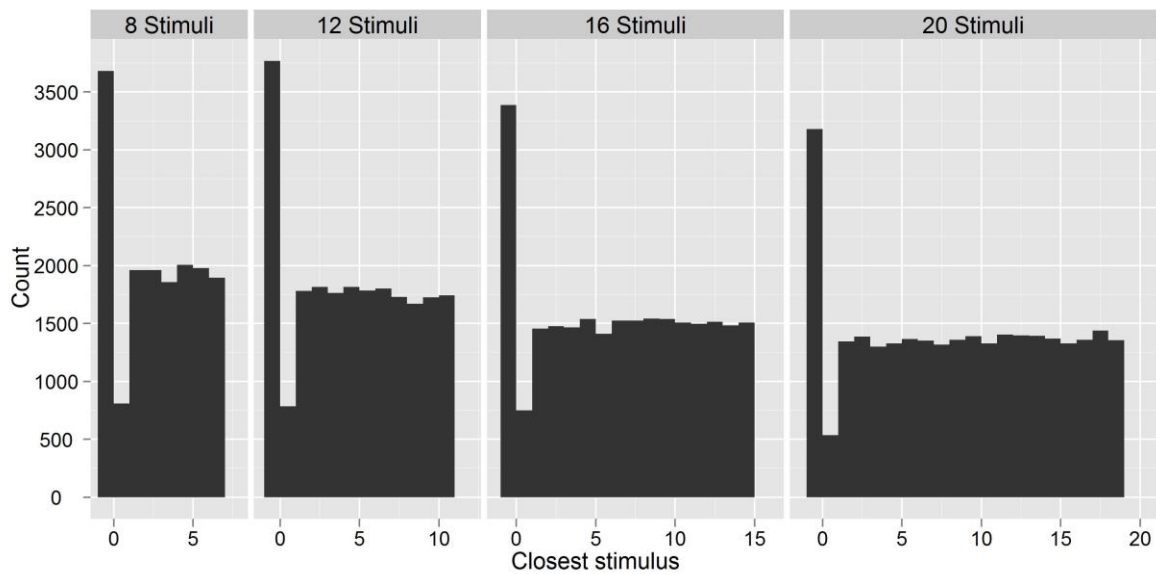
#### 5.1.4 Eye Tracking Results

Reaction time analysis allows us to examine the overall speed and efficiency of search; eye tracking analyses give us the tools to investigate the mechanisms of such speed and efficiency effects. In this section, I describe the results of several lines of investigation into the eye movements in this task.

There were several basic eye tracking results that I checked for to confirm the validity of my measures. In a search task like this one, it would be expected that participants would fixate on or near all distractors at an equal rate, and would fixate on or near the target at a greater rate. Figure 5-3



shows all fixations categorized by the nearest stimulus.



**Figure 5-3:** Closest numbered stimulus for each fixation, categorized by number of stimuli on screen. Stimulus 0 is the target; stimulus 1 is the second cue (which stays on screen for the duration of the search).

Stimulus 0 is the target; as expected, it is fixated more than other stimuli. Stimulus 1 is the second cue which remained on screen for the entirety of each trial. It is fixated at 30-40% the rate of the distractor stimuli. This is likely due to its unique shape (circle instead of diamond) making it discriminable without foveation. Because of this drastically reduced rate of fixation, the second cue is not considered a stimulus/distractor in the analyses that follow.

I also expected participants to preferentially fixate on areas of the screen containing more stimuli, and I expected fixations near more stimuli to have a longer duration. For both of these analyses, a measurement of the proximity or density of stimuli to a fixation is required.

Two such measurements were used to analyze the data. The first was a count of the number of stimuli within 3 degrees of visual angle of the fixation. Though simplistic, this count gave a rough idea of the number of stimuli included in each fixation. However, it ignored both the local density of stimuli (how close/far each stimuli was to the center of the 3 degree radius) and the relative density outside of, but still near, the 3 degree radius (ie., the presence or absence of nearby stimuli that could be reached with a short saccade).

The second measure addressed these issues. For each fixation, a score  $S$  was computed for each fixation by the following function, where  $d_{stim}$  represents the distance to a particular stimulus:

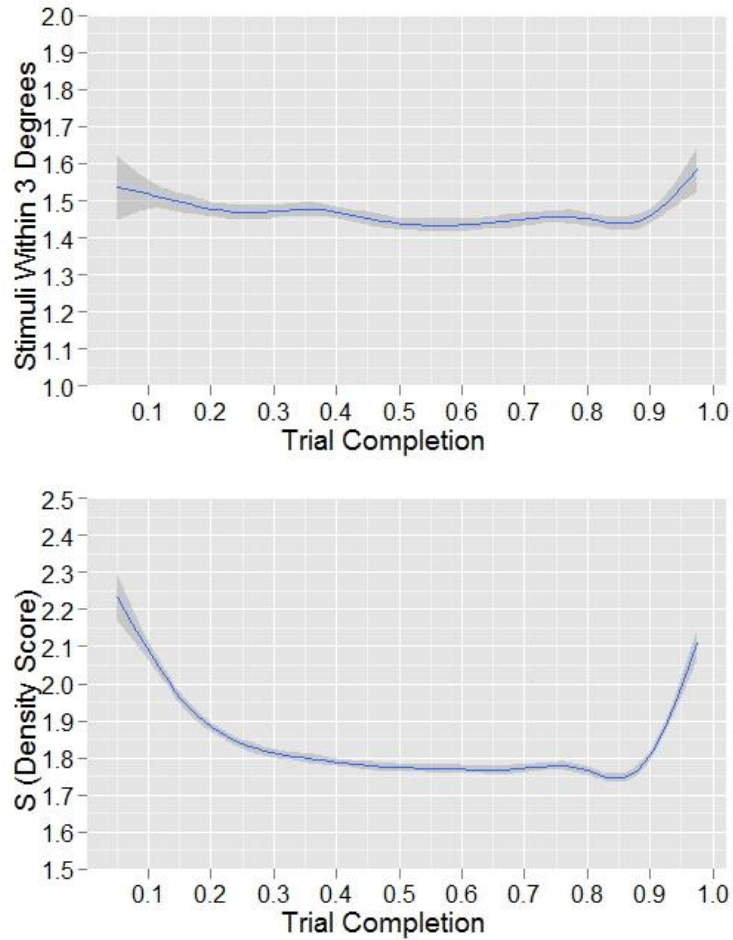
$$S = \sum_{stim} \frac{1}{1 + d_{stim}}$$

