EXPECTATIONS INFLUENCE VISUAL SEARCH

A Dissertation
Presented to
The Academic Faculty

by

Carolyn Hartzell

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
MAY 2017

COPYRIGHT © 2017 BY CAROLYN HARTZELL

EXPECTATIONS INFLUENCE VISUAL SEARCH

Approved by:

Dr. Rick Thomas, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Wendy Rogers
School of Psychology
Georgia Institute of Technology

Dr. Daniel Spieler
School of Psychology
Georgia Institute of Technology

Date Approved: April 21, 2017

ACKNOWLEDGEMENTS

I would especially like to thank my advisor, Dr. Rick Thomas, for all of the late nights teaching me the skills I needed to complete this thesis. Without his above and beyond assistance, this thesis would not have been possible. I would also like to thank Nick Internicola for being so understanding and supportive of me as I have worked endless hours to finish this thesis.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	**iv**
LIST OF TABLES	**vi**
LIST OF FIGURES	**viii**
LIST OF SYMBOLS AND ABBREVIATIONS	**ix**
SUMMARY	**x**
CHAPTER 1. Introduction	**1**
1.1 Working Memory and Visual Search	**4**
1.2 Hypothesis-Guided Search	**5**
CHAPTER 2. Experiment	**7**
2.1 Method	**8**
2.1.1 Participants	8
2.1.2 Stimuli and Apparatus	8
2.1.3 Procedure	10
2.2 Results	**15**
2.2.1 Search A: First Target Search Results	15
2.2.2 Search B: Second Target Search Results	17
2.3 Discussion	**27**
CHAPTER 3. Model	**32**
3.1 Model Explication	**33**
3.1.1 Hypothesis Generation Model (HyGene)	33
3.1.2 Fixation Selection and Inhibition	39
3.1.3 Accumulator	41
3.1.4 Stopping Rules	42
3.2 Model Fitting	**43**
3.3 Results	**46**
3.3.1 Fit Results	46
3.3.2 Comparison with Empirical Data	52
3.4 Model Discussion	**55**
CHAPTER 4. General Discussion	**58**
REFERENCES	**62**

LIST OF TABLES

Table 1	Overview of Experiment Ecology	11
Table 2	Detailed Explication of Experiment Ecology	12
Table 3	Expectation Definitions and Corresponding likelihoods	18
Table 4	Results of GLM Analyses for Search B Response Times for Any Second Target	20
Table 5	Mean Search B Response Times by Likelihood	20
Table 6	Results of GLM Analyses for Search B Response Times for Specific Second Target	21
Table 7	Search B Response Times Based on Likelihoods for Specific second target, conditional on First Target	21
Table 8	Second Target Dependent Measures based on Target 1 ID	22
Table 9	Summary of Statistical Model Results for Miss Errors	25
Table 10	Odds of Error based on Likelihoods for Specific Second Target	26
Table 11	Odds of Error given Specific Cue-Target Sets	27
Table 12	Model Parameters	44
Table 13	Summary of Rankings for Model Fit Results	46
Table 14	HGS-All Fit Metrics and Parameters	48
Table 15	Comparison of HGS-All Model and Empirical Fixation Data	50

Table 16	Model Fit Statistics and Best Fitting Parameters for Cue 1 Trials	51
Table 17	Model Fit Statistics and Best Fitting Parameters for Cue 2 Trials	53
Table 18	Model Comparison for Miss Errors	55
Table 19	Model Comparison for Odds Ratios for Miss Errors	55

LIST OF FIGURES

Figure 1	Overview of Cues, Disease, and Targets	11
Figure 2	Visualization of Ecology	13
Figure 3	Fixation Selection Conditional Probabilities	24
Figure 4	Overview of the Hypothesis-Guided Visual Search Model (HGS).	33
Figure 5	Overview of the HyGene model (Thomas et al., 2008) used in the present hypothesis-guided search model (HGS).	35
Figure 6	Schematic of implementation of episodic and semantic memory stores. Each color represents a possible target color. The gray mini-vector represents the no-target hypothesis. Note that the mini-vectors in the episodic traces are degraded relative to the semantic traces.	36
Figure 7	Overview of the visual search portion of the model, including fixation selection and inhibition.	41
Figure 8	Conditional probability of fixating each color given the first target found for the HGS-All model and the empirical data	54

LIST OF SYMBOLS AND ABBREVIATIONS

AIC	Akaike information criterion
C1 through C4	C1, C2, C3, and C4 refer to the four colors that were used for the experiment and the model. Color assignments differed by participant, so specific color names are not used.
HGS	Hypothesis-guided search model. HGS-All and HGS-SOC refer to two variations of the model that were run.
HyGene	Hypothesis generation model
LCA	Leaky, competing accumulator model
$L_{CA}(E)$	Likelihood based on trial cue and first target found for the existence of a second target
$L_{A}(E)$	Likelihood based on just the first target found for the existence of a second target
$L_{CA}(B)$	Likelihood based on trial cue and first target found for the specific target remaining in the trial.
$L_{A}(B)$	Likelihood based on only the first target found for the specific target remaining in the trial.
QIC	Quasi-likelihood under the independence model criterion
SOC	Set of contenders, a WM-like component of HyGene
SSM	Subsequent search miss (error)
T1 or T2	First target found during search (T1) or the second target present during search (T2), whether it was found or not
WM	Working memory

SUMMARY

Satisfaction of search errors, also called subsequent search misses, are a costly visual search problem, particularly in radiology. To date, research on causes and interventions for satisfaction of search errors has focused on properties of the stimuli and the mechanics of the search process. I present evidence to support a new theoretical understanding of some of the underlying cognition that drives search behavior and that can predict visual search errors.

An eye-tracked experiment that manipulated participant expectations of target characteristics and number of targets demonstrated that participant expectations, generated based on environmental cues and long-term memory, influence search behavior. Through exemplar training, participants learned to associate cues with target sets that varied in color of target and number of targets. Participants were instructed to utilize these learned relationships to facilitate their visual search. Analysis of response time, fixation data, and miss errors indicated that expectation was a significant predictor of search behavior, with lower expectations for secondary targets being associated with shorter response times, more miss errors, and fewer fixations to unexpected colors.

In a first step towards understanding the cognitive mechanisms behind visual search misses for secondary targets, a cognitive process model was developed. This model integrated hypothesis-guided search with visual search to predict participant behavior. The model was tested against the empirical data and successfully captured the high-level results of the experiment. Future iterations of the model will seek to better fit the more subtle complexities of the empirical results.

CHAPTER 1. INTRODUCTION

When the number of visual search targets is unknown, a given target is more likely to be missed if other targets are present and identified first—an effect originally labeled as *satisfaction of search* by radiology, though it occurs in all domains (Berbaum et al., 1990; Tuddenham, 1962). Although the name implies that the error is due to early search termination once the observer is satisfied that all targets have been found, more recent research has shown that this does not account for all search errors (Cain, Adamo, & Mitroff, 2013). To this end, the errors have been re-labeled *subsequent search misses* (SSMs; Cain et al., 2013). Visual search errors, particularly search misses, have been the cause of many medical malpractice lawsuits over the years (Berlin, 1997). According to one review, over 30% of the medical malpractice lawsuits at one hospital were due to radiological diagnostic errors with an estimated 60% due to perceptual errors (Berlin, 1997). In a different study, about 80% of radiological diagnostic errors were perceptual (Donald & Barnard, 2012). Furthermore, up to 30% of these perceptual errors may be due to satisfaction of search effects (Berbaum, Franken, Caldwell, & Schartz, 2010). For most of these perceptual errors, radiologists are later able to clearly perceive abnormalities that they had previously missed (Berbaum et al., 2010), implying that there was something in the original search that was fundamentally different from the later search.

Many explanations for SSM effects focus on the parameters of the search itself, such as the perceptual characteristics of the targets and distractors, or the working memory resources used to store target information (Cain et al., 2013; Cain & Mitroff, 2013; Körner & Gilchrist, 2007). I argue and test the idea that higher-level cognition can significantly affect search behavior and search errors, particularly for subsequent search misses. Specifically, I present evidence from a research experiment and a computational cognitive model that expectations that are generated from long-term memory drive visual search, and can account for search behavior both before and after an initial target is located. Previous work in the field of radiology has demonstrated that clinical information and radiologist experience affect visual search patterns, but little work has been done to understand the cognitive mechanisms affecting SSM errors (Manning, 2010).

Development of a cognitive process model that can be fit to the empirical data can lead to improved understanding of how and why experience and expectations can lead to SSM errors. From there, work can be done to begin to develop effective interventions to reduce errors. This account integrates literature on multiple target search, working memory-guided search, and hypothesis-guided search to improve understanding of how expectations influence visual search behavior. My theoretical and computational account provides a new lens through which to consider visual search and should motivate new interventions to reduce search errors.

Historically, research into satisfaction of search and SSMs has focused on how the mechanics of search break down, as opposed to why they break down, despite the extensive literature on the top-down influence of cognition on search (e.g., Beck, Hollingworth, & Luck, 2012; Treisman & Sato, 1990; Wolfe, 2007). Research in radiology has generally employed signal detection theory, receiver operating characteristic curves, and time course curves to analyze how targets get missed (Berbaum et al., 2010). Although useful tools, they are descriptive, and do not address, for example, *why* discrimination or bias may be dependent on the number of targets. Psychological research has explored how attention and working memory may cause SSM errors, but as explained further below, these explanations typically focus on resource utilization or set biases without addressing cognitive functioning (Adamo, Cain, & Mitroff, 2013; Biggs, Adamo, Dowd, & Mitroff, 2015; Cain et al., 2013; Körner & Gilchrist, 2007). Perceptual set theory is a non-cognitive theory, that has been proposed, which states that once a target is found, the visual system is biased toward finding targets that appear similar (Cain et al., 2013). Evidence has been found to support perceptual set theory in some cases but not all, and the theory does not account for all SSMs (Biggs et al., 2015; Cain et al., 2013).

A number of interventions have been tested to reduce SSM errors, but, because there is still no real understanding of *why* SSMs occur, most of these interventions have shown little success. Setting a minimum time for search, removing the first target found, directing attention to unexamined areas, checklists, and a few other interventions have all failed to significantly reduce search misses (Berbaum, Franken, Caldwell, & Schartz, 2006; Cain & Mitroff, 2013; Wolfe et al., 2009). To date, the intervention with the most success was an unanticipated effect that occurred while studying a different aspect of SSM errors, namely, requiring radiologists to think aloud as they examine x-rays (Berbaum et al., 2010). The reason for the success of this method in reducing SSMs is still not well understood. It is clear that we need to derive a better understanding of the cognition underlying the errors. Only with an understanding of why they occur can we expect to develop effective interventions.

1.1 Working Memory and Visual Search

Research into how cognition affects SSM errors has focused primarily on the role of working memory. The contents of working memory (WM) have been shown to influence and guide visual search behavior, even when the contents are not relevant to the search in question (e.g., Beck et al., 2012; Olivers, 2009; Olivers, Peters, Houtkamp, & Roelfsema, 2011; Soto & Humphreys, 2007). Under the assumption of WM-guided search, it has been proposed that WM depletion is the cause of SSM errors.

There are three mechanisms by which this has been proposed to occur, all of which presuppose that WM is used to conduct efficient visual search. First, if multiple target search also has multiple target representations, more WM resources are used to maintain the representations, leaving less available to guide search. It has been shown that search efficiency degrades when multiple representations are being maintained for visual search (Stroud, Menneer, Cave, & Donnelly, 2012). For the second and third theories, research suggests that observers automatically encode certain information about targets once they are found, thus utilizing WM resources (Cain & Mitroff, 2013). Some have proposed that memory is pre-allocated for this information, citing evidence that multiple target search is less efficient from the beginning of search (Körner & Gilchrist, 2007). Others believe that WM is utilized once the first target has been found, citing evidence that observers often re-fixate already-found targets (Cain & Mitroff, 2013). This last theory suggests that cognition changes after the first target is found, but it does not address higher order cognitive functioning, just the depletion of a cognitive resource. Both the multiple target representation and the WM pre-allocation theories obliquely assume that participants have generated expectations about how many targets there are and what the targets may be, but they fail to address how such expectations are generated, largely because in most visual search research, these expectations are provided to the participants directly. The resource utilization theories also do not address *how* cognition (and expectations) may change as search proceeds.

Studies in visual search often provide participants with the identity or a direct representation of the target for which they will be searching (e.g., Beck et al., 2012; Hout, Walenchok, Goldinger, & Wolfe, 2015), but in many applied tasks, it is up to the observer to generate a set of possible search targets based on a visually unrelated cue. For example, a radiologist will use a patient's clinical history to generate possible diagnoses, and then use these diagnoses to guide search for relevant abnormalities in an x-ray (Manning, 2010). The diagnoses generated are based on the radiologist's long-term memory, from extensive training and experience. More experience leads to better diagnoses (Manning, Ethell, Donovan, & Crawford, 2006). Furthermore, more experience has been shown to influence the mechanics of search behavior. Radiologists with more experience have longer average saccade lengths and fewer fixations around the landing area of the saccades (Manning, 2010). It is thus clear that experience influences visual search. It is also clear that observers often use their experience and external cues to generate predictions about what they will find (Manning, 2010). Tasks that do not include this generation step may be missing a critical aspect of what guides search and how subsequent search behaviors are affected.

