The Dissertation Committee for Anqi Zhang
certifies that this is the approved version of the following dissertation:

Human Visual Detection and Search in Natural Backgrounds

Committee:

Wilson S. Geisler, Supervisor

Ernst-Ludwig Florin, Co-supervisor

José R. Alvarado

Michael P. Marder

Human Visual Detection and Search in Natural Backgrounds

by
Anqi Zhang

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin
December 2024

Dedication

To:

Jesus the Son of the Living God,

my Lord and my God,

who purposefully created me in His image,

my King and my Shepherd,

who searched for me and found me,

my Judge and my Savior,

who forgives me and cleans my heart,

my Life and my Light,

who faithfully sustains me with constant goodness,

my Friend and my Love,

who loves me and never leaves me,

my Joy and my Peace,

who awes me with countless marvelous works,

and my Source of Hope,

who daily returns my attention to eternity.

Epigraph

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

—A paraphrase of Reverend Thomas Bayes, 1763

An Essay Towards Solving a Problem in the Doctrine of Chances

THOMAS

Unless I see in His hands the imprint of the nails, and put my finger into the place of the nails, and put my hand into His side, I will **not believe**.

JESUS

Place your finger here, and **see** My hands; and **take** your hand and **put** it into My side; and do not continue in disbelief, but **believe**.

THOMAS

My Lord and my God!

JESUS

Because you **have seen** Me, have you **now believed**? Blessed are they who **did not see**, and **yet believed**.

—John 20:25–29

Acknowledgments

I cannot emphasize enough the love and care, nurture and support from my family for me to become the person I am and complete the strenuous work in my PhD program. My loving mother always believes in me and models for me the upstanding work ethics. She sacrificed much of her potentials and opportunities for me, so I could ever receive the education I have been cherishing. My grandmother, though unable to continue after primary education due to brutal war and chaos, always supports me to pursuit my untiring interest in mathematics and natural sciences. I also want to memorize my grandfather, who passed away during my sophomore year in college. He regularly encouraged me to stay curious and seek truth, to live with integrity and stand for justice, and to persevere with contentment.

I want to express my deep gratitude to my research supervisor, Dr. Wilson S. Geisler. As one of the world's best scientists who has been advancing vision science in the last five decades, he insists that I, a graduate student slightly older than his grandson, to just call him as Bill. I am forever inspired by Bill's passion and excellence in pursuing the knowledge and understanding of visual neuroscience, behavior, and computation. Without his extremely patient teaching and mentoring, I will never be able to make serious progresses in Bayesian visual detection and search. I have lost count how many invaluable and insightful discussions we have had, and how many hours and thoughts Bill has sacrificially invested in me to grow as an early career scientist. Furthermore, I am much grateful for his kind support and flexibility with my research work when I unexpectedly experienced the impact of transsphincteric fistula in the past year or two, especially during the periods of surgeries and ensuing recovery.

I offer my thanks to fellow researchers in and beyond my lab who directly or indirectly support the completion of this dissertation, including Calen Walshe, Abrahnil Das, Can Oluk, and Eric S. Seemiller. I commend all participants for their hard work in my experiments, during which much friendship formed and grew; they are Abrahnil Das, Can Oluk, Hayden Stegall, Andrew Eastland, Matthew X. Velez, and Damion Fisher. Furthermore, I thank National Institutes of Health, especially the National Eye Institute, for their financial support to obtain the fundamental knowledge of the human visual system in this dissertation, which will eventually be applied to reduce eye diseases and eliminate vision loss. I also thank the Vision Science Society for hosting the annual conference, where I presented my most results in this dissertation as posters and talks, and received helpful feedback and encouragement.