For both these analyses, the first and last fixations in a trial were excluded. The first fixation simply recorded the spot where the eyes were fixated when the search array came onscreen. While this data was useful in validating the calibration of the eye tracker, (since I would expect that participants would fixate on the second cue before each search) it contained no information about the search itself. The final fixation occurred after the motor response to end the trial had been initiated, and for this reason was often both unrepresentative of the search behavior and artificially short in duration (since it was terminated when the trial ended with the button press).

Relative fixation order within trials was used as an index of preferential looking; that is, I operate under the assumption that participants directed their gaze to highly salient or highly probable areas on the screen before directing their gaze to less salient or probable areas. There are two measures which capture this relation: time from trial start, and fixation number within the trial. I chose to use fixation number as a measure of progression through trials, because it provides a series of intuitively discrete events, whereas sampling gaze position at selected time points is highly arbitrary, since there is no clearly correct frequency for sampling.

Total number of fixations was quite variable from trial to trial. Since the third fixation on a trial with three total fixations is obviously qualitatively different from the third fixation on a trial with nine total fixations, I normalized this measure of relative order within trials by calculating  $trial\ completion = \frac{fixation\ number}{total\ fixations}$ . Trial completion gives a measure of the relative within-trial temporal position of fixations on a 0.0 – 1.0 scale and allows direct comparison of fixations that come from trials with differing total numbers of fixations (as in the above example with a three fixation trial and an eight fixation trial).

As expected, participants preferentially fixated high density areas of the screen: Figure 5-4 shows the relation of trial completion to both number of stimuli within 3 degrees (top) and density score (bottom).



**Figure 5-4:** Number of stimuli within 3 degrees of visual angle (top) and density score (bottom) by trial completion. The shaded area shows standard error.

These data were analyzed using two simple linear models: Model 1 was *Stimuli Within 3 Degrees ~ Trial Completion* and Model 2 was *Density Score ~ Trial Completion*. Both models had statistically significant coefficients for their respective index of stimulus density. Model 1 yielded  $R^2 = 0.0000769$ ,  $F(1, 70905) = 6.453$ ,  $p < 0.05$ . Model 2 yielded  $R^2 = 0.00312$ ,  $F(1, 70905) = 222.9$ ,  $p < 0.001$ . Table 5-1 lists the intercepts and coefficients for each model.

<b>Model 1: Duration = <math>\alpha + \beta_1(\text{Stimuli Within 3 Degrees})</math></b>					
	value	SE	df	t	significance
$\alpha$	1.479	0.009	70905	172.13	$p < 0.001$

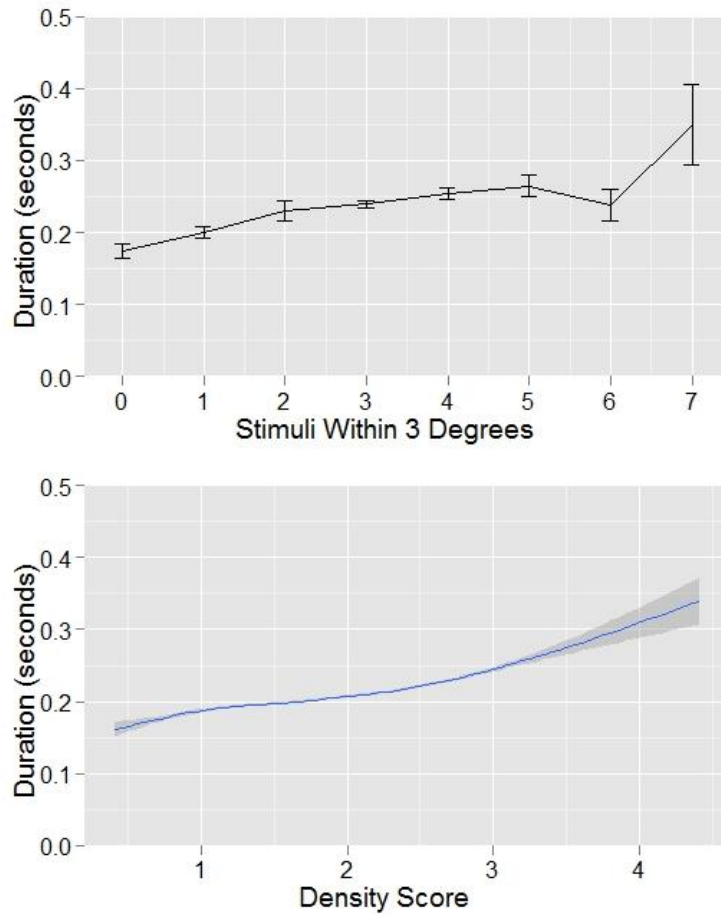
$\beta_1$	-0.037	0.014	70905	-2.54	$p < 0.05$
<b>Model 2: <math>rt = \alpha + \beta_1(\text{Density Score})</math></b>					
	value	SE	df	T	significance
$\alpha$	-1.880	0.005	70905	345.85	$p < 0.001$
$\beta_1$	-0.136	0.009	70905	-14.93	$p < 0.001$

**Table 5-1:** Intercepts and coefficients for the *Stimuli Within 3 Degrees ~ Trial Completion* and *Density Score ~ Trial Completion* models.