1.2 Hypothesis-Guided Search

Thomas et al. (2008) proposed the idea of *hypothesis guided search,* where the hypotheses one generates guide search behavior, which can account for both the role of experience in search and the generation process for uncertain targets. Although originally this hypothesis-guided search was used to explain information search behavior, Buttaccio, Lange, and Thomas (2014) have recently extended the idea to a visual search task in which possible target representations are generated from long term memory based on external cues. This theory directly applies to domains like radiology where the radiologist will use patient information and symptoms to generate hypotheses for possible abnormalities that may appear in an x-ray. The details of how hypotheses are generated using external cues to prompt long-term memory, and how these hypotheses subsequently guide search will be explained in more detail in CHAPTER 3. To date, hypothesis-guided search has only been applied to the *retrieval guidance paradigm*—single-target searches that can be accomplished in relatively few fixations. The current study extends the retrieval guidance paradigm and the concepts of hypothesis-guided search to account for the influence of participants' generated expectations on visual search behavior, both before and after first-target acquisition, in order to provide a new theoretical understanding of errors like SSM.

The remainder of this paper is divided into three chapters: experiment, model, and general discussion. The experiment chapter describes the procedure and results of a laboratory experiment designed to reveal how expectations can influence search behavior. The model section combines a model of hypothesis generation with a model that simulates certain features of visual search to help reveal the key cognitive processes involved in influencing search behavior. The final section discusses the implications of the experiment and the corresponding model and suggests new avenues of research to improve understanding of how higher order cognition affects visual search.

CHAPTER 2. EXPERIMENT

The present experiment combines the tools and stimuli of basic research with concepts from applied domains, with a focus on radiology. In radiology, SSMs have been categorized into three different mechanisms of error: scanning errors, where the target is never fixated; recognition errors, where the target is fixated but for a duration insufficient to make a decision about it; and decision errors, where the target is fixated long enough for a decision to be made regarding its status (Kundel, Nodine, & Carmody, 1978; Nodine, Mello-Thoms, Kundel, & Weinstein, 2002). The threshold between recognition and decision errors has traditionally been defined as the average dwell time the participant has used to make correct decisions about other targets, or, as an empirically-derived heuristic, about 1000 ms (Berbaum et al., 2010). The experiment reported in this paper eliminates decision errors, using targets that are easily distinguishable from distractors once they are fixated. Furthermore, because there were no a priori predictions regarding the differential influence of expectation on scanning errors compared to recognition errors, these two error types were combined. Here, SSM errors were defined as participants' failure to report a present second target, regardless of whether it was fixated. Subsequent research will attempt to differentiate the cognitive mechanisms between the two error types.

During the study, participants first learned an environmental ecology for cue and target set relationships. This is similar to a radiologist's acquired experience with the relationship between symptoms on a patient's medical chart and possible abnormalities in an x-ray. After training, participants were presented with just the cue and then were eye-tracked as they searched for one or two targets in an array of distractors. I predicted that initial search behavior would be driven by the hypotheses most highly correlated with the cue, where hypotheses are defined as possible target characteristics – in this case, color. I predicted that search occurring after the first target would be driven by the hypotheses most highly correlated with the integrated information from the cue and the first target found. Measures of search behavior included response times, fixations by object color, and miss errors for secondary targets.

2.1 Method

2.1.1 Participants

Thirty-six Georgia Tech undergraduate students (53% female) participated for partial course credit. Participants had self-reported normal or correct-to-normal vision.

2.1.2 Stimuli and Apparatus

2.1.2.1 <u>Display and Eye-tracker</u>

The display was 33.9° by 19.4° visual angle with a resolution of 1920x1080, and participants were 88 cm away. The background was gray (RGB: 180, 180, 180). The eye-tracking apparatus was the desktop-mounted EyeLink 1000 Plus (SR Research Ltd., 2015). Participants' dominant eye was tracked using a high resolution camera with sampling rate of 1000 Hz. Participants used a chin rest to maintain stability and viewing distance. Calibration was performed using a nine-point dot array, with a maximum acceptable validation error of 1° of visual angle. Participants were checked for drift at the end of the first block of trials and had to maintain a fixation less than 1° deviation from the drift check point for 1000 ms in order to proceed. If they could not, they were recalibrated. The threshold for saccadic amplitude was 1.0° visual angle. Fixations were defined as fixed eye gaze position for at least 50 ms located within 1.5° visual angle of an object, to allow for some incidental drift during each experiment block. Fixations to empty space were not analyzed. The minimum distance between display objects was 2.2° visual angle.

2.1.2.2 Cues

Two simple geometric shapes, a circle and a square, were used as cues. Cues were black line drawings that fit within a square of 1.1° visual angle. During training, cues were visible for the duration of each training trial. During test, cues were visible until the onset of the search array.

2.1.2.3 Targets and Distractors

Targets were rotated Ts and distractors were rotated Ls, both sized to fit within a square of 1.2° visual angle. These objects could each be one of four different colors that loosely corresponded to the labels red (RGB: 244, 67, 123), green (57, 153, 103), blue (120, 130, 255), and yellow (216, 115, 38) and were equalized for luminance.

2.1.2.4 Search Array

Eight objects of each color appeared in every visual search trial for a total set size of 32, one or two of which were targets. The search display was divided into a 5x9 grid, such that there was a maximum of one object per grid square. No objects could appear in the central grid square. Object grid locations for each trial were generated by random assignment and then jittered within each grid square. Each participant saw the same object locations as if each search array were an x-ray or a "case," such that each participant saw the same cases, but in a random order.

2.1.3 Procedure

During initial training, participants were presented with example stimuli allowing them to learn associations and likelihood relationships between cues and targets. These associations included both the likely number of targets based on cue, and the likely color of the target(s), based on cue. For each test trial, a cue was presented, which prompted the generation of hypotheses about the target(s) for which to search. I predicted that the generated expectations, about both color and number of targets, would drive subsequent search behavior.

2.1.3.1 Ecology

Cues were presented to the participants 40 times each. Each cue was strongly associated with a "disease." Participants were informed that there were two underlying diseases, with each disease defined by two "symptoms," or colors, and so over the 80 exemplar trials they learned that each cue was highly associated with two of the four colors. Specifically, for 80% of the cue-1 trials, target sets were from disease-1, defined as combinations of color-1 (C1) and color-2 (C2). For 20% of cue-1 trials, target sets were from disease-2, defined as combinations of C3 and C4. The reverse pattern was true for cue-2 trials. Cue, disease, and target set combinations are shown in Figure 1. Specific conditional likelihoods can be seen in Table 1, and are depicted in Figure 2 for additional clarification. Not every possible target combination was used in order to simplify the task structure enough to be learnable in a one-hour session. For each cue, at least one of each of the associated colors appeared in an equal number of trials, so that participants would be motivated to search for both of the likely colors, rather than favoring one. Another characteristic of the ecology, as shown in Table 1, is that cue-1 was more likely to be matched with a dual target search trial than cue-2 was. Thus, when participants saw cue-1, they should have had a higher expectation of finding two targets.

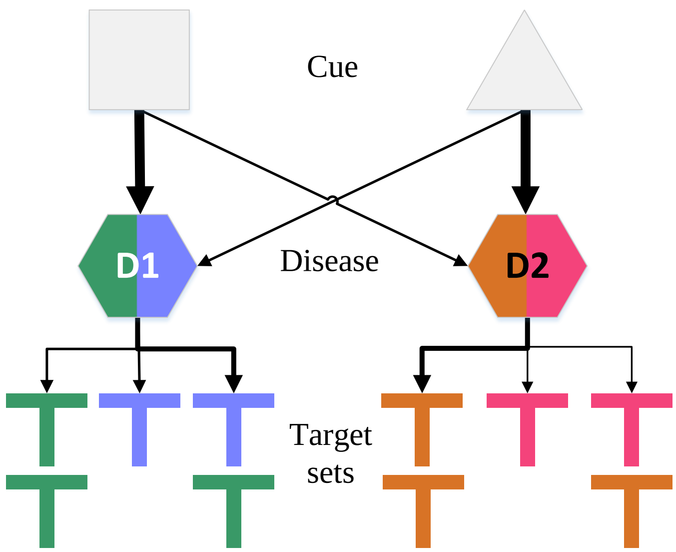

Figure 1. This figure shows the possible cue, disease and target combinations. Cues are represented by the geometric shapes. Arrow width loosely corresponds to likelihood of combination. Note that actual colors were varied between subjects.

Table 1.

Summary of Environmental Ecology

	Probability								
			Target Sets						
	Disease			Dual Target			Dual Target		
Given	D1	D2	C_1	C_1C_2	C_2C_2	C_3	C_3C_4	C_4C_4
Cue1	.80	.20	.15	.50	.15	.075	.05	.075
Cue2	.20	.80	.05	.10	.05	.375	.05	.375

What Table 1 does not readily convey is that this ecology engenders an additional set of conditional dependencies between the targets, based on the target that was found first. These conditional probabilities for the second target are displayed in Table 2. For example, given a cue-2-C3-C4 target set, if a participant finds the C3 target first, they would expect that it is likely to be the only target in the array. If they find the C4 target first, they are more likely to search for a second target, with the expectation that it is probably a C4 target. These conditional probabilities allow us to distinguish between search behavior based on perceptual characteristics and raw base rates from behavior based on learned environmental structure.

Table 2.

Conditional Likelihood of Second Target

		Likelihoods					
		Target 2					
Given		C1	C2	C3	C4	No Target	Any Target
Cue-1							
First Target =	C1	.	.769	.	.	.231	.769
	C2	.769	.231	.	.	.	1.0
	C3	.	.	.	.400	.600	.400
	C4	.	.	.400	.600	.	1.0
Cue-2							
First Target =	C1	.	.667	.	.	.333	.667
	C2	.667	333	.	.	.	1.0
	C3	.	.	.	.118	.882	.118
	C4	.	.	.118	.882	.	1.0
Cue ignored/forgotten							
First Target =	C1	.	.75	.	.	.25	.75
	C2	.75	.25	.	.	.	1.0
	C3	.	.	.	.18	.82	.18
	C4	.	.	.18	.82	.	1.0

Notes. Probabilities equal to 0 have been replaced with a period to reduce visual clutter.

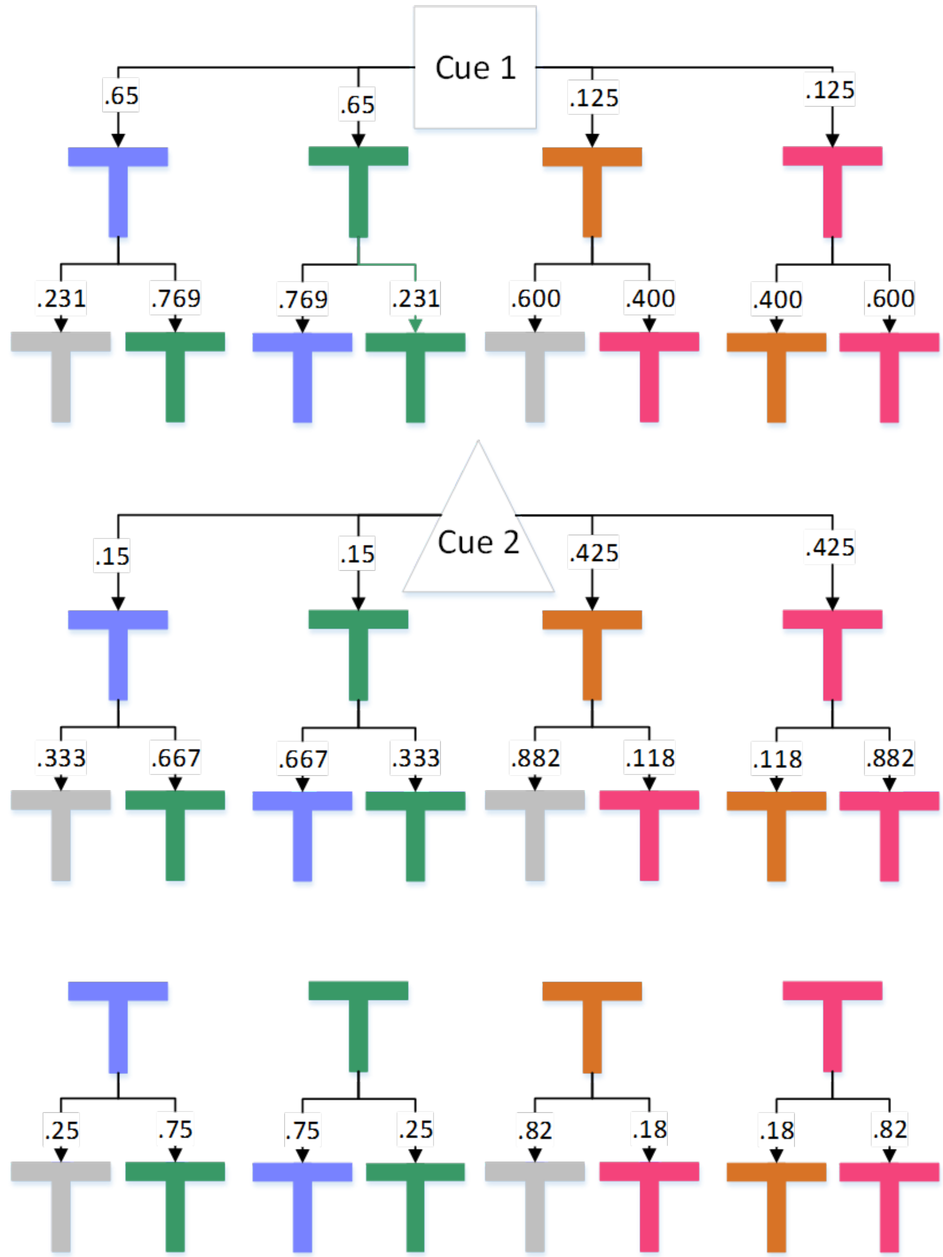

Figure 2. Visualization of Table 2, showing cue, first target found, and the possible second targets. Between cue and first target found are the conditional probabilities of encountering a trial containing at least one target of that color, given the cue. Probabilities between first and second target represent the likelihood that the second target will be present given the cue and the first target. The bottom target sets (without cue) show the likelihoods collapsed across cue. Colors of the first targets represent, from left to right, C1-C4. Gray targets indicate there is no second target.