I am grateful for the strong and close communities at the University of Texas at Austin (UT). Here I highlight the positive impacts on my research work. In the Department of Physics, Dr. Steven Weinberg revived my mathematical skills with his Quantum Mechanics class in my first semester at UT; my co-supervisor Dr. Ernst-Ludwig Florin gave me valuable opportunities for and feedback on academic presentations, supported my advancement to doctoral candidacy at the early period of the COVID-19 pandemic, and processed the extra administrative work for my interdepartmental research; Dr. Philip J. Morrison and Dr. Aaron Zimmerman trained me with fascinating math that I hope to use one day for vision science, in Fluid Mechanics and General Relativity, respectively (shout out to the metric tensor and Riemann curvature tensor!); Dr. José R. Alvarado and Dr. Michael P. Marder kindly served in my dissertation committee. In the Department of Statistics and Data Sciences, Dr. Mary Parker taught me two semesters of Mathematical Statistics with great clarity; Dr. Arya Farahi helped me understand and apply the procedure of

Scientific Machine Learning (later used in Chapter 5). In the Center for Perceptual Systems, Dr. Lawrence K Cormack taught me Bootstrap Statistics; Dr. Robbe Goris broadened my understanding of Systems Neuroscience, and provided feedback on my covert search project. In my student organization Longhorn Chi Alpha Christian Fellowship, I received consistent and abundant love, encouragement and support from Pastor Kelly Brown and many UT students (e.g., Cohovi Aimihoue, Nathaniel Degen, James Albritton, Joseph Richiuso, and Brent Sordo), especially during the most difficult moments of my PhD program, and during the writing "marathon" of this dissertation.

I greatly appreciate my girlfriend Yadira Plata for her unwavering support, unceasing encouragement, and unrevealed prayers. She is one of the first persons who read my dissertation and provided constructive feedback.

Last but foremost, I want to praise God and give all glory to Him. As the Author and Source of all knowledge and wisdom, God faithfully and lovingly guides me to detect, search, and find brilliant and bounded computation of the human visual system. My best friend Jesus "God with us" has been with me through the mountains and valleys of this PhD program, sharing joy and sorrow. Since knowing Him, I have enjoyed even more studying and researching topics in Physics and vision science, because His selfless sacrifice for me (and for the world) and His physical resurrection opened my eyes and heart. I was hopeless, meaningless, and purposeless, but now I know the hope, meaning, and purpose of my life!

Preface

I am excited for you to read this dissertation. As you search for the information that might be relevant and helpful to your work, may you find it quickly, or quit searching neither too early nor too late. As the title points out, in this dissertation we focus on the topic of human visual detection and search in natural backgrounds, summarize the effects of various factors to human visual behavior, and formulate Bayesian decision-making for detection and search tasks.

Growing up, I was quite addicted to computer games. I would constantly search for more (virtual) resources, better strategies, stronger allies, weaker opponents, and easier battles for a final victory. That brought me much experience and made me a "professional" practitioner in visual search, but in digital backgrounds. For some reason, I am already a professional visual searcher in the real world, like many other people, but why and how? Therefore, I chose to re-search this topic to discover and understand the relevant, fascinating mechanisms and computation in the human visual system, and equally importantly, to apply the results to improve visual performance of people and machines for good purposes.

I would never be able to complete this work without the guidance and support of my research supervisor Dr. Wilson S. Geisler. He demonstrated the first-principles approach of applying the ideal observer and its derivatives to vision science research. Our journey of research is loaded with surprises and challenges. We have encountered multiple moments when the freshly analyzed experimental results were simply puzzling. Then we followed up with plans to further investigate and evaluate the reliability/credibility of those surprises, because I do still make coding mistakes. Even-

tually, we overcome. This scientific process iterates just like the theorem of Reverend Thomas Bayes. Intuition and prior knowledge are verified or re-evaluated on the basis of the likelihood of hypotheses given observation. If the observation is insufficient, we search fore more. Scientists search for objective facts and correspond subjective opinions to them, not for surprises (mismatch between prior and likelihood) or their absence.

Now, are you ready to join me in the follow chapters, and search among what I have searched and found?

Abstract

Human Visual Detection and Search in Natural Backgrounds

Anqi Zhang, PhD
The University of Texas at Austin, 2024

SUPERVISORS: Wilson S. Geisler, Ernst-Ludwig Florin

Visual detection and search in complex natural backgrounds are fundamental and ubiquitous tasks for humans. Nevertheless, the information processing of the human visual system in those tasks have not been well understood. In this dissertation, we applied Bayesian decision theory to analyze and model visual detection and search. For detection, we found human observers fully exploit spatial modulation of background contrast, but only partially the background spectrum. We observed a target was more detectable when its phase was more similar to that of the background, and explained this effect with the interaction between phase similarity and intrinsic position uncertainty. Bridging from detection to covert search, human search performance was surprisingly better than the ideal searcher with the measured human detectability map, despite substantial loss of detectability in the fovea. Correlated internal noise is a plausible explanation. More importantly, we found extremely simple heuristic decision rules for covert search are almost optimal. Our systematic analysis revealed factors that significantly affect the performance lag of heuristic searchers. Furthermore, heuristic compositions that result in the same accuracy can be distinguished by patterns in location-dependent statistics. Overall, our discoveries deepen