There are several points to be made concerning the two models. First, both show significant, effects, establishing that the Density Score measure is a valid one. Second, the coefficient,  $R^2$  value, and p-values in Model 2 exceed those in Model 1 by several orders of magnitude, suggesting that simply counting the number of stimuli within three degrees of visual angle is an imperfect measure. Comparing the two models directly (as I compared linear models in Chapter 4) is non-trivial, because unlike the previous linear models, these are not nested; given that preferential looking to high density areas of the screen in this task is confirmed, I see no need to make the direct comparison here.

Third and finally, both linear smooths in Figure 5-4 show a marked upward trend in their respective density measure during the last 10% of trial completion. Though this upwards trend is not sufficient to cancel out the net negative direction of the linear fit, it does suggest that looking behavior changes during this last portion of the trial. I would speculate that this corresponds to a direct fixation on the target, which would increase the density score by minimizing that particular distance (between fixation and target). It also might increase the average number of stimuli within three degrees simply by guaranteeing the presence of at least one stimulus (the target) within that radius, whereas other fixations throughout the trial might have a greater likelihood of containing no stimuli within that radius.

Also as expected, high density fixations were of greater duration. Figure 5-5 shows fixation duration as a function of stimuli within 3 degrees (top) and density score (bottom).



**Figure 5-5:** Fixation duration by number of stimuli within 3 degrees (top) and density score (bottom).

Since there were a fixed number of discrete values for Stimuli Within 3 Degrees, the data shown in the top panel of Figure 5-5 was analyzed using one-way ANOVA. Number of Stimuli Within 3 Degrees increased fixation duration,  $F(7, 127) = 16.837$ ,  $p < 0.001$ . Linear modeling of  $Duration \sim Density\ Score$  (the relation shown in the bottom panel of Figure 5-5) confirms the positive relation between Density Score and duration.  $R^2$  for the model was 0.016,  $F(1, 70905) = 1124$ ,  $p < 0.001$ ; Table 5-2 lists the intercept and coefficient.

<b>Model 1: Duration = <math>\alpha + \beta_1(\text{Stimuli Within 3 Degrees})</math></b>					
	value	SE	df	t	significance
$\alpha$	1.479	0.009	70905	172.13	$p < 0.001$
$\beta_1$	-0.037	0.014	70905	-2.54	$p < 0.05$

**Table 5-2:** Intercept and coefficient for the  $Duration \sim Density\ Score$  model.

So far, I have established that participants in my task preferentially fixate on areas of the display with a higher density of stimuli, and maintain fixation longer the more stimuli there are in the vicinity. I have also confirmed that my processing of the raw eye tracking data yields valid data, and that my density score for each fixation is a valid measure of stimulus density at a particular point on the display. With these basic analyses complete, I address the motivating question of this chapter: how does conditional probability affect eye movements in search?

Having established that high conditional probabilities result in benefits in search efficiency, a simple question follows: are such benefits the result of spending less time looking at low probability stimuli, or (to an extent) avoiding looking at low probability stimuli at all, and instead fixating on or near higher probability stimuli.

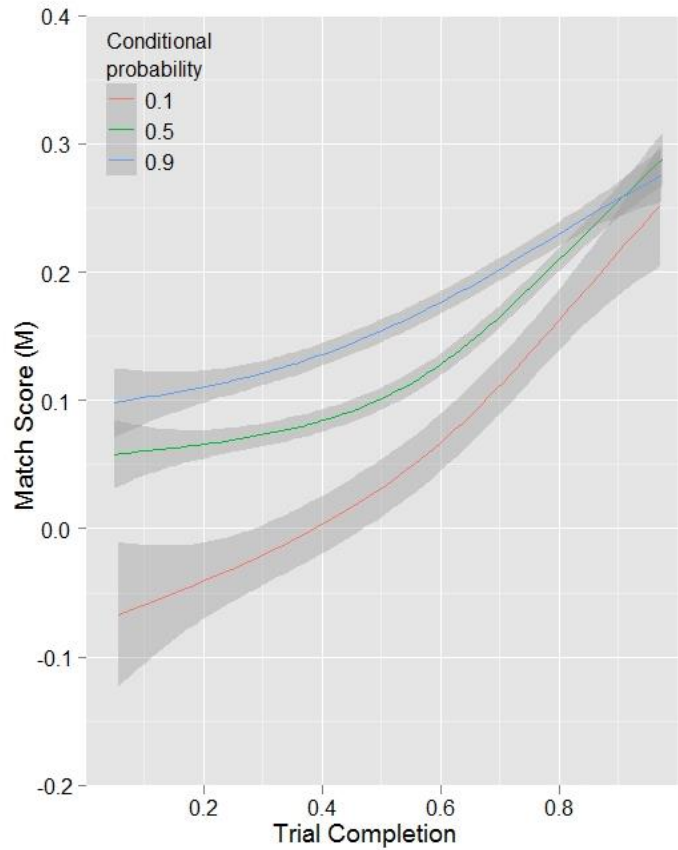
Average fixation duration did not differ between conditional probability conditions,  $F(2, 46) = 1.168, p = 0.320$ , suggesting that the latter possibility, that searchers preferentially fixated higher probability stimuli, is correct. This is particularly interesting in light of my earlier linear modeling results, which showed that participants were not simply ignoring the low probability stimuli and exclusively searching the higher probability stimuli. It seems, rather, that stimuli are weighted (at least in part) by their probability.