2.1.3.2 Training

Participants were given two blocks of 40 trials of exemplar training. By the end of training they had seen the entire environmental ecology described in the previous section, so that under perfect learning, they would have internalized the likelihoods for each cue-target(s) sets. They were told to pay attention to the relationship between the cue and the number and color of target(s). They were not told any specific probabilities or likelihoods. During the training trials, participants were first shown the cue, followed by the first target, and then, if there was another target for that trial, the second target. Items appeared in series but stayed visible for the entire training trial so that the cue and target(s) were all visible by the end of the trial. Presenting targets sequentially was done to facilitate participant learning of the conditional dependencies between targets, explained further above in the ecology section.

2.1.3.3 Test

Participants then moved to the visual search phase. Participants were instructed to hit a button when they located the first target, at which point they were queried about the target's orientation. During the orientation question, a visual mask covered the search array. Once they responded to the query, they were returned to the search array until they hit a button to terminate search, either because they had found the second target or because they had decided that none existed. After the button press the search array was blocked with a mask, and participants were asked to indicate the orientation of the second target if one existed. Each visual search trial was preceded by a display of a cue for 1000 ms. As stated previously, in each trial, 32 objects were presented, one or two of which were targets.

Participants were told that there was at least one target in every trial. For each of the four colors, there were eight objects, regardless of the number of targets in the display.

2.2 Results

Dependent variables included response time (RT), number of fixations, and miss errors. Analyses were performed separately for the first half of the trial, while the participant searched for the first target, and the second half of the trial, while the participant searched for the second target. The first half of the trial (Search A) was defined to be from the onset of the search array to when participants pressed a button indicating that the target was found. The second half of the trial (Search B) was defined to be from the onset of the search array following the orientation question to when participants pressed a button to indicate either that a target was found or that they believed there to be no additional target.

2.2.1 Search A: First Target Search Results

2.2.1.1 Response Time Results

Trials were not analyzed if any of three conditions occurred: if RT was less than 200 ms, if participants responded with the incorrect orientation, or if participants did not fixate the target prior to pressing the button. The first criterion was to ensure that participants had time to look at the display before responding. The second criterion served as an attention check, under the assumption that incorrect responses indicated that the participant had not actually seen the target. The third criterion was because analyses would only be informative of search RT if there was reason to believe that the participant actually found an initial target. Combined, trials matching these criteria account for 9.2% of the data, with most errors stemming from trials in which the eye-tracker did not record consistent fixation data. Response times were analyzed using GLM for repeated measures, using as predictors the number of targets and a binary variable representing whether the first target was an expected color or not. A color was considered to be “expected” if it was one of the two colors highly associated with the displayed cue. It was predicted that RTs would be faster for trials with two targets ($n_t = 2$) due to the reduction in distractor set size. It also was predicted that RTs would be faster if the target color was expected. Number of targets and target expectation were predicted to interact such that RTs in trials with two expected-color targets would be faster than RTs in trials for a single predicted-color target.

The results yielded a significant main effect of number of targets, $F(1,35)=186.88$, $p<.0001$, $M(n_t=1)=10.868$ sec, SE=.258; $M(n_t=2) = 6.593$ sec, SE= .177). There was also a significant main effect of expectation, $F(1,35)=99.93$, $p<.0001$; M(not expected) = 10.294 sec, SE =.281; M(expected) = 7.167 sec, SE = .138. The interaction was also significant, $F(1,35) = 9.20$, $p=.0045$. These results supported the prediction that search was faster for expected targets, indicating that participants were using information about target likelihoods, as prompted by the cue to guide their search.

2.2.1.2 Fixation Results

The measure of interest for fixation data was the relative proportion of fixations to objects of each color. As mentioned above, the fixations that were analyzed were defined as a fixed eye position held for at least 50 ms within 1.5 degrees visual angle of an object. The minimum distance that could separate any two objects was 2.2 ° visual angle. Fixations to empty space were not analyzed for the purposes of this study. It was predicted that there would be more fixations to colors prompted by the cue. Cue was used as a proxy for expectation for this analysis.

Analyses for fixations were performed using GLM with a logit link function and a binomial distribution, using color and cue as predictors. There was a significant interaction between color and cue (χ^2 (3) = 23.42, $p<.0001$), with expected colors receiving a significantly higher proportion of fixations than unexpected colors. Odds ratios were alos calculated to determine the odds of fixating one color (or combination of colors) compared to any other color. For cue-1 trials, participants were 1.592 times more likely to fixate C1 or C2 versus C3 or C4 ($p<.0001$). For cue-2 trials, participants were 1.685 times more likely to fixate C3 or C4 versus C1 or C2 ($p<.0001$). These results supported the prediction that participants' expectations, prompted by the cue, would guide fixations to expected target colors.

2.2.1.3 <u>Miss Error Results</u>

Miss errors were not analyzed for Search A, because they can only occur for the second target, in Search B.

2.2.2 Search B: Second Target Search Results

For Search B, several different ways of defining participants' expectations were identified. Each definition corresponded to a different set of likelihood values for the Search B targets, so each definition was analyzed to determine which had the best fit to the participants' search behavior. The definitions and likelihood values are explicated in Table 3, though it should be noted that the same values can be seen in the explication of the ecological structure in Table 2. Likelihoods may be calculated using as the given information different ecological elements, and may be calculated for different events. In this experiment, likelihoods may be conditional on the trial cue *and* the first target found or just the first target found. Likelihoods may be calculated for the *existence* of a second target or for the *specific* target remaining in the trial. All combinations of these definitions are plausible, so each was assessed for best fit for response time and error analyses. For $L_{CA}(B)$ and $L_A(B)$ single target trials, the likelihoods represent the likelihood of a single-target trial (the No T column of Table 2) given the specific information. Although the metric likelihoods are shown in Table 3, these expectations were treated as categorical predictor values for analysis. We also considered the possibility that participants did not update their expectations after the first target found, so we tested models that just used likelihoods based on cue as the predictors, but no cue-only-based model was significant, indicating that participants did update their expectations after finding the first target.

Table 3

Expectation definitions and corresponding likelihoods separated by number of targets per trial

Likelihoods	Given information	Event	Values for n_t=1	Values for n_t=2
$L_{CA}(E)$	Trial cue and first target found	Existence of a second target	0.118, .4, .667, .769	0.118, .4, .667, .769, 1.00
$L_A(E)$	First target found only	Existence of a second target	0.18, .75	0.18, .75, 1.00
$L_{CA}(B)$	Trial cue and first target found	Specific target remaining in the trial	0.232, .333, .6, .882	0.118, .231, .333, .4, .6, .667, .769, .882
$L_A(B)$	First target found only	Specific target remaining in the trial	0.25, .82	0.18, .25, .75, .82

2.2.2.1 Response Time Results

All trials were analyzed for which the RT for the first target acquisition was valid, i.e., RT_A greater than 200 ms, correct orientation response, and a valid first target identified. The rationale for these restrictions was that the RT for Search B would only be relevant if we could be confident that they found a target in the first half of the trial.

Response times were analyzed using GLM for repeated measures. Results were analyzed separately for $n_t = 1$ and $n_t = 2$ due to the dependence between number of targets and certain values of the predictor likelihoods. It was predicted that for all trials, higher likelihood values would be associated with longer RTs, as participants would spend more time searching for a more-expected target. This effect was predicted to be mitigated in two target trials, however, because search could also terminate quickly due to target acquisition.

Analyses of response times based on the expectation of the *existence* of a second target indicated that for single target trials, both sets of likelihoods ($L_{CA}(E)$ and $L_A(E)$) were significant predictors of response time. The Akaike information criterion (AIC) values indicate that the model using the likelihoods conditional on the integrated cue and first target ($L_{CA}(E)$) was a better fit to the data. Model results and comparisons can be found in Table 4. The mean RTs for the best-fit model ($L_{CA}(E)$) indicated that as predicted, participants had significantly longer RTs when they had higher expectations of a second target (and could not find one), but surprisingly, the relationship was not monotonic. As shown in Table 5, participants spent more time searching for a target at $L_{CA}(E)$=.118 than for $L_{CA}(E)$=.4. This is addressed in the discussion section.

For two-target trials both sets of likelihoods were also significant predictors of RT. The best-fit model for two-target search trials was also $L_{CA}(E)$, as shown in Table 4. The mean RTs show a non-monotonic pattern, where the longest RT was the middle likelihood value. This pattern was likely due to the two search termination mechanisms operating, one where participants terminated search based on expectation that no additional target was present, and one based on locating the second target. The longer RTs for the higher likelihood values relative to lower values are consistent with the prediction that participants would expend more time and effort to find the target remaining if they had a stronger expectation for its presence.

Table 4

Results of GLM analyses for Search B RTs using likelihoods for existence of a second target as predictors

Variable	DF	F	p-value	AIC
Two-target search				
$L_A(E)$	2/67	5.5	.0062	31697.1
$L_{CA}(E)$	4/121	2.85	.0266	*31666.4
One-target search				
$L_A(E)$	1/35	46.89	<.0001	17272.3
$L_{CA}(E)$	3/105	16.55	<.0001	*17239.7

Notes. *Indicates best fitting statistical model. Note that models for one- and two-target search cannot be compared because they are based on different data.

Table 5

Mean RTs broken out by likelihood based on cue and first target

One-target trials			Two-target trials		
$L_{CA}(E)$	Mean (ms)	Std Err.	$L_{CA}(E)$	Mean (ms)	Std Err.
0.118	6595.0	246.5	0.118	3658.8	687.0
0.4	5722.3	559.6	0.4	4190.6	714.3
0.667	8754.9	673.3	0.667	5824.0	543.1

0.769	9404.2	391.7	0.769	5596.9	239.9
			1	5125.7	128.1

Analyses of response times based on the expectation of the *specific* second target were only performed on two-target trials, because results for one-target trials would be the same as for the previous analysis for one-target trials. Both sets of likelihoods, $L_{CA}(B)$ and $L_A(B)$, were significant predictors of RT, as shown in Table 6. The model based on both cue and the first target ($L_{CA}(B)$) was a better fit to the data. As with the models for existence of a second target, the pattern of mean RTs was not monotonic with likelihood values, shown in Table 7. The means and standard errors were therefore calculated for each target set and target order to understand the response times for these analyses better, as shown in Table 8. A comparison between this table and Table 2 indicate that although C3-C4 trials and C4-C3 trials had the same conditional likelihood for the second target, RTs for C4-C3 were faster than C3-C4, indicating a qualitatively different search process between the two. This is addressed further in the discussion section.

Table 6

Results of GLM analyses for Search B RTs using likelihoods for the specific, remaining target as predictors

Variable	DF	F	p-value	AIC
Two-target search				
$L_A(B)$	3/104	4.00	.0097	31683.1
$L_{CA}(B)$	7/237	2.06	.0481	*31622.3

Table 7

Search B Response Times based on likelihoods for specific second target, conditional on T1.

	RT (ms)	
Likelihood	Mean	SE
.118	4760.7	534.8
.231	5712.7	372.6
.333	6150.7	678.6
.400	5236.4	534.8
.600	5167.7	434.5
.667	5649.8	376.8
.769	5358.6	172.7
.882	4630.3	206.0

Table 8

Second Target DVs based on Target 1 ID. RT also based on Target 2 ID.

Target		RT (ms)	
T1	T2	Mean	Std. Err.
1	0	9240	290
1	2	5634	239
2	1	5174	244
2	2	5814	355
3	0	6453	193
3	4	3914	538
4	3	6516	637
4	4	4729	202

2.2.2.2 Fixation Results

Analyses for fixations were performed using GLM, with a logic link function and a binomial distribution. Similar to the RT data, two models were run, one using color and the first target identity as predictors, and one using color, cue, and first target identity as predictors. Trials were only discarded if the identity of the first target could not be ascertained, e.g., if there was no record of a fixation to a target during Search A. This accounted for 11.5% of the data and was primarily due to trials in which the eye-tracker did not record data. It was predicted that the relative proportion of fixations to each color would be based on expectations generated by the first target identified. Thus, if the participant found a C1 target in the first half of the trial, they would be likely to search primarily C2 objects for a second target.

The GLM with the logit link and a binomial distribution uses generalized estimating equations, which is not a likelihood-based method. Therefore, rather than using the AIC to compare the expectation models, the QIC, (quasi-likelihood under the independence model criterion) was used to determine the best fitting model (SAS Institute, 2017). The model using color and first target identity was a better fit to the data (QIC=5019) than the model using color and first target identity *and* cue (QIC=5056). Using the model with only color and the first target identity as predictors, the interaction between color and first target identity was significant, χ^2 (9) = 27.70, p=.0011. This indicates that the proportion of fixations to each color depended on the identity of the first target. The main effects of color and first target were also significant: color χ^2 (3) = 23.27, p<.0001, and first target χ^2 (3) = 18.08, p=.0004. These main effects were most likely due to the underlying ecology of the experiment, as some colors had more two-target trials than others did, and target identity was correlated with color. This ecology characteristic could account for higher numbers of fixations for those two predictors. For each color, conditional likelihoods of fixating that color compared to any other color were calculated for each first target identity. Conditional likelihoods for fixations to each color based on the first target found are shown in Figure

2, superimposed on the ecological conditional likelihoods. The data show that participants were sensitive to the likelihoods for the specific target remaining. The fact that C1 and C2 were generally both higher or lower than C3 and C4 also indicates that participants learned the underlying disease, for example, disease-1 only has C1 or C2, never C3 or C4, and vice-versa for disease-2.