understanding of human visual detection and search, and will spawn applications in numerous industries, such as medical image perception, human-computer interface, and artificial vision.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation to study human visual detection and search

Have you ever tried to look at the sky and clouds to spot an impending rain or storm? Have you ever tried to monitor the change of a traffic light to cross the street or resume driving? Have you ever tried to find a family member or a friend in a crowd of people? Have you ever tried to locate a grocery item in a supermarket? Have you ever tried to unearth a key, a phone, or a document from a messy desk?

Visual detection and search in complex natural backgrounds are fundamental and ubiquitous daily tasks for humans, and many other animals. Though we perform these tasks in a seemingly effortless fashion, the associated neural mechanisms, computational and behavioral strategies remains a rich mine of scientific discoveries and life-changing applications. Understanding how humans detect and search

- provides insights into how our brain processes and applies visual information;
- inspires efficient and robust detection and search algorithms in computer vision;
- guides the development of cognitive models for perception, attention, and decision-making;
- informs visual design of human-computer interfaces, such as those in medical imaging, advertising, ergonomics, and security monitoring;
- advises training programs that aim to improve visual detection and search performance, including in natural images, medical images, satellite images, or thermal images.

1.2 Roadmap of my dissertation

I aspire to understand the complex human visual search behavior in the real world through experiments, simulations, and theories along this progressing order: (1) detection: how human observers detect the presence of a target at a singular, fixed location; (2) covert search: how human observers localize a target among multiple locations without eye and head movements; (3) overt search: how human observers choose the direction, magnitude, and frequency of eye and head movements to localize a target; (4) embodied search: how human observers navigate in and make change to the environment of visual content to localize a target; (5) social search: how human observers communicate within a group to localize a target together.

My dissertation presents our discoveries in human visual detection and covert search. The current overarching research questions I seek to answer are: (1) What factors affect human performance in visual detection and search? How? (2) What computation processes do humans use in visual detection and search? How good are they? To address the first question, I measured performance with psychophysics under varying conditions of target and background. For the second question, I employed a hybrid approach based on Bayesian Decision Theory (BDT) and Signal Detection Theory (SDT), to model, explain, and predict human visual detection and search in natural backgrounds. This normative, first-principle approach allows intuitive understanding of human behavior (with concepts such as prior, likelihood, criterion), and the extraction of some of the underlying computation principles human observers are using.

For a majority of the visual tasks I will elaborate in the following chapters, the statistically optimal algorithms are attainable. I will show that this family of algorithms, coined the ideal observer [2], serves as a pivotal performance benchmark for

human observers. Nevertheless, human observers often do not replicate the calculation of the ideal observer, due to limited computation capacity, biological constraints, and/or operational sufficiency of simple heuristics. When the BDT is used to model human visual behavior, both the ideal and heuristic observers need to be analyzed and tested.

The current chapter sketches the current understanding of human vision detection and search most relevant to my research. I will first define the tasks of visual detection and search. Next, I will briefly point out how the human visual system (HVS) and natural images specify the scope of computation in those two tasks. The following discussion will focus on the details of design and analysis of psychophysical experiments to measure and quantify visual behavior. Last but not least, I will lay out the theoretical framework of the BDT for visual detection and search.

In Chapter 2, I will address how well human observers make use of background information in space and spatial frequency to detect targets. To start with, I will describe the $1/f$ noise and its application to approximate natural images. Then I will derive the ideal observer for detection in any filtered Gaussian background with spatially modulated contrast. In short, this optimal algorithm applies template matching, a popular technique in computer vision, after whitening/flattening the spatial frequency and the local contrast. Lastly, given the experimental results, I will introduce biological components, such as the contrast sensitivity function (CSF) and intrinsic position uncertainty (IPU), to model and explain human detection performance.