While preferential looking at high probability stimuli can be inferred from the fixation duration analysis above, eye tracking allows us to directly examine the phenomenon as well. For this purpose, I calculated a match score,  $M$ , which measured the degree to which stimuli around a fixation matched the color of the target on that trial.

$$M = \sum_{\substack{\text{target} \\ \text{color} \\ \text{stims}}} \frac{1}{1 + d_{stim}} - \sum_{\substack{\text{other} \\ \text{color} \\ \text{stims}}} \frac{1}{1 + d_{stim}}$$

The calculation for match scores was similar to the one used to compute density scores, with the exception that stimuli whose color did not match the target color were assigned a negative score; this generates a scale with negative values representing predominantly non-target color stimuli around a fixation, and positive values representing predominantly target color matching stimuli around a fixation. Greater magnitude indicates greater degree of matching/non-matching stimuli around the

fixation. Assigning positive/negative scores for matching/non-matching stimuli normalizes the measure across displays which differ in number of stimuli, allowing direct comparison for all data. Figure 5-6 shows a linear smooth of match score as a function of trial completion for each conditional probability.



**Figure 5-6:** Match score by trial completion for each level of conditional probability. Searchers look preferentially towards the areas of the display containing stimuli of the cued color. The cued color matches the target color when conditional probability is higher, and does not match when conditional probability is low.

Match score was modeled as a function of trial completion and conditional probability using simple linear regression.  $R^2$  was 0.017,  $F(5, 70901) = 244.3$ ,  $p < 0.001$ . Table 5-3 lists the intercept and coefficients for the model.

<b>Match Score ~ Trial Completion * Conditional Probability</b>					
	value	SE	df	t	significance
$\alpha$ (intercept)	-0.134	0.019	70901	-7.179	$p < 0.001$

$\beta_1$ (Trial Completion)	0.366	0.031	70901	11.624	$p < 0.001$
$\beta_2$ (0.5 Conditional Probability)	0.124	0.020	70901	6.306	$p < 0.001$
$\beta_3$ (0.9 Conditional Probability)	0.192	0.020	70901	9.542	$p < 0.001$
$\beta_4$ (Trial Completion : 0.5 Conditional Probability)	-0.102	0.033	70901	-3.054	$p < 0.01$
$\beta_5$ (Trial Completion : 0.9 Conditional Probability)	-0.156	0.034	70901	-4.591	$p < 0.001$

**Table 5-3:** Intercept and coefficients of the Match Score ~ Trial Completion \* Conditional probability linear model.

In addition to the tests of the coefficients in isolation, I also used the linear hypothesis function from the car (companion to **applied regression**) package in R to test the null hypotheses that  $\beta_2 = \beta_3$  and  $\beta_4 = \beta_5$ . Both null hypotheses were rejected, with  $F(1, 70901) = 45.711$ ,  $p < 0.001$ , and  $F(1, 70901) = 10.407$ ,  $p < 0.01$ , respectively. This additional analysis establishes that not only are all factors in the model different from the base condition (0.1 conditional probability), but 0.9 conditional probability has a differential effect than 0.5 conditional probability both in isolation ( $\beta_3 > \beta_2$ ) and in combination with trial completion ( $\beta_5 < \beta_4$ ).

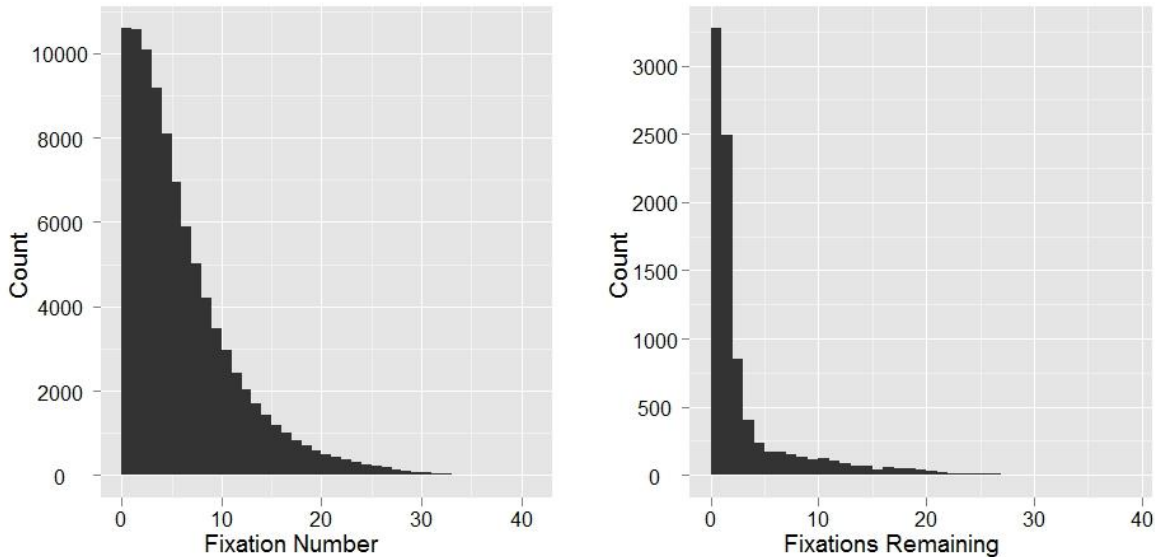
It is notable that the scores for the 0.5 conditional probability condition begin above the neutral 0-score that one might expect when no information on the color of the target is available. This asymmetry is due to the procedure used to generate the distractor arrays. Distractors were generated using an algorithm which quickly converged on a uniform distribution. However, these distractors were generated independently of the target, which resulted in, on average,  $n$  non-target colored items and  $n+1$  target colored items in each search trial.