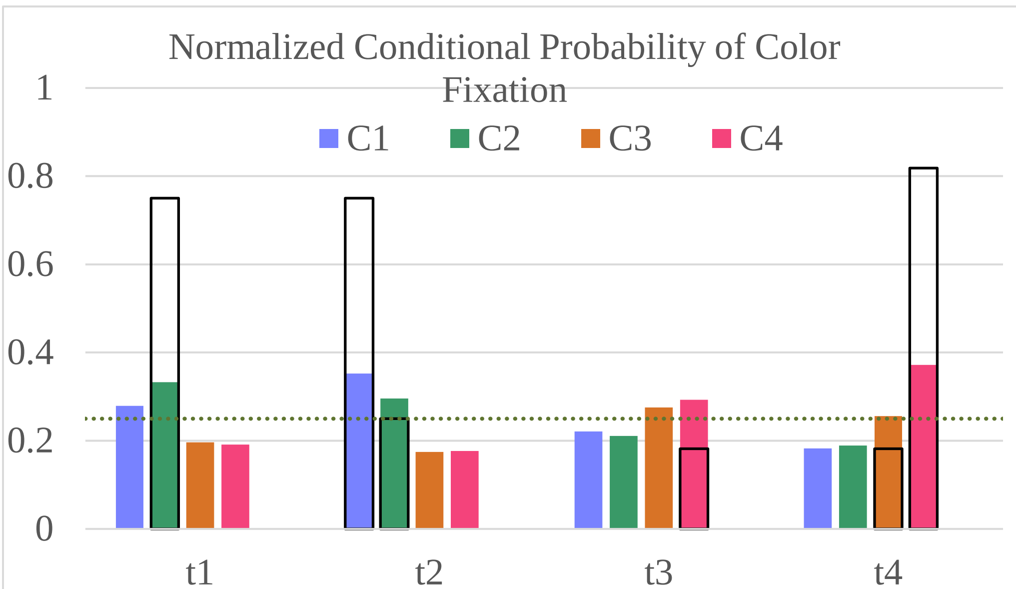

Figure 3. Conditional probability of selecting one color object (C1-C4) over any other color, given the identity of the first target found (T1-T4). The gray dotted line represents chance. The black outlines represent the ecological conditional probability.

2.2.2.3 <u>Miss Error Results</u>

Miss errors were defined as either a) participants never fixating the target or b) participants responding that no second target was present on two-target trials. The distinction between the two definitions is that the participant may have fixated the target in the latter condition but not recognized it as a target. It was predicted that lower expectations for the existence of a target or for the specific target remaining would result in a higher chance of error. Miss errors were analyzed using GLM with a logit link function and a binomial distribution and were only analyzed for two-target trials.

Expectations for the *existence* of a second target significantly predicted the likelihood of committing a miss error, for both models for expectation $L_{CA}(E)$ and $L_A(E)$. Model $L_{CA}(E)$ had the best fit using the QIC, as shown along with the statistical analysis results in Table 9, indicating that participants utilized both cue and first target information when generating expectations for Search B.

Table 9

Summary of model results for miss errors.

Predictor	DF	χ^2	p-value	QIC
$L_A(B)$	3	71.45	<.0001	1919.5
$L_{CA}(B)$	7	77.95	<.0001	1924.2
$L_A(E)$	2	20.53	<.0001	1969.8
$L_{CA}(E)$	4	35.1	<.0001	1965.7

On trials where participants had low expectations for a second target ($L_{CA}(E)$=.118 or $L_{CA}(E)$ = .4), they were roughly three times more likely to commit a miss error compared with trials on which they had high expectations for a second target (odds=3.326, p<.0001), suggesting that lower expectations degrade search. . Interestingly, this effect is primarily due to $L_{CA}(E)$=.4 trials, corresponding to the cue-1-C3-C4 target set, a point which is addressed in more detail in the discussion section. It was predicted that low expectations would cause errors more frequently, but it is unclear why a mid-range likelihood for a specific cue-target set permutation would cause a more significant number of errors than sets with lower likelihoods.

Expectations generated for the particular target remaining in the search array also significantly predicted odds of miss errors for both models ($L_A(B)$ and $L_{CA}(B)$). In these analyses, $L_A(B)$ was the best fitting model as shown in Table 9 above, and indeed, the best fitting model of the four. Participants were significantly more likely to commit a miss error with lower expectations for the target and they were less likely to commit a miss error with higher expectations for that target. Odds ratios for each of the four possible $L_A(B)$ expectation values were not monotonic. On trials for which the likelihood of the remaining target was $L_A(B)$=.25, participants were more likely to make an error compared with trials for which the likelihood was $L_A(B)$=.18. This can be seen in Table 10.

Table 10

Odds of Error based on likelihoods for a specific second target, based on the first target

Likelihood	Odds	SE	p-value
0.18	1.771	0.372	0.006
0.25	2.512	0.433	<.0001
0.75	0.551	0.061	<.0001
0.82	0.408	0.051	<.0001

Notes. Odds ratios are for each likelihood versus all others.

To explore the non-monotonicity further, additional post-hoc analyses were run to examine odds of committing an error for various conditions. It should be noted, however, that these analyses partitioned the error trials into highly specific categories, where each category had relatively low number of trials. In addition, due to the exploratory nature of the investigation, family Type-1 error was not controlled. The full results are shown in

Table 11. The most interesting comparisons are to consider cases in which the only difference in the trial type was the order in which the targets were found. For trials with C1 and C2, the cue and the order that the targets were identified (which in turn would affect likelihoods), did not significantly increase the odds of committing an error. However, for trials with C3 and C4, cue and target order had a significant impact on the likelihood of committing an error, with cue-1-C3-C4 trials having a much higher chance of an error than any other combination, including cue-2-C3-C4, and cue-1-C4-C3. These results suggest that although expectations play an important role in predicting odds of committing an error, there are additional nuances and considerations that need to be investigated.

Table 11

Odds of committing an error given the identified cue-target set.

$L_{CA}(B)$	Cue	T1	T2	Odds	Std Err	χ^2	p-value
.118	2	3	4	1.368	.517	.69	.4068
.118	2	4	3	2.129	.837	3.7	.0545
.231	1	2	2	2.363	.4704	18.65	<.0001
.333	2	2	2	.657	.3232	4.13	.0422
.400	1	3	4	5.901	2.0374	26.43	<.0001
.400	1	4	3	.202	.125	6.69	.0097
.600	1	4	4	.766	.1869	1.19	.275
.667	2	1	2	.812	.2179	.6	.4367
.667	2	2	1	.526	.1791	3.56	.0593
.769	1	1	2	.754	.1255	2.89	.0894
.769	1	2	1	.550	.0951	11.95	.0005
.882	2	4	4	.458	.0762	22.02	<.0001

2.3 Discussion

Results indicate that participants' visual search behavior is influenced by their expectations about the targets. Participants were faster to locate the first target when it aligned with their expectations, and they displayed more fixations to the object colors that matched the most likely target color. Participants were also more likely to make errors when they did not expect a target or when the target color was something they did not expect, quite possibly because they did not fixate the low-likelihood colors as often. It is clear, however, that the mechanisms driving search behavior are more nuanced than can be accounted for by simple likelihood values.

The disease-2 trials, with a target permutation of C3-C4 in particular, demonstrate some underlying complexities not accounted for by the likelihood values alone. The ecology was designed so that for most trials where C3 was located in Search A, there would be no additional target, resulting in extremely low likelihood values for a second target. However, response times showed that in single-target trials, participants searched longer in the lowest likelihood condition, $L_{CA}(E)=.118$, than they did for $L_{CA}(E)=.4$ (though it should be noted that both RTs were shorter than those for higher likelihood values). Both of these likelihood values are associated with trials in which the first target found was C3 or C4. The difference between them is the former is associated with cue-2 trials, and the latter with cue-1 trials. The cue-2 trials were the ones most strongly associated with disease-2. Thus, participants would have had higher exposure to *cue-2*-C3-NoT trials than *cue-1*-C3-NoT trials, so my *a priori* prediction was that they would terminate search earlier for cue-2 trials due to a high expectation of there being no additional targets. It is possible, however, that rather than interpreting cue-2-C3 trial RTs as longer than predicted (M=6.6 s), it could be that cue-1-C3 trial RTs are shorter than predicted (M=5.7 s). Perhaps because C3 was an unexpected target for cue-1, participants terminated search exceptionally early after finding it.

This explanation would also help us understand an unanticipated finding in the error analysis for disease-2 trials. Only for cue-1 trials, C3-C4 target permutations were far more likely to have miss errors than C4-C3 trials, despite the fact that the $L_{CA}(B)$ likelihoods were the same for both. Because the same pattern did not occur for cue-2 trials, the result must be related to cue in a manner not accounted for by likelihoods alone. Because disease-2 is not predicted by cue-1, the effort the participants expended in Search A may have influenced their motivation and/or ability to continue search during Search B. Whatever the underlying cause of the differences in RT and errors, it is clear that expectations based only on the likelihood values cannot explain all of search behavior. It seems that a valuable place to start additional investigation would be the role the cue and Search A play in Search B behavior.

Another possibility for the unexpected results identified above, is that they are due to individual differences in participants' ability to learn the ecology. Ecological likelihoods were used to represent participant expectations, but if participants did not learn the ecology well or correctly, their expectations would not match the likelihoods. Additionally, participants could be using a different strategy or heuristic to remember target sets, which could also affect the validity of using likelihood as a proxy for expectation. A limitation of this study is that participants' internalization of the ecology was not explicitly measured, only inferred from their behavior. There are also known differences in eye movement patterns between novices and experts in certain fields (Manning, 2010), so it would useful in the future to be able to map participants to novice or expert categories based on the extent to which they have learned the ecology.

It has been argued that people are less likely to miss second targets if they share perceptual similarities to the first target (perceptual set theory, Biggs et al., 2015). Evidence in support of this theory was not obtained in the present experiment. If perceptual similarity were a factor, participants should be less likely to commit an error for C2-C2 trials compared to C2-C1 trials. In fact, participants were 4.57 times more likely to commit a miss error for C2-C2 trials than C2-C1 trials. This is easily explained by considering the likelihood values from Table 2. Even if perceptual similarity plays a role, the expectations that participants generated from their experience with the target sets overpowered any perceptual effect, suggesting that top down components of visual search can be more important than bottom-up components for predicting search errors. Importantly, the results of this experiment demonstrate that expectations can both facilitate and hinder search performance, depending on if the expectations align with reality. It is therefore extremely important to understand the influence of top-down cognitive mechanisms on visual search.

In order to begin to elucidate the specific cognitive processes that influence the search behaviors shown in this experiment, I developed a computational cognitive model combining expectation or hypothesis generation with visual search behaviors. Cognitive process models are valuable tools to explicitly test theories and assumptions about human cognition. They also allow researchers to gain new insights into the cognitive processes involved in a task and how they interact with each other by generating novel predictions about participant performance. Importantly, cognitive models can not only generate predictions for the original task that the model was designed to fit but also for new tasks. Parameterizing key cognitive processes also allows researchers to explore how changes in a process affect human behavior. For SSM errors, a cognitive process model can provide insight into what processes guide visual search, and more importantly, which processes contribute the most to miss errors. Once the cognitive mechanisms contributing to search errors are better understood, interventions to reduce error can be developed that specifically target these processes.

CHAPTER 3. MODEL

The present hypothesis-guided search (HGS) model integrates a model of hypothesis generation, a modified leaky accumulator model, simple inhibition, and a fixations selection model. The hypothesis generation model is used to simulate how participants use the cue and later the first target found to generate expectations about the remainder of search from prior experience in the task (Thomas, Dougherty, Sprenger, & Harbison, 2008). The accumulator model is used to simulate how participants' expectations change as search progresses. This permits participants that begin a search trial expecting targets from disease-1 to shift their expectations to disease-2 colors after repeated failure to find a target from disease-1. Selected objects that are not targets are inhibited, which results in a penalty to the accumulator corresponding to that object's color. The fixation selection model is a simplified model of visual search[′1], but does not incorporate bottom-up activations or geometric eye movements. It simulates how participants choose an object to fixate, probabilistically based on their expectations for what color(s) the targets may be. A high-level overview of the model is shown in Figure 3. The following sections walk through the details of the model, dividing it into the major sections identified above in the order in which they are called in the model.

[′ 1]Although a more complex model of visual search that included basic visual processes was proposed, the target acquisition model (TAM; Zelinsky, 2008), I determined that TAM in its current state is unable to locate targets from the stimuli and object layout used in the experiment. It was decided that it was beyond the scope of this paper to modify TAM to work properly.