As suggested by the pattern of results in Figure 5-6 and confirmed by the above analyses, searchers begin the task biased to look to areas of the display containing the cued color, and over the course of the trial increasingly (and additively with their initial bias) fixate the true target color. The magnitude of this increase in target color fixations is proportional to the magnitude of the initial bias towards the cued color, with trials in which participants were cued to the non-target color showing the most dramatic shift, uncued color trials showing a less dramatic shift, and correctly cued trials showing only a slight shift.

It is not enough for searchers to quickly foveate the target stimulus; they also must correctly identify it once present in the foveal region (within 3 degrees of visual angle of the fixation). Figure 5-7 shows the cumulative distribution of fixation number for all fixations from all trials (on the left); on the right, it shows the number of subsequent fixations in the trial (ie., the number of fixations it



took to locate the target after fixating within 3 degrees) for fixations *which were located within 3 degrees of the target*.



**Figure 5-7:** Fixation number (left) for all fixations from all trials, and fixations remaining (right) for fixations (from all trials) falling within 3 degrees of visual angle of the target.

If fixating near the target had no effect on locating it, we might expect the distributions to look roughly equivalent. However, subjective experience, eye tracking data, and common sense all tell us that this is not the case. The distribution of fixations remaining suggests that such fixations often result in target identification and trial termination within 1 – 2 subsequent fixations. For the purposes of analysis of missed identification, however, I classified all fixations within 3 degrees of the target which did not result in trial termination immediately or on the next fixation as misses. Although Figure 5-7 suggests that fixating within 3 degrees of the target can influence detection beyond the next fixation, such fixations which do not directly and immediately result in target identification and trial termination are not target identifications or ‘hits’ in the strictest sense.

To test whether conditional probability might affect target identification, the subset of fixations where the target was within 3 degrees was analyzed using logistic regression. Unlike in the previous analyses, here the final fixation in a trial was included, since by taking the subset of fixations which fell within 3 degrees of the target I remove fixations away from the target after it has been located. Fixations were coded as successful identifications if they were the last or second to last fixation in the trial, or unsuccessful identifications if more than one fixation followed, as discussed

above. The distance between the target and fixation is also included in the model, since the directness of the fixation on the target is likely to influence detection. The results of the logistic regression are detailed in Table 5-4.

<b>Logistic regression on Detection ~ Conditional Probability + Target Distance</b>					
	value	SE	df	t	significance
$\alpha$ (intercept)	1.281	0.077	17203	16.636	$p < 0.001$
$\beta_1$ (0.5 Conditional Probability)	0.194	0.072	17203	2.696	$p < 0.01$
$\beta_2$ (0.9 Conditional Probability)	0.258	0.073	17203	3.559	$p < 0.001$
$\beta_3$ (Target Distance)	-0.556	0.022	17203	-25.574	$p < 0.001$

**Table 5-4:** Logistic regression coefficients for the model Detection ~ Conditional Probability + Target Distance

The interaction between conditional probability and target distance did not reach significance when included in the model, and was thus dropped from the analysis. Target distance was significant, confirming that the directness of the fixation near or on the target is a factor in detection. Both conditional probability conditions were significant, indicating that when cued to the non-target color, searchers have a reduced likelihood of successfully identifying the target when fixating near it. The 0.9 conditional probability and 0.5 conditional probability coefficients were tested with the linear hypothesis function used above. The difference between coefficients was not significant at the  $p = 0.05$  level, but was very close,  $\chi^2(1, N = 17204) = 3.6511$ ,  $p = 0.056$ ; I speculate that a stronger result between the 0.5 conditional probability and the 0.9 conditional probability might be possible in a paradigm which explicitly controls distance between stimuli and fixation.

## Chapter 6

### General Discussion

#### 6.1 Recap of Results

In Experiments 1 – 3, I demonstrated that participants cued to the probable color of the target in a visual search task search more efficiently for high probability targets and less efficiently for low probability targets; I also showed that participants did not employ a strategy of exhaustively searching high probability stimuli before searching any low probability stimuli. In Experiment 4, I replicated my earlier results showing that participants were faster to locate high conditional probability targets, but only when they were explicitly informed of the probability relationships between cues and targets. When uninformed about the cue-target relationships, participants demonstrated no ability to learn these relationships or to effectively make implicit use of them; when misled about the cue-target relationships, participants showed some sensitivity to them, but were unable to make use of them to facilitate their search.

Experiment 5 examined eye movements during the search task employed in Experiments 1 – 4; below, I recap the significant eye movement findings. Participants in this task show a variety of expected eye movement behaviors: they fixate near distractors at a roughly equal rate, they fixate the target at a greater rate than the distractors, and they fixate visually distinct non-target items (the second cue) less often than distractors. Participants preferentially fixate higher density (in terms of stimuli) areas of the display, as I demonstrate with more precision than simply counting the number of stimuli near each fixation using the density score measure covered in Chapter 5. Using this new measure, I established that searchers in this task fixate longer on higher density areas of the screen. I also used a variation on this measure (match score) to measure the preponderant color of the stimuli near each fixation, as well as the magnitude of that preponderance. Using match score, I analyzed fixation behavior in each conditional probability condition, finding that participants started the search highly biased to look to areas of the display matching the color they were cued to, but over the course of the trial this bias shifted towards the color of the target, with the magnitude of the shift in proportion to the magnitude of the initial bias. Finally, I investigated the factors which caused searchers to fail to identify targets they had fixated, finding that while, as expected, the distance between fixation and target influenced identification, the conditional probability of the target also played a role above and beyond simply producing more or less efficient eye movements; searchers

tended to be less likely to identify a low probability fixated target, and trended towards being more likely to identify high probability fixated targets.