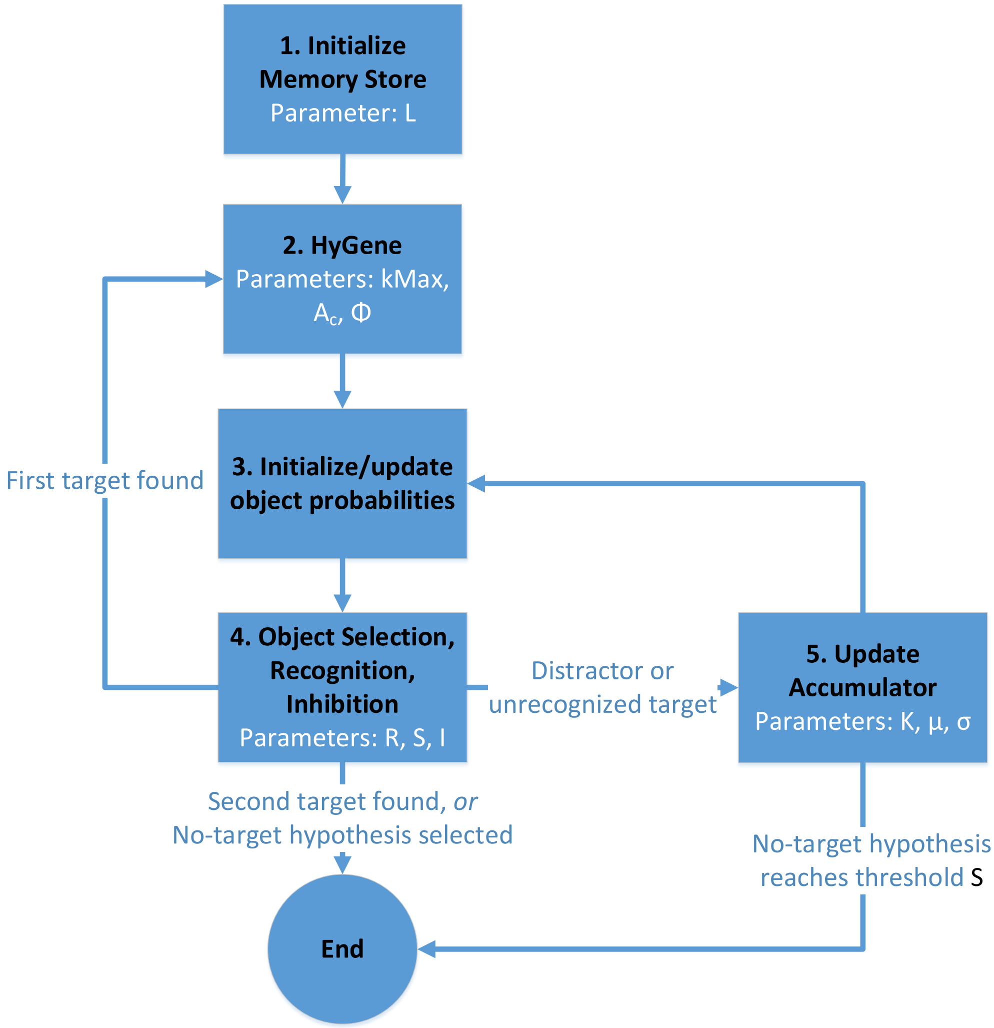

Figure 4. Overview of the Hypothesis-Guided Visual Search Model (HGS).

3.1 Model Explication

3.1.1 Hypothesis Generation Model (HyGene)

The full details of the hypothesis generation model (HyGene) can be found in Thomas et al. (2008), but the model is described below in how it relates to the present HGS visual search model. There are three key elements to HyGene: memory storage, global memory matching, and set of contenders (SOC) generation. Two different versions of the model were tested, one that only utilized the memory store and global matching modules

(HGS-All), and one that also included the SOC generation (HGS-SOC). A third version of the model (Random-GS) eliminated HyGene, effectively eliminating use of experience to aid search. Instead, equal object likelihoods were used to initiate the search model.

At a high level, HyGene takes in data from the external environment (D_{obs}) and uses a global matching method to compare it to an episodic and a semantic store in order to calculate activations for each possible hypothesis in the semantic store. Based on the semantic activations, hypotheses are probabilistically generated into the set of contenders (SOC), a component which is functionally similar to working memory. The contents of the SOC are used for any additional cognitive processes. Traditionally, the aforementioned processes were from the decision-making literature, such as probability or confidence judgements. Here, the SOC output is used to bias object fixation selection. An overview is shown in Figure 4, followed by a more detailed description.

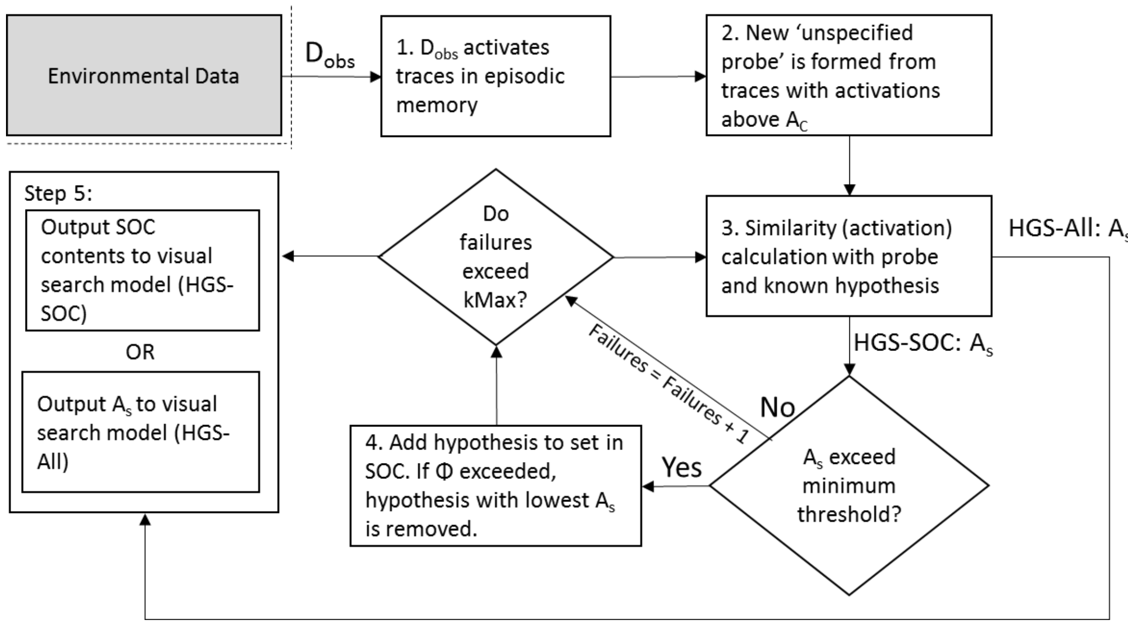

Figure 5. Overview of the HyGene model (Thomas et al., 2008) used in the present hypothesis-guided search model (HGS).

3.1.1.1 Memory Store

HyGene's declarative memory is composed of an episodic store and a semantic store. A schematic of the memory storage is shown in Figure 5. Each store is composed of memory traces, represented by one dimensional arrays, called vectors, populated by the integers -1, 0, and 1. The episodic store is based on the experimental ecology to simulate a participant's experience, whereas the semantic store contains all possible hypotheses to simulate a participant's schematic knowledge. To model the present experiment, the episodic store held separate traces for Search A and Search B, where each trace had a context component concatenated with a target component. In Search A, the context was simply a vector representing the cue. In Search B, the context was a concatenation of a vector representing the cue and a vector representing the first target found. The Search B context vector was done in this way because the best-fitting statistical models of the empirical data incorporated information from both cue and first target.

There were separate trace vectors for every trial that participants saw in the experiment. To mimic imperfect encoding of cue-target sets in the experiment, HyGene employs a parameter (L) that degrades each trace, by changing vector elements to zero with a probability of 1-L. The semantic store traces represent each color that a target can be. Participants generate hypotheses about what color the target(s) will be based on a cue. There is also a semantic trace for no-target, which represents the hypothesis that there is no additional target. The no-target trace only becomes relevant for Search B.

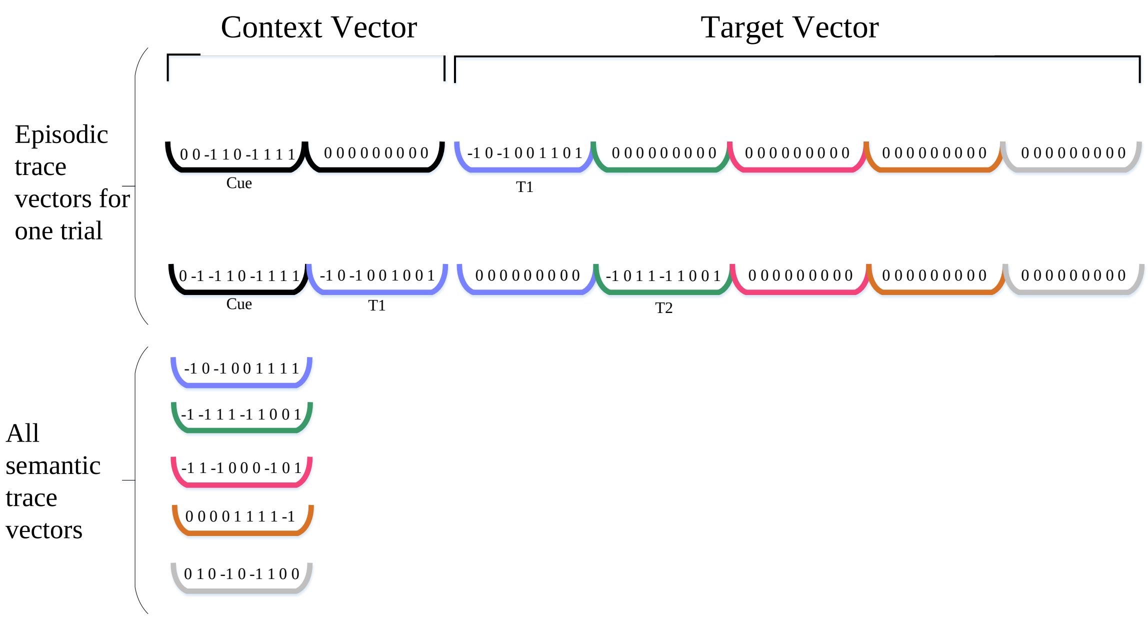

Figure 6. Schematic of implementation of episodic and semantic memory stores. Each color represents a possible target color. The gray mini-vector represents the no-target hypothesis. Note that the mini-vectors in the episodic traces are degraded relative to the semantic traces.

3.1.1.2 Global Matching

In the global matching phase, the HyGene model uses a trace from the external environment and compares it with the context component of each episodic trace. For Search A, the external trace is the cue, and in Search B, the external trace is a combination of the cue and the first target found. The comparison between the external trace and the episodic traces generates an activation value for each trace, calculated using Equation (1) below. Equation (1) calculates the activation for any two vectors of the same length, where N_{12} is the number of elements of $Vector_1 \neq 0$ and $Vector_2 \neq 0$.

$$Activation = \left(\frac{\sum Vector_1 \cdot Vector_2}{N_{12}}\right)^3 \tag{1}$$

Traces above a parameterized minimum threshold (A_c) are combined proportionally with their activation value and condensed into a single probe by performing a weighted sum of all above-threshold traces and then normalizing. The A_c parameter only allows traces that are similar (activated) enough to the current trial to contribute to the semantic retrieval. For this model, A_c was set to a constant .25, allowing traces that were approximately 65% similar (where similarity is the cube root of activation) to be activated and used in the probe. This probe is then compared with each trace in semantic memory, generating semantic activation values. The semantic activations are used for the next component of the model. For the first model that was tested, HGS-All, the model proceeds from here to the fixation selection module. For the second model that was run, HGS-SOC, the next step is the generation of hypotheses into the set of contenders.

3.1.1.3 Set of Contenders Generation

The SOC is functionally similar to working memory. It is capacity-limited, with the capacity set by a model parameter (Φ), and represents the information used in online processes. Because participants' working memory was not measured for this experiment, this parameter was set to a constant of four, which is consistent with the approximate maximum number of hypotheses or chunks of information that a person can maintain (Simon, 1973). In an iterative process, an attempt is made to generate a hypothesis into the SOC, with a generation likelihood for each hypothesis equal to its relative semantic activation. In order to be successfully generated into the SOC, the hypothesis' activation must be lower than that of the hypothesis in the SOC with the current lowest activation, and the currently generated hypothesis must not already be represented in the SOC. These criteria for generation make it so that if one hypothesis is significantly more active than another is, it is more likely to be generated into the SOC. Furthermore, once a highly active hypothesis is generated, few to no competitors are likely to be generated, because they will not have sufficient activation. If any of the generation criteria are not met, a retrieval failure counter is increased. Once the number of retrieval failures surpasses a parameterized threshold (kMax), the generation process ends. The kMax threshold parameter has previously been varied to fit time pressure data, but due to the fact that there was no time pressure in this experiment, this value was set to a constant, kMax = 8. Once the generation process is complete, only the contents of the SOC are used for further cognitive processing.

The effective difference between the HGS-All model and the HGS-SOC model is that the SOC mechanics can impose limits on the number of hypotheses generated, depending on the order in which they were generated. This is in line with current research on human behavior for generation from long-term memory (Thomas, Dougherty, & Buttaccio, 2014). However, this task differs from those used in previous studies with HyGene, in that all possible hypotheses are readily visible on every trial, with each of the four color possibilities visually represented on the screen. Furthermore, this task is not a one-time decision task, but rather an ongoing visual selection process. Given that participants know that there is always at least one target, if they continuously fail to find C1 or C2 targets, despite their high likelihood, they will eventually have to entertain the hypothesis that the target is C3 or C4. The inhibition mechanism discussed below in combination with the accumulators permits hypotheses to change over time, but HGS-All allows the less likely hypotheses to have a higher starting value than when the SOC limits generation. Both models were tested to determine which of them provides the best account of participant behavior.

3.1.2 Fixation Selection and Inhibition

The next portion of the present model is fixation selection, which is depicted in Figure 6. This model assigns a likelihood of fixation to each object visible on the display, and then probabilistically selects an object to be "fixated." To allow for the expectation that no additional target is present in Search B, an invisible additional object is added to the array, representing the no-target hypothesis identified above. If it is drawn, search is likely to be terminated as described in the search termination section. Prior to the initial hypothesis generation, each object is assigned a likelihood equal to $1/nObjects$ (Step 1 in Figure 1). For the Random-GS model, these baseline equal weights are the starting weights, and HyGene is not called. After hypothesis generation, the output of HyGene is used to modify the weights (Step 2). For HGS-All, the output is the normalized semantic activations for each of the five hypotheses. The HGS-SOC model only outputs relative activation values for the contents of the SOC. In both cases, the hypotheses' likelihoods are evenly distributed among the objects of the corresponding color (plus the invisible no-target object) and added onto the existing object activations (Step 3).

The new object weights are renormalized and then used to probabilistically select an object for fixation (Step 4). If the object is a distractor, that item is inhibited (Step 5). The value of inhibition is parameterized (I). If it is a target, the model will recognize it as a target with likelihood equal to the parameter R. This recognition parameter is a temporary means of allowing the model to commit recognition errors. As mentioned earlier, for the present experiment, the definition of SSM errors does not distinguish between scanning and recognition errors. Later exploratory analysis of the data with comparison to this parameter will allow us to develop predictions for the relative likelihood of a scanning error compared with a recognition error.

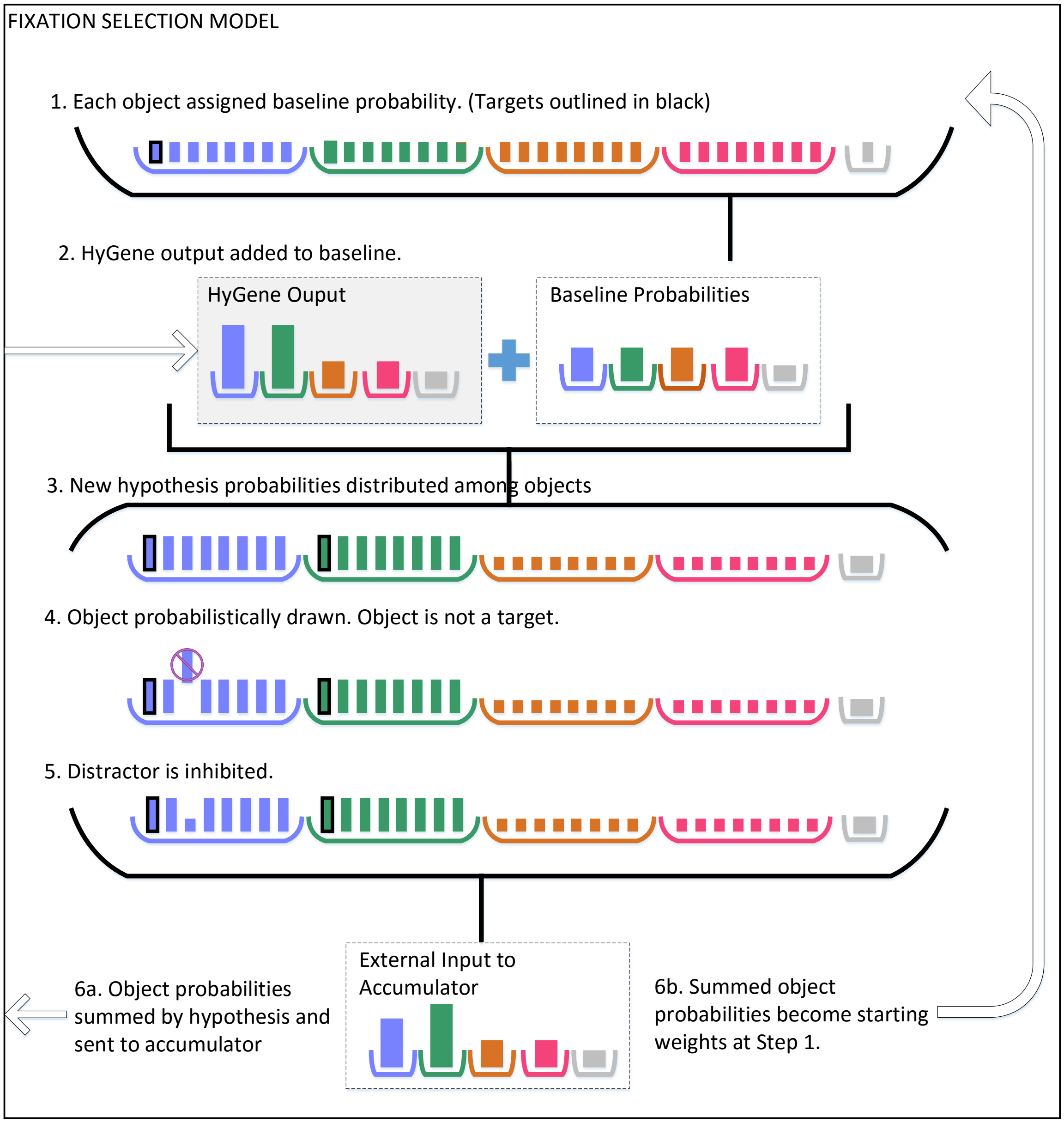

Figure 7. Overview of the visual search portion of the model, including fixation selection and inhibition.

3.1.3 Accumulator

Once an object has been inhibited, the accumulator module is called. The accumulator module uses Equation (2. This equation is a modified version of Equation 4 from Usher and McClelland's leaky, competing accumulator model (LCA; 2001). An important difference between the HGS accumulator and most accumulator models, however, is that there are no thresholds for the hypothesis accumulators, *except* for the no-target hypothesis, discussed more later. The other accumulators here are used to model change in expectations over time and do not directly result in search termination. Usher and McClelland's LCA model can be broken out into four main components, with at least one parameter associated with each: recurrent activation, decay, lateral inhibition, and noise. However, the parameters for recurrent activation and decay are applied to the same value, specifically, the initial accumulator weight. In addition, because the accumulator values sum to one in the HGS model, lateral inhibition is also a function of the starting weight of the accumulator. The three parameters associated with these functions were thus combined into a single parameter, K. *Old weight* in Equation 2 represents the previous output from the accumulator module. *External input* represents the current activations of each object summed across hypothesis color. Thus, for every hypothesis except for the one that had just been inhibited, the external input should be the same as the old weights. Effectively, the combination of the inhibition and the accumulator module allow participant to switch between hypotheses upon continued failed object selections from the original most likely hypotheses.

$$Acc.Weight = K * [Old\ weight] + [External\ input] + Noise(\mu, \sigma) \tag{2}$$

3.1.4 Stopping Rules

There are three basic means of terminating search. The first, most obvious one is when the model has found both targets. This results in immediate search termination. The second option is based on the main prediction of the experiment discussed above as well as the present HGS model. Specifically, when the no-target hypothesis is sampled (Step 4 of Figure 6), there is a probability equal to R that search will be terminated. The third stopping option was implemented when it appeared that the model would essentially time out without finding a second target on too many occasions. This option added a threshold (parameter S) to the accumulator for the no-target hypothesis. When the no-target accumulator reaches this threshold, search is terminated. Higher values of the S parameter would indicate that the second stopping mechanism, the sampling-based stopping rule, is likely being employed more often than the threshold stopping rule. The time out mechanism indicated above was set to 500 iterations, which is well above the maximum, outlying, experimentally obtained number of fixations in a given trial (number of fixations = 347). Model iterations that timed out were not assessed.

3.2 Model Fitting

Due to the high number of parameters, I performed a grid search for best fitting parameters, selecting a small number of reasonable values for each parameter. I tested these parameters for each of the three models described above: HGS-All, HGS-SOC, and Random-GS[2]. The parameters and the values that were tested are listed in Table 12. For model fitting, certain parameters are more associated with individual differences between participants, such as the HyGene parameters. For other parameters, it was unclear whether a given individual would maintain the same parameter value for every trial type. Trial type is defined by the starting cue and every permutation of target set (e.g., cue-1-C1-C2, cue-1-C2-C1, cue-1-C1-noT, etc.). For example, the inhibition value may be higher for trials

[2] Later work will use genetic algorithms to more fully explore the parameter space.

in which participants were more likely to switch between hypotheses quickly. For this reason, and given the large number of model fit outputs for the three models, 36 participants, 16 trial types, and 486 parameter sets identified, model data was fit in two different ways, by collapsing across participant and fitting to trial type, and by collapsing across trial type and fitting to participant.

Table 12

Model Parameters

Parameter	Description	Values Tested in Grid Search		
HyGene				
L	Encoding - higher values indicate better encoding	.25	.55	.85
A_c	Minimum activation - sets threshold on how similar an episodic trace must be to be incorporated into the probe	.25	-	-
kMax	Generation termination - sets threshold on how many retrieval failures are allowed	8	-	-
Φ	Capacity of the SOC	4	-	-
Accum.				
K	Parameter applied to previous accumulator value, combines effects of recurrent activation, decay, and competition	0	.8	1.2
nMu	Mean of the noise added to the accumulator	0	-	-
nSD	Standard deviation of the noise added to the accumulator	.05	.15	-
Visual Search				
I	Inhibition applied to distractors	0	-.1	-.3
R	Recognition probability - model has R probability of accurately recognizing a target or terminating search	.8	.9	1
S	Stopping threshold - if accumulator value reaches S, model proceeds to search termination phase	.5	.75	1

The models generated fixation and error data (but not RT data) for each trial, which were used to compare with participant data. Error data was the proportion of miss errors that occurred for a particular trial type. Fixation data included the number of fixations to each color for Search A and Search B, resulting in eight different fixation measures. Because fitting dependent measures requires each measure to use the same scale, separate fits were calculated for fixation data and for error data.

For each model (HGS-All, HGS-SOC, and Random-GS), the sum of squared deviations (SS) were calculated between participant data and each dependent measure for every combination of the parameter values listed in Table 12. Parameter sets were ranked by the SS values for fixation data and similarly ranked for error data. The model with the lowest combined rank was determined to be the best-fitting model for the condition in question. As mentioned, separate fits were performed based on trial type and based on participant. It was assumed that either of the HGS models would be the best fitting model for each condition, as opposed to the Random-GS model, given that the empirical results supported predictions that participants would extract information from cues and use that to guide their search. Of the two HGS models, the HGS-All model appeared to perform better during model development, so the best fitting parameter sets for HGS-SOC and Random-GS were indexed into the overall HGS-All rankings to determine how they performed, compared to HGS-All. If the best-fit models for HGS-SOC or Random-GS achieved rank 1 in these combined rankings, it would indicate that it was the best-performing model, not the HGS-All model.

3.3 Results

3.3.1 Fit Results

Results indicate that the baseline model, HGS-All, had the highest combined rank for most of the fit analyses performed. HGS-SOC had the second highest combined rank for most of the fit analyses. The ranking results are summarized in Table 13. The high ranks of the HGS models indicate that a model incorporating participant experience and expectations generated from that experience fits the empirical data better than a model that starts with equal likelihoods for each hypothesis. Cognitive models that incorporate participant experience and expectations can therefore help us begin to understand what cognitive mechanisms involved in expectation generation may contribute to search misses. Understanding of these mechanisms can lead to novel predictions about participants' search behavior and then perhaps also lead to novel approaches to interventions to reduce search misses. Exploring the specific parameter values that resulted in the best fit can help guide future research for SSM errors.

Table 13

Summary of rankings for model fit results. How many times each model achieved rank one, two, or three is displayed for each model for each analysis.

		Number of times rank was achieved		
Fit Type	Rank	HGS-All	HGS-SOC	Random-GS
Fit by Participant (out of 36)	1	36	0	0
	2	0	31	5
	3	0	5	31
Fit by Trial Type (out of 16)	1	12	4	0
	2	4	10	2
	3	0	2	14

A complete listing of the parameter values that best fit each participant is listed in Table 14. This table also shows the average sum of squared deviations for the fixation fit and for the proportion of miss errors for each participant. As shown, there was a lot of variation between participants, making it difficult to point to a single parameterized mechanism that may be important in predicting search misses. The variability suggests that it is possible that participants were deploying different processes across the different conditions than are currently represented in the model.

Table 14

HGS-All fit metrics and best fitting parameters for each participant.