## **6.2 Discussion**

Previous work, much of which I review in Chapters 2 and 3, has been based on search displays which are simple (e.g., orienting tasks), regular (e.g., with stimuli always appearing in fixed locations, often in a grid arrangement), or both (e.g., Posner-style tasks in which only one or two stimuli appear in fixed locations). Such experiments can provide evidence for the effectiveness of probability as a graded attentional cue, but do not necessarily generalize beyond the limited search types or the display structures they employ.

My results provide evidence that people can make use of probability information even under conditions of greater complexity and variation. Despite conditions of near-total randomness for the spatial positions of the target and distractors, and despite significant trial-to-trial variation of distractor number and distractor color ratio (congruent:incongruent), participants in my task made use of conditional probabilities to effectively guide search. What's more, the modulation of efficiency I observe is not a result of a wholesale change in search strategy, but rather is the graded change that would be expected of a probability manipulation. Not only can probability guide search, but it does so even when constraints on complexity and spatial relations are abolished.

I interpret my results as demonstrating that cues can manipulate participants' estimates of the likelihood of forthcoming target features, in my case color, and not that they bias participant expectations. The distinction between probabilistic biases and expectations is subtle, but important. As reviewed in Summerfield and Egnér (2009), this distinction is often ignored or conflated in many studies on attention and the effect of informative cues. Summerfield and Egnér (2009) describe two main effects of participant expectations. First, violations of expectations may direct participants to prioritize inspection and evaluate preferentially surprising locations or objects. Second, expectations may bias the interpretations of sensory information. My protocol did not examine either of these sorts of effects. While my participants did expect, in the colloquial sense, that a magenta target would follow two magenta cues, this only served, in the framework of Summerfield & Egnér (2009) to bias attention. In my task, the search arrays always mixed two colors of items in roughly equal proportions. While participants could expect the target to be of a certain color, the appearance of the search array itself provided no opportunity to violate this expectation and therefore no opportunity to prioritize some locations or elements over others on the basis of such a violation. The other function

of expectation, to bias interpretation, was also not assessed in my task. My targets were all identified by virtue of a missing corner. My participants had no expectation over which corner would be missing, so there was no information available to them that could influence their interpretation of potential targets. Had I used a search array in which items were variously colored between the extremes of magenta and cyan, or if I had used cues that gave information as to a target's missing corner, I might have seen expectation effects, but I did not use such stimuli or cues. In short, the cues I used gave information about the likely color of the target, and as such allow us to interpret my results as probabilistic cues influencing attentional prioritization.

I have presented evidence that probability guides search; however, I find virtually no evidence that participants in this task can learn the conditional probability relationships through exposure alone, something that might be better characterized as implicit statistical learning to distinguish it from my explicit probabilistic cuing. I found this surprising for two reasons. First, the relation between cue color and target color is simple and straightforward – more cues of a certain color predict higher probability for the target to take that color. Second, as highlighted in the opening of this paper, there is a preponderance of evidence in support of the ubiquity and automaticity of visual statistical learning. In this section, I will highlight some of this relevant evidence and how it might relate to my own results.

One explanation for the inability of uninformed participants to learn the relationships of the cues to targets may be the complexity of the task. There are two cues, not one; I changed the shape and position of the second cue in order to increase the chance that it is noticed as distinct and attended. However, these changes were irrelevant for the cue-target contingency. In such a complex situation, it may be that greater experience than I gave here is necessary to discover the relevant contingencies. This relates to another possible explanation for the inability of my uninformed participants to learn the conditional probability relations between cues: the non-adjacency of the visual presentation. The cues are presented at the beginning of the task, and the task ends when the target is located; a demanding (in terms of visual resources) search interrupts the sequential perception of the statistically related cues and target. Turk-Browne, Jungé, & Scholl (2005) investigated statistical learning using a sequential presentation task in which shape stimuli were presented in an attended color and an unattended color, and found that nonadjacent relations were learned only for stimuli in the attended color. From this, they concluded that statistical learning is gated by selective attention. Pacton & Perruchet (2008) employed a task in which participants viewed a sequence of digits and performed an arithmetic operation on a pair of digits either immediately

succeeding (adjacent pair) or surrounding (nonadjacent pair) a target digit. Their findings and conclusions were similar to Turk-Browne et al.; they found that statistical relationships were only learned for digit pairs that were necessary for task completion, and they concluded that joint processing was necessary for the learning of such dependencies.

My results are in general agreement with this basic idea, though the term ‘selective attention’ might be too narrow. To complete my task, it was obviously necessary to perceive and attend to the target. My participants also certainly perceived the cues; each cue was presented alone on the screen for a full second. However, there were no competing stimuli that would require my cues to be selectively attended in the binary sense implied by Turk-Browne et al. Perhaps a more descriptive requirement for implicit learning of statistical relationships would be that the related stimuli be effortfully processed or interacted with. Uninformed participants in my task had little incentive to process or interact with the cues beyond their inevitable passive perception; thus, with this context it is not surprising to me that they fail to discover the connection between cues and target.

Ono, Jiang, & Kawahara (2005) investigated statistical learning in search using a paradigm quite similar to Chun & Jiang (1998). They repeatedly presented pairs of trials in which target position on the current trial was cued by the target position, distractor configuration, or a combination of both on the previous trial. They found that when all features of a previous trial were held constant (target position on no distractor trials, distractor configuration on targetless trials, and both target position and distractor configuration on target present trials with distractors) participants learned the relation between the previous trial’s characteristics and the location of the target on the current trial. However, when any aspect of the previous trial was allowed to vary (resulting in one predictive feature and one random feature) learning was abolished. Ono et al. explain these results using a signal:noise framework; predictive features generate signal, random features generate noise, and statistical learning requires a minimum signal:noise ratio.

My data support such an explanation. By the above definition, there is a great deal of noise inherent in my paradigm. Half of my trials (the 0.5 conditional probability condition) feature non-predictive cues, and every trial involves a search of some length that can also be considered as noise (since the distractors have no predictive value). This explanation also aligns with the selective attention/joint processing account of implicit statistical learning outlined earlier: selectively attending or processing a certain subset of stimuli can be equated to enhancing the signal of the processed set while simultaneously filtering out the noise of the unprocessed set.

I suspect that this is the crux of my participants' inability to learn the cue/target relations in my task. In the uninformed condition, the cues are the aspect of the display least salient to completion of the task; this could lead to a reduction in signal and even, possibly, partial filtering of them as noise. Conversely, the search display itself, the 'real' noise (as far as probability learning is concerned), requires the most intensive processing to complete the search task.

It is also possible that the difficulty my participants experienced in learning the probability relations in my task is a consequence of dealing with information about a target feature, and not target location. Although there is some evidence to the contrary (Chun & Jiang, 1999) the preponderance of evidence suggests that spatial information enjoys a significant advantage in the realm of implicit statistical learning. For example, Endo and Takeda (2004) showed that contextual cuing effects disappeared when the cued information was changed from spatial position to target identity. In their task involving four clusters of stimuli, each containing different numbers of potential targets, Williams et al. (2009) showed that spatial differences between clusters of stimuli (in terms of likelihood to contain the target) were detected and leveraged quickly and automatically. Surprisingly, participants in their task did not implement search strategies based on features, even though such strategies would be expected to yield faster RTs than the spatial probability-based search patterns that they were quick to adopt.

Though participants in this task did not learn the probability relations through exposure alone, when they were informed of those relations they searched more efficiently by preferentially fixating on areas of the screen containing predominantly target-colored stimuli. Recall from chapter 4 that I demonstrated that participants did not simply disregard the uncued color and exhaustively search the cued color in this task, as might be predicted by a binary on/off characterization of attention. Rather, participants show a consistent, graded effect of probability on their search behavior, fixating target color matching areas of the display preferentially but not exclusively. Such behavior in response to an endogenous probability cue supports the idea that the attentional allocation of visual resources acts on a continuum, and characterizations of attention can benefit from defining behaviors and environments in probabilistic, rather than absolute, terms.

Though my task is less than ideal for investigating the accuracy of target detection, as discussed briefly in Chapter 5, searchers' ability to successfully identify a fixated target was still modulated by probability. While I cannot say what the mechanism of this change in detection rate might be, the difference in likelihood of successfully identifying a fixated target based on conditional probability alone suggests that it is graded rather than binary.

This guidance of eye movements by probability (implemented in the task via color) is consistent with other recent work, e.g. Williams and Reingold (2001), who found that color effectively guided eye movements during a triple conjunction visual search task. Not only does the work presented here reaffirm their result, it also builds on it. By examining the guidance of eye movements in search through the lens of probability, I provide a quantifiable and comparable measure of attentional effects. Such a measure is useful not only because it offers a more precise description of the relevant phenomena, but also because it provides a method for comparing and equating disparate paradigms, or even disparate elements within paradigms. For example, Williams and Reingold found that color was the feature which most influenced eye movements (more so than shape and orientation) in their triple conjunction task. Couching this finding in the language of probability would allow comparison between their work and this one (and a plethora of other work on search and attention) and would also enable them to examine to precisely what degree color had more influence on eye movements than other features.

The above comparison and critique is included, not to dismiss the validity of the work, but to provide an example of the overarching goal of this thesis – that is, to demonstrate the value of precise, quantitative analysis of attentional phenomena in search, and to put forward probability measures as a method for achieving such analysis. It is my belief that such an approach is beneficial to work in the field, and it is my hope that I have provided a compelling example herein.



## Appendix A:

### Eye Tracking Data Processing Details

A great deal of the time and effort involved in this work was dedicated to developing the communication protocol for the S2 eye tracker and also to processing the raw data received from the S2 to identify fixations and saccades. This appendix details the communication and analysis methods for the S2 and the data it produced, with accompanying code samples; it is intended as both a reference for readers of this manuscript and a resource for others working with the S2 eye tracker or analyzing raw eye tracking data. Complete code for any or all of the protocols described in this appendix can be obtained from Bryan Cort (bryancort@gmail.com).

As previously mentioned, the S2 eye tracker is designed for use with the Windows operating system, and is not compatible out of the box with Linux, which all experimental protocols were configured and run on. This necessitated that I develop a protocol for communication between the Linux machine running the experiments and the S2 (connected to a separate machine running Windows).

The S2 operates as a server, responding to queries from a connected client. Connection is established via a TCP/IP connection, and queries/commands are sent (and responses received) as XML strings. In the communication protocol, the connection is established using the following (Python) code:

```
class Miramatrix:
    def __init__(self, server='130.127.49.153', port=4242, loggers = []):
        try:
            # Connect to the server
            self.sock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
            self.sock.connect((server, port))
            print 'connected to',server, '!'
            self.state = 'connected'

            self.calibrated = 0
            self.eyeDict = {}
            self.logging = 0

            self.loggers = loggers

            # Start a thread for receiving
            self.receiving_thread = threading.Thread(target=self.communication_loop)
            self.receiving_thread.start()

        except:
            print 'whoops! couldn\'t connect to',server,'on port',port
```