	Sum of Sq. Error		Best Fitting Parameters					
ID	Fixations	p(Miss)	L	I	K	nSD	R	S
0	232.0	0.024	.25	-.3	0	.05	.9	1
1	149.4	0.066	.25	-.1	0	.05	.8	.75
2	564.7	0.193	.55	0	0	.15	.9	.5
3	107.9	0.049	.55	-.1	0	.15	.8	1
4	309.1	0.127	.85	-.1	1.2	.15	1	.5
5	134.7	0.161	.55	-.1	.8	.15	.8	.5
6	218.0	0.175	.25	0	0	.15	1	.5
7	334.5	0.027	.25	-.3	1.2	.05	.8	1
8	321.6	0.127	.85	-.1	.8	.05	.8	.5
9	152.2	0.059	.85	0	0	.15	.9	.5
10	267.5	0.059	.55	0	.8	.15	.9	1
11	175.6	0.124	.55	-.3	1.2	.05	1	1
12	111.5	0.160	.85	0	0	.05	.9	.5
13	741.5	0.139	.25	-.3	.8	.15	1	1
14	102.0	0.036	.55	0	1.2	.15	1	1
15	299.3	0.168	.85	-.1	1.2	.15	.9	.5
16	182.2	0.137	.55	-.1	1.2	.15	1	.5
17	243.8	0.074	.55	-.3	0	.15	.9	.5
18	430.5	0.140	.25	-.1	0	.15	.9	1
19	229.7	0.114	.55	0	0	.15	1	.75
20	210.8	0.189	.85	0	0	.15	.8	.5
21	275.2	0.028	.85	-.3	1.2	.15	1	.5
22	146.3	0.242	.85	-.1	1.2	.15	.9	1
23	140.6	0.085	.25	-.3	.8	.05	1	.5
24	339.7	0.092	.85	-.1	.8	.05	1	1
25	127.3	0.129	.25	-.1	0	.05	.8	1
26	219.5	0.088	.85	-.1	1.2	.15	1	1
27	351.8	0.106	.25	-.1	0	.15	.8	1
28	148.4	0.121	.55	-.3	1.2	.05	1	1
29	116.5	0.091	.55	-.1	.8	.15	.8	.75
30	138.1	0.089	.55	-.1	.8	.05	.9	.5
31	121.3	0.142	.85	-.3	0	.15	1	.5
32	152.8	0.156	.25	-.1	0	.15	1	.75
33	118.2	0.134	.55	0	0	.15	.8	.5
34	384.7	0.056	.55	0	0	.15	.9	1
35	550.0	0.203	.85	-.3	1.2	.05	1	.75

An investigation of the best-fitting parameters when model data was fit to trial type, collapsed across participant, reveals some intriguing consistency in parameter values. Full results are shown in Table 15 and Table 16. They have been split by cue due to the large amount of data. As an example, the K parameter for self-recurrent activation, decay, and lateral inhibition between hypotheses, was 0 for most of the Rank 1 models, suggesting that this construct may not be necessary to model. Relatedly, however, the standard deviation of the noise parameter in the accumulator was most frequently the higher value of .15. Higher noise but no K-parameter may indicate that the Accumulator model in its current state does not explain participant data very well.

Looking at Table 14, Table 15, and Table 16 together, it appears that some broad additional conclusions about the parameters can be drawn. First, participants' data generally reflects at least a moderate encoding value (L), supporting the concept that they are able to learn cue-target set combinations. Second, most participants and trial types were best fit by a model that included inhibition, suggesting that top-down inhibition of unsuccessful hypotheses is a promising construct for modeling participant behavior. The variation in the stopping rule, and the fairly high number of the lowest stopping rule value (.5) indicate that stopping mechanism of sampling the no-target hypothesis is not sufficient to explain participants' stopping behavior. This is particularly noticeable in the variation in the different trial conditions. The sampling stopping mechanism was designed to be able to account for all trial types, which ideally would have led to a consistently high stopping mechanism for all trial types. More investigation and further development of the model are necessary to gain better insights.

Table 15

Model fit statistics and best fitting parameters for Cue 1 trials.

		Mean SS Error		Best Fitting Parameters					
Trial Type	Model	Fixation	p(Miss)	L	I	K	nSD	R	S
Cue 1									
C1-noT	HGS-All	1455.0	-	.55	-.3	0	.15	.8	.5
	***HGS-SOC**	**650.9**	**-**	**.55**	**-.3**	**0**	**.15**	**1**	**.5**
	Random-GS	2259.6	-	-	-.3	0	.15	1	.5
C1-C2	***HGS-All**	**154.0**	**.040**	**.85**	**-.3**	**0**	**.15**	**1**	**.75**
	HGS-SOC	154.2	.040	.55	-.1	0	.15	.9	1
	Random-GS	208.7	.040	-	-.3	0	.15	1	1
C2-C1	***HGS-All**	**119.8**	**.046**	**.55**	**-.3**	**0**	**.15**	**.9**	**1**
	HGS-SOC	110.1	.047	.85	-.3	0	.15	1	.5
	Random-GS	288.2	.046	-	-.3	.8	.15	.9	1
C2-C2	***HGS-All**	**213.5**	**.155**	**.85**	**-.3**	**.8**	**.15**	**.8**	**.75**
	HGS-SOC	186.1	.122	.25	-.3	0	.15	1	.5
	Random-GS	262.0	.146	-	-.3	0	.15	.9	.5
C3-noT	HGS-All	421.7	-	.85	-.3	0	.15	1	.5
	***HGS-SOC**	**418.5**	**-**	**.85**	**-.3**	**0**	**.05**	**.9**	**.5**
	Random-GS	1693.1	-	-	-.3	0	.15	.9	.5
C3-C4	***HGS-All**	**399.2**	**.172**	**.85**	**-.3**	**0**	**.15**	**.9**	**1**
	HGS-SOC	399.1	.171	.85	-.3	.8	.15	.8	1
	Random-GS	402.8	.272	-	-.3	1.2	.15	1	.75
C4-C3	***HGS-All**	**683.9**	**.113**	**.85**	**0**	**0**	**.15**	**.8**	**.5**
	HGS-SOC	685.7	.113	.25	-.3	.8	.05	.8	.75
	Random-GS	682.8	.114	-	0	1.2	.05	1	.75
C4-C4	***HGS-All**	**197.2**	**.090**	**.55**	**-.1**	**1.2**	**.15**	**1**	**.5**
	HGS-SOC	202.0	.090	.55	-.3	0	.15	.8	.75
	Random-GS	211.4	.090	-	-.3	.8	.15	1	.5

Notes. Target sets are listed in the order in which the targets were found, making C1-C2 different from C2-C1. Only free parameters are listed in this table. Refer to Table 12 for parameters with fixed values. Bolded lines were Rank 1 models.

Table 16

Model fit statistics and best fitting parameters for Cue 2 trials.

		Mean SS Error		Best Fitting Parameters					
Trial Type	Model	Fixations	p(Miss)	L	I	K	nSD	R	S
Cue 2									
C1-noT	HGS-All	1085.7	-	0.55	-0.3	0	0.15	1	0.5
	***HGS-SOC**	**538.6**	**-**	**0.55**	**-0.3**	**0**	**0.15**	**0.9**	**0.5**
	Random-GS	1761.7	-	-	-0.3	0	0.15	1	0.5
C1-C2	***HGS-All**	**245.6**	**0.094**	**0.25**	**-0.3**	**1.2**	**0.05**	**1**	**0.75**
	HGS-SOC	245.7	0.094	0.55	-0.3	0	0.05	1	1
	Random-GS	248.8	0.095	-	-0.1	0.8	0.15	1	1
C2-C1	***HGS-All**	**236.1**	**0.108**	**0.25**	**-0.3**	**0**	**0.15**	**0.9**	**0.75**
	HGS-SOC	244.5	0.108	0.85	-0.1	0	0.15	1	0.75
	Random-GS	239.3	0.108	-	-0.3	0.8	0.15	0.9	1
C2-C2	***HGS-All**	**383.2**	**0.231**	**0.25**	**-0.3**	**0.8**	**0.05**	**1**	**1**
	HGS-SOC	382.7	0.227	0.55	-0.3	0	0.15	1	0.75
	Random-GS	382.7	0.243	-	-0.3	0	0.15	0.9	0.5
C3-noT	HGS-All	335.3	-	0.85	-0.3	0	0.15	1	0.5
	***HGS-SOC**	**213.6**	**-**	**0.55**	**-0.3**	**0**	**0.15**	**0.8**	**0.5**
	Random-GS	3641.0	-	-	-0.3	0	0.15	0.9	0.5
C3-C4	***HGS-All**	**544.8**	**0.227**	**0.55**	**-0.1**	**1.2**	**0.05**	**0.8**	**1**
	HGS-SOC	545.0	0.227	0.55	0	1.2	0.05	1	1
	Random-GS	565.0	0.228	-	-0.3	0	0.15	1	1
C4-C3	***HGS-All**	**589.8**	**0.226**	**0.55**	**-0.1**	**1.2**	**0.15**	**0.9**	**0.5**
	HGS-SOC	590.0	0.234	0.55	-0.3	0	0.15	0.9	1
	Random-GS	589.8	0.234	-	-0.3	0	0.15	1	0.75
C4-C4	***HGS-All**	**164.8**	**0.026**	**0.55**	**0**	**0.8**	**0.15**	**0.8**	**1**
	HGS-SOC	175.7	0.026	0.55	-0.3	0	0.15	1	1
	Random-GS	508.1	0.026	-	-0.1	0.8	0.15	1	0.75

Notes. Target sets are listed in the order in which the targets were found, making C1-C2 different from C2-C1. Only free parameters are listed in this table. Refer to Table 12 for parameters with fixed values. Bolded lines were Rank 1 models.

3.3.2 Comparison with Empirical Data

To assess how well the best-fit model actually performed, the best-fitting parameter sets for each participant were used to generate data from the HGS-All model. To give the models a better chance, each participant's best parameter set was run ten times, reducing some of the vagaries due to the random effects in the model. The HGS-All data was then subjected to the same statistical analyses that were performed for the participant data for fixations and miss error likelihoods. The model achieved all of the main effects reported earlier. It also replicated many of the patterns of the actual data, including odds ratios and relative fixation proportion frequencies. In general, it performed more similarly to analyses based on likelihoods for the *existence* of a second target rather than those for the specific target remaining. It did not produce the counter-intuitive results seen with the C3-C4 target sets described in the experiment's discussion section. More investigation and model development is required to determine what may be causing that result.

Table 17 shows the results from the HGS-All fixation data along with the comparable empirical data results.

Table 17

Comparison of HGS-All model and empirical data for fixation data

	Model Data			Empirical Data		
	χ^2 Test			χ^2 Test		
Statistical Test	DF	χ^2	p-value	DF	χ^2	p-value
Search A						
Cue	1	1.1	0.3035	1	1.35	.246
Color	3	11.4	0.0099	3	2.75	.432
Cue * Color	3	224.8	<.0001	3	23.42	<.0001
	Odds Ratios			Odds Ratios		
	Odds	p-value		Odds	p-value	
Cue 1: (C1 or C2) vs (C3 or C4)	1.812	<.0001		1.592	<.0001	
Cue 2: (C3 or C4) vs (C1 or C2)	1.747	<.0001		1.685	<.0001	
	χ^2 test			χ^2 test		
Search B	DF	χ^2	p-value	DF	χ^2	p-value
Target 1	3	100.2	<.0001	3	18.08	0.0004
Color	3	163.7	<.0001	3	23.27	<.0001
T1 * Color	9	203.5	<.0001	9	27.7	0.0011

As shown above, the HGS-All model successfully reproduces the omnibus assessments of the fixation data, though it appears that the model may display a bias toward some colors, given that there was a main effect of color in Search A. Figure 7 below compares the relative likelihood of fixation for participants to the model's performance on the same metrics. As shown, if anything, the model does not utilize expectations when selecting fixations to the extent that participants do, suggesting that there may be too much inhibition or too much noise in the model with the current parameter values.

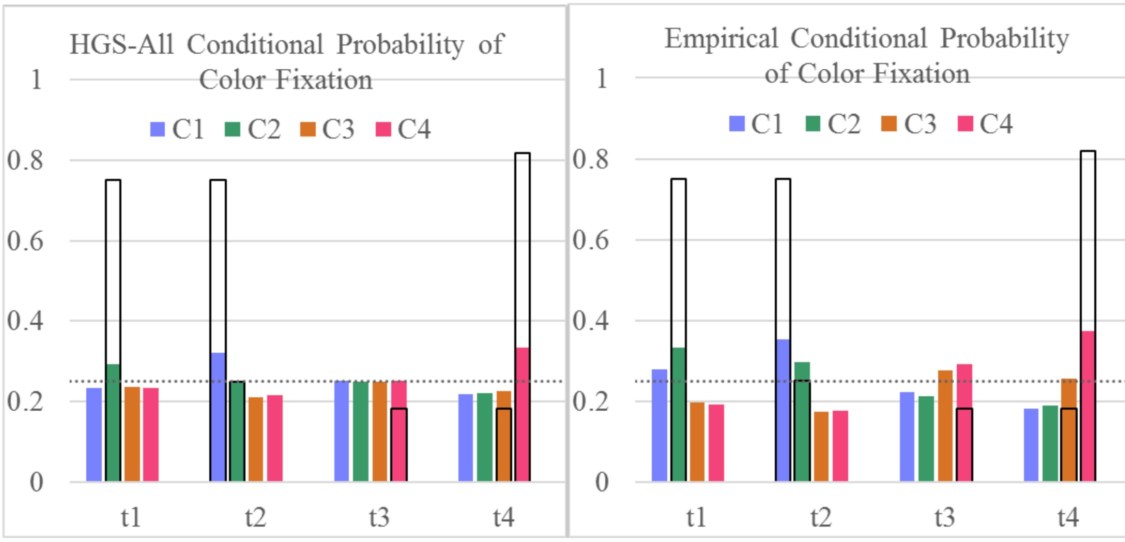

Figure 8. Conditional probability of fixating each color given the first target found for the HGS-All model and the empirical data

Table 18 shows the results of the statistical analyses for the error data along with the empirical results for comparison. As shown, although the HGS-All model had the same overall omnibus results, the best-fitting model in this case was based on both cue and first target and was for the expectation of any second target (L_{CA}(E)), rather than a specific second target. This indicates that the model can represent broad error data reasonably well, but is missing some of the nuances of how participants generate expectations for specific targets. Table 19 compares the odds ratios for the model with the odds ratios for the empirical data, using the best-fitting empirical statistical model, L_A(B), and the best-fitting HGS-All statistical model, L_{CA}(E).

Table 18

Model Comparison for Miss Errors.