After the connection is established, the protocol repeatedly queries the S2:

```

def communication_loop(self):
    self.sock.settimeout(1.0)

    while self.state != 'shutdown':
        try:
            # receive a message
            message = self.receive(4096)
            self.handle_message(message)
        except:
            # It's possible we're shutting down and we tripped the
            # socket time out. See if that was the case: state will be shutdown.
            if self.state == 'shutdown':
                return

```

It handles each reply it receives by splitting the XML strings into a dictionary object containing name:value pairs (one pair per variable received):

```

def handle_message(self, msg):

    msg = unicodedata.normalize('NFKD', msg).encode('ascii','ignore')
    msg = msg.replace('"', '')
    # Take the xml data and split it into the tokens
    split = msg.split(' ')
    # remove the equals signs, the <REC token, and any leftover XML formatting
    pairs = [ i.replace('/>\r\n','').split('=') for i in split[1:] ]
    # Look at all of the key/value pairs
    for p in pairs:
        #build a dictionary with the XML tags as keys and the eyetracking data as value
        if(len(p) == 2):
            self.eyeDict[str(p[0])] = str(p[1])

    for logger in self.loggers:
        try:
            if logger.logging:
                logger.write()
        except:
            print 'Data logging error'

```

Data is recorded by one or more logger objects attached to the base ‘Mirametrix’ object instantiated by the communication protocol:

```

class Logger:
    def __init__(self, tracker, outnamespace, log, varListET=['CNT','FPOGD','FPOGID','FPOGS','FPOGV','FPOGX','FPOGY','TIME'], varListEXP=[]):
        '''log is the file to log data to, tracker is the Mirametrix object receiving eyetracking data, varListET is the list of XML tags to
        get data for from the tracker, varListEXP is a list of variable names in the environment to monitor and record in addition to
        eyetracking data. outnamespace is the namespace to look for the variable in varListEXP in; in general, should be 'globals()',
        ie., loggerDemo = Logger(..., globals(), ...). Do not add quotes around globals(), it is not a string.'''

        self.tracker = tracker
        self.outnamespace = outnamespace
        self.log = log
        self.varListET = varListET
        self.varListEXP = varListEXP

        self.filehead = self.varListET + self.varListEXP
        self.log = log
        self.log.write(str(self.filehead).replace('[','').replace(']','').replace(' ',''))

        self.logging = 0

```

This communication protocol runs in parallel with the experiment. After all the data has been received and recorded, substantial processing is necessary to generate usable saccade/fixation classifications from the raw gaze data. The code to classify this raw data was also written in Python.

The transformation of the raw gaze data into fixations and saccades took place in three steps. First, the x- and y-coordinate data was smoothed by replacing each value with the median value of the

five element set including that particular value, and two preceding values, and the two following values. For example,  $x_{10}$  would be replaced by  $\text{median}(x_8, x_9, x_{10}, x_{11}, x_{12})$ .

Second, velocity was calculated for each observation and observations for which velocity exceeded 100.00 degrees of visual angle per second were classified as saccades; the remaining values were classified as fixations. Fixation coordinates were calculated by taking the averages of the x and y coordinates of each raw data point; fixation duration was a trivial calculation of the difference between the timestamp of the last raw data point in that fixation, minus the timestamp of the last raw data point preceding that fixation (ie., the last data point in the previous saccade).

Third, fixations which were very close (less than 1 degree of visual angle) in position were combined, and fixations of very short duration (less than 60ms) were eliminated. This correction was applied to eliminate unrealistically short fixations, and to correct for single fixations that had been erroneously classified as separate, very close fixations.

Though these operations are quite simple in principle, their implementations in Python are too lengthy to include here. However, as mentioned above, I am happy to provide those implementations, in their entirety, to any interested party.

## Bibliography

- Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, 2
- Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. *Vision Research*, 41(25), 3613-3625.
- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106(3), 1382-1407.
- Berkley, M. A., Kitterle, F., & Watkins, D. W. (1975). Grating visibility as a function of orientation and retinal eccentricity. *Vision Research*, 15(2), 239-244.
- Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks. I. contrast detection and discrimination. *JOSA A*, 20(7), 1341-1355.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4), 487.
- Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes extracting categorical regularities without conscious intent. *Psychological Science*, 19(7), 678-685.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97(4), 523.
- Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, 112(2), 291.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13), 1484-1525.

- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28-71. doi: 10.1006/cogp.1998.0681
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360-365. doi: 10.1111/1467-9280.00168
- Eckstein, M. P., Beutter, B. R., & Stone, L. S. (2001). Quantifying the performance limits of human saccadic targeting during visual search. *PERCEPTION-LONDON-*, *30*(11), 1389-1402.
- Endo, N., & Takeda, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Attention, Perception, & Psychophysics*, *66*(2), 293-302.
- Estes, W., & Taylor, H. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences of the United States of America*, *52*(2), 446.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499-504.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489-1506.

- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, 20(11), 1254-1259.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35-B42.
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434(7031), 387-391.
- Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3)
- Neisser, U. (1964). Visual search. *Scientific American*,
- Peterson, W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Information Theory, IRE Professional Group On*, 4(4), 171-212.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, 32, 531-556.
- Snodgrass, J. G., & Townsend, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception and Performance*, 6(2), 330.
- Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 25(2), 168-199.

- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136.
- Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134(4), 552.
- Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 399.
- Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review*, 8(3), 476-488.
- Williams, C. C., Pollatsek, A., Cave, K. R., & Stroud, M. J. (2009). More than just finding color: Strategy in global visual search is shaped by learned target probabilities. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 688-699. doi: 10.1037/a0013900
- Wolfe, J. M. (1994). Guided search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202-238.
- Wolfe, J. M. (2007). Guided search 4.0. *Integrated Models of Cognitive Systems*, 99-119.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419.

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention* Oxford University Press, USA.