	HGS-All Model Data				Empirical Data			
Predictor	DF	χ^2	p-value	QIC	DF	χ^2	p-value	QIC
$L_A(B)$	3	270.3	<.0001	19276	3	71.45	<.0001	1919.5*
$L_{CA}(B)$	7	293.8	<.0001	19177	7	77.95	<.0001	1924.2
$L_A(E)$	2	339.05	<.0001	19069	2	20.53	<.0001	1969.8
$L_{CA}(E)$	4	352.3	<.0001	19010*	4	35.1	<.0001	1965.7

Note. QIC values cannot be compared between model and empirical data. *Indicates best-fitting statistical model

Table 19

Model Comparison for Odds of Error

	Model Odds Ratios			Empirical Odds Ratios		
Likelihood	Odds	SE	p-value	Odds	SE	p-value
$L_A(B)$						
0.18	2.672	.178	<.0001	1.771	0.372	0.006
0.25	.911	.047	.0730	2.512	0.433	<.0001
0.75	.812	.037	<.0001	0.551	0.061	<.0001
0.82	.506	.025	<.0001	0.408	0.051	<.0001
$L_{CA}(E)$						
0.118	3.568	.408	<.0001	1.074	.416	.854
0.400	2.148	.230	<.0001	5.732	1.910	<.0001
0.667	1.008	.096	.9327	.590	.152	.0400
.769	.440	.039	<.0001	.542	.116	.004
1.00	.294	.020	<.0001	.508	.101	.0007

Notes. Odds ratios are for each likelihood versus all others.

3.4 Model Discussion

The results for the models reported here represent early efforts to develop a robust model of hypothesis-guided visual search that can be used to identify the cognitive mechanisms involved during secondary search misses. Once the cognitive mechanisms are identified, interventions can be developed to help mitigate their impact in causing SSM errors. Model fitting results and comparisons to empirical data suggest that the hypothesis-guided search model is able to capture visual search behavior at a high level. Now that this baseline has been established, work can be done to refine the model and begin to replicate some of the more complex aspects of the empirical data.

One of the first steps to improve the model will be to test it against several different mechanisms of generating expectations, such as base rate information or the raw likelihood values. It could be that a simpler heuristic than HyGene can capture the data, but it seems likely that the most successful version of expectation generation will need to consider higher order cognition somehow. HyGene provides a sophisticated way of describing participant expectations and allows us to dig down into the possible cognitive mechanisms, such as retrieval and working memory, involved in influencing visual search behaviors.

Another important step is to perform a detailed assessment of how changes in the parameter values affect search performance, whether these changes improve the model's ability to fit the empirical data or degrade it. For the initial model assessments, a simple grid search was performed to fit the model to the data, using only two to three values per parameter. A more in-depth analysis will likely lead to better model fits, which may better account for some of the subtler aspects of the empirical data, such as the differences between disease-1 and disease-2 search behavior. In addition, I plan to explore model changes to areas of apparent weakness in the current HGS model. For example, the model currently resets once an initial target is found, using the cue and the just-found target as inputs into HyGene. Based on the empirical results, there may be some carry-over effects from the cue that cannot be accounted for by the simple likelihoods for additional targets.

Another advantage to systematically manipulating parameter values at a finer scale than what was used in the present model assessment is that it will help to reveal trends in search performance based on parameter changes. It will be especially important to identify the parameter trends that increase likelihood of search errors, which could lead to novel, testable predictions about what would cause participants to miss more targets in practice. Much work remains to fully test the model and identify its strengths and weaknesses in its ability to fit or predict participant performance. Although the hypothesis-guided search model is still in the early stages of assessment, its ability to capture the high-level empirical results is a promising sign of its future utility in identifying causes and interventions for SSM errors.

CHAPTER 4. GENERAL DISCUSSION

Subsequent search misses have been studied for decades, particularly in the field of radiology, where they can have costly consequences. Most of the work to date has investigated the mechanics of the visual search process to determine how the errors happen. Studies have explored characteristics such as dwell times for identified and missed targets, saccadic characteristics of experts and novices, and individual differences in sensitivity and bias to better understand what happens during SSM errors, but there is still very little understanding of why these errors occur and what cognitive processes are involved. Without an understanding of why errors occur, it is difficult to develop effective means of reducing them, and indeed, very few interventions tested to date have significantly reduced SSM errors (Berbaum et al., 2010; Cain et al., 2013).

For the present experiment and corresponding cognitive process model, I proposed a theoretical framework that can begin to improve our understanding of the higher-level cognition involved in visual search behaviors and errors. I predicted that observers would use external cues, as well as their experience and knowledge of the ecological structure of the environment to generate expectations or hypotheses about what target(s) may be present. I further predicted that observers would use these expectations to guide their visual search. Observers would be more likely to fixate items corresponding to their expectations and would expend more time and effort searching for targets when they had higher certainty about their identity and existence. Under this framework, when observers' expectations aligned with the true state of the world, search would be more accurate, but if the true state of the world were unexpected, observers would be more likely to miss secondary targets.

The present experiment generally supported the main predictions that response times would be shorter for low expectations of a target, that likelihood of committing an SSM error would increase with decreasing expectations of a target, and that fixations would be biased toward objects corresponding to the hypotheses that participants were likely to be favoring, based on their trained experience. One limitation of the experiment, however, is that the ecological target likelihoods were used as a proxy for expectation, and the results of the experiment for certain types of search trials did not perfectly align with these ecological likelihoods. It would then appear that either other mechanisms are at work or that a different way of representing participants' expectations should be investigated. For example, there may be characteristics of the initial part of search (Search A in the experiment) that affect behavior in subsequent search (Search B) that are not completely accounted for by the likelihoods. Examples of these other characteristics of Search A may be length of search or effort expended.

These other possible mechanisms can be implemented and tested in future iterations of the cognitive model of hypothesis-guided search that was presented here (HGS). The present HGS model has already successfully reproduced the higher-level results of the empirical experiment. Yet comparisons with participant data have already suggested possible improvements to the model to bring it more in line with the empirical results. For example, as addressed above, the best-fitting parameters for the accumulator module of HGS indicate that it is not necessarily functioning as expected, with a key parameter often fitting best with a value of zero, and a high noise parameter value being the most frequently ranked as best-fitting. This suggests that a simpler module or a different method of simulating change in predictions over time would better serve the HGS model. Additionally, results of the present experiment will help guide further development of two of the new parameters introduced in this model, namely the recognition parameter (R) and the stopping threshold (S). The fact that they appear to change between trial types suggests that there is some information that participants may be picking up from the trial that is changing their cognition in ways that are not currently captured by the model. Future iterations of the model will seek to be able to answer what it is about certain type of trials that cause a lower stopping threshold and thus earlier search termination than predicted by the hypothesis likelihoods. Future iterations of the model will also differentiate scanning and recognition errors and seek to compare the model predictions to the empirical results.

One of the limitations of the present model and experiment is that it requires highly controlled environmental ecologies. In order to make accurate predictions about participant behavior, the HGS model has to be trained with the same experience as participants receive. In applied domains, it is often impossible to control or to totally account for participants' experience and is therefore difficult to program that experience into the model. It is thus highly challenging to predict the hypotheses that they would generate given external cues. In addition, laboratory experiments may not be generalizable to the applied domains. The results of the present experiment suggest detrimental outcomes to visual search when participants' expectations do not align with the true state of the world. However, studies in radiology have not found a negative impact when observers are provided with a patient's clinical history, even when it does not correctly predict search targets (Berbaum & Franken, 2006). This discrepancy, however, is exactly why a better understanding of the basic cognition underlying SSM errors is necessary. Being able to identify what it is about the task tested in the radiological studies that makes the cognition different from the present laboratory study will help future research into development of effective means of reducing search errors in any domain or task. The HGS model does not need to be tied to a particular experimental paradigm, and it can therefore generate predictions about novel search tasks with different stimuli in other domains. In fact, this model of observer expectations' influence on visual search may explain an effect found in single target visual search studies called the low prevalence effect (Wolfe et al., 2009). Simply, research has shown that participants are more likely to miss targets if they have low prevalence. The HGS model can account potentially account for this effect using experience-based expectation generation.

The long-term goal of the work presented in this thesis is to identify effective interventions for reducing visual search errors of all types. Visual search is a crucial aspect of many important domains, including the oft-mentioned radiology, but also including baggage screening, search and rescue, and other types of medical imaging. The results of the present experiment and model suggest that one way of reducing error is to identify ways of adjusting observer expectations. A decision support tool is one way this could potentially be implemented, where the tool could direct observer attention to hypotheses that it has reason to believe the observer may not be considering. The tool could measure certain characteristics of the observer's behavior to infer their mental state, or it may be trained to have similar experience as the observer, thus allowing it to predict what hypotheses the observer has generated and which hypotheses to which they may not be giving sufficient attention. A simpler intervention, however, may be simply to have a second observer study the same stimulus without any obvious external cues (such as a patient's clinical history), or with a cue that would bias them toward less likely hypotheses. More experimental and modeling work needs to be done to determine the most effective means of adjusting observer expectations, and to determine what other possibilities exist for effective interventions. The present results, however, indicate that we are heading the right direction.

REFERENCES

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2013). Self-Induced Attentional Blink: A Cause of Errors in Multiple-Target Search. *Psychological Science*, *24*(12), 2569–2574.

Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous Control of Attention by Multiple Working Memory Representations. *Psychological Science*, *23*(8), 887–898. http://doi.org/10.1177/0956797612439068

Berbaum, K. S., & Franken, E. A. (2006). Commentary: Does clinical history affect perception? *Academic Radiology*, *13*(3), 402–403. http://doi.org/10.1016/j.acra.2005.11.031

Berbaum, K. S., Franken, E. A., Caldwell, R. T., & Schartz, K. M. (2006). Can a checklist reduce SOS errors in chest radiography? *Academic Radiology*, *13*(3), 296–304. http://doi.org/10.1016/j.acra.2005.11.032

Berbaum, K. S., Franken, E. A., Caldwell, R. T., & Schartz, K. M. (2010). Satisfaction of search in traditional radiographic imaging. In E. Samei & E. A. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (pp. 107–138). New York: Oxford University Press.

Berbaum, K. S., Franken, E. A. J., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., … Lu. (1990). Satisfaction of search in diagnostic radiology. *Investigative Radiology*, *25*(2).

Berlin, L. (1997). Malpractice Issues in Radiology: Perceptual Errors. *ARRS*, (October), 943–946.

Biggs, A. T., Adamo, S. H., Dowd, E. W., & Mitroff, S. R. (2015). Examining perceptual and conceptual set biases in multiple-target visual search. *Attention, Perception, & Psychophysics*, *77*(3), 844–855.

Buttaccio, D. R., Lange, N. D., Thomas, R. P., & Dougherty, M. R. (2014). Using a model of hypothesis generation to predict eye movements in a visual search task. *Memory & Cognition*, *43*(2), 247–265.

Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2013). A taxonomy of errors in multiple-target visual search. *Visual Cognition, 21*(7), 899–921.

Cain, M. S., & Mitroff, S. R. (2013). Memory for found targets interferes with subsequent performance in multiple-target visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(5), 1398–1408.

Donald, J. J., & Barnard, S. A. (2012). Common patterns in 558 diagnostic radiology errors. *Journal of Medical Imaging and Radiation Oncology, 56*(2), 173–178.

Hout, M. C., Walenchok, S. C., Goldinger, S. D., & Wolfe, J. M. (2015). Failures of Perception in the Low-Prevalence Effect : Evidence From Active and Passive Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 977–994.

Institute, S. (2017). QIC goodness of fit statistic for GEE models. Retrieved January 1, 2017, from http://support.sas.com/kb/26/100.html

Körner, C., & Gilchrist, I. D. (2007). Memory processes in multiple-target visual search. *Psychological Research*, *72*(1), 99–105.

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology*, *13*, 175–181.

Manning, D. J. (2010). Cognitive factors in reading medical images. In E. Samei & E. A. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (pp. 91–106). New York: Cambridge University Press.

Manning, D. J., Ethell, S., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography*, *12*(2), 134–142. http://doi.org/10.1016/j.radi.2005.02.003

Nodine, C. F., Mello-Thoms, C., Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology*, *179*(4), 917–923. http://doi.org/10.2214/ajr.179.4.1790917

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(5), 1275–1291. http://doi.org/10.1037/a0013896

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: when it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. http://doi.org/10.1016/j.tics.2011.05.004

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, *4*(3–4), 181–201. http://doi.org/10.1016/0004-3702(73)90011-8

Soto, D., & Humphreys, G. W. (2007). Automatic Guidance of Visual Attention From Verbal Working Memory. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 730–737. http://doi.org/10.1037/0096-1523.33.3.730

Stroud, M. J., Menneer, T., Cave, K. R., & Donnelly, N. (2012). Using the dual-target cost to explore the nature of search target representations, *38*(1), 113–122.

Thomas, R. P., Dougherty, M. R., & Buttaccio, D. R. (2014). Memory Constraints on Hypothesis Generation and Decision Making. *Current Directions in Psychological Science*, *23*(4), 264–270. http://doi.org/10.1177/0963721414534853

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, *115*(1), 155–85. http://doi.org/10.1037/0033-295X.115.1.155

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(3), 459–478. http://doi.org/10.1037/0096-1523.16.3.459

Tuddenham, W. J. (1962). Visual Search, Image Organization, and Reader Error in Roentgen Diagnosis. *Radiology*, *78*(5), 694–704.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. http://doi.org/10.1037/0033-295X.108.3.550

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated Models of Cognitive Systems* (1st ed., pp. 99–119). New

York: Oxford University Press.

Wolfe, J. M., Horowitz, T. S., Wert, M. J. Van, Kenner, N. M., Place, S. S., & Kibbi, N. (2009). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology*, *136*(4), 623–638. http://doi.org/10.1037/0096-3445.136.4.623.Low

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