We also investigated the effect of amplitude-spectrum similarity and phase similarity on human detection performance. In Chapter 3, I will show that the interaction between phase similarity and IPU is able to explain the surprising detection

pattern of human observers.

Chapter 4 bridges visual detection and visual search. I will unearth the factors affecting performance in covert search. Specifically, we used an ancillary detection experiment to measure the detectability map of individuals, and incorporated the map into Bayesian search. Human observers outlandishly outperformed the prediction of Bayes-optimal decision rule, despite having some loss of detectability in the fovea. The spatial correlation of internal noise serves as a plausible cause. Also, we discovered that a wide range of simple heuristics for covert search can achieve near-optimal overall performance.

I will present, in Chapter 5, a systematic analysis of Bayesian heuristics in covert search. I will first define the sensory and heuristic spaces, and then show the degree of performance changes due to varying heuristic search rules. In the last chapter, I will summarize major findings in my dissertation and explore their methodological limitations, theoretical implications, and practical applications.

1.3 Tasks of visual detection and search

Detection typically describes the aim and activity of determining the presence (existence within the range of concern) of a physical object or particular information. The word "detect" comes from "detegere" in Latin, which means to uncover, reveal, and expose. Visual detection is detecting based on the sensing and perceiving of light within the field of vision.

Search typically describes the aim and activity of reducing uncertainty about the location of a physical object or the distribution of particular information. The word "search" has a longer history of origin. The words "kirk" in Proto-Indo-

European, “kirkos” in Greek, and “circus” in Latin have the meanings to bend, surround, go around. Then “serchen” in Old French means to go through and examine carefully and in detail. Visual search is searching based on the sensing and perceiving of light within the field of vision.

Visual detection can be regarded as a specific instance of visual search, a perspective supported by many theoretical frameworks, including the Bayesian decision-making model that is the focus of my dissertation. Moreover, from cognitive, neural, and behavioral perspectives, visual detection represents a simplified yet essential component of the broader visual search process.

The physical object or the particular information constitutes the target set, a specified definition of the “signal” in SDT. For example, breast cells and any sign of incidental findings form the target set of search in mammography. A path from the start node to the goal node is the target of the A^* search [3]. The most relevant and rigorous literature forms the target set of the search step in a systematic topic review. If a search reduces the uncertainty of the location of a physical object or the distribution of particular information to a sufficiently limited scale, then we “find” or “localize” the target.

The objects or information in the environment (that do not contain the target set) constitutes the background set, a specified definition of the “noise” in SDT. For example, the X-ray images of healthy patients are background medical images. Literature that is not relevant or rigorous is background information.

Both the target set and the background set can consist of from a single, deterministic element up to an infinite numbers of elements that follow an empirical and extremely complex distribution. In a visual task, the most common cases of a target set include a singular, deterministic projection of an object (e.g., a specific

American flag with a specific luminance, contrast, rotation, scale, shape, texture, material, color, etc.), a single object with some of those image dimensions relaxed, a semantically defined category (e.g., dogs), and multiple categories (e.g., tumor mass, microcalcification, cyst, and abscess). The most common cases of a background set include a single background, a statistically stationary family (e.g., $1/f$ noise), a semantically defined category (e.g., grass), and multiple categories (e.g., images of various natural objects and scenes).

The stimulus set combines target and background, representing all possible observable inputs in detection, search and other behavioral tasks. A stimulus in visual detection and search can have multiple targets, or no target. A present target, when being combined with background, can be additive, occluding, multiplicative, and more.

Figure 1.1 gives an example of a target, a background, a stimulus, and a trial that are typical in our studies. The target is a simple wavelet with fixed spatial frequency, orientation, phase, and contrast. The background is a cropped sample from a grayscale natural image data set. The stimulus is typically generated on a rectangular monitor. The target is additive (matrix addition of pixel values) to the background, as shown in Figure 1.1c. A trial begins with the human observer focusing at the center of the display with the help of a visual cue, and a blank background shows up to prevent the visual cue from leaking into the stimulus (temporal integration). For a detection task, the target is present at a known location, or absent. For a search task, the target is present at one of the potential locations, or absent. Lastly, sufficient time is given for the observer to make their decision and have their response recorded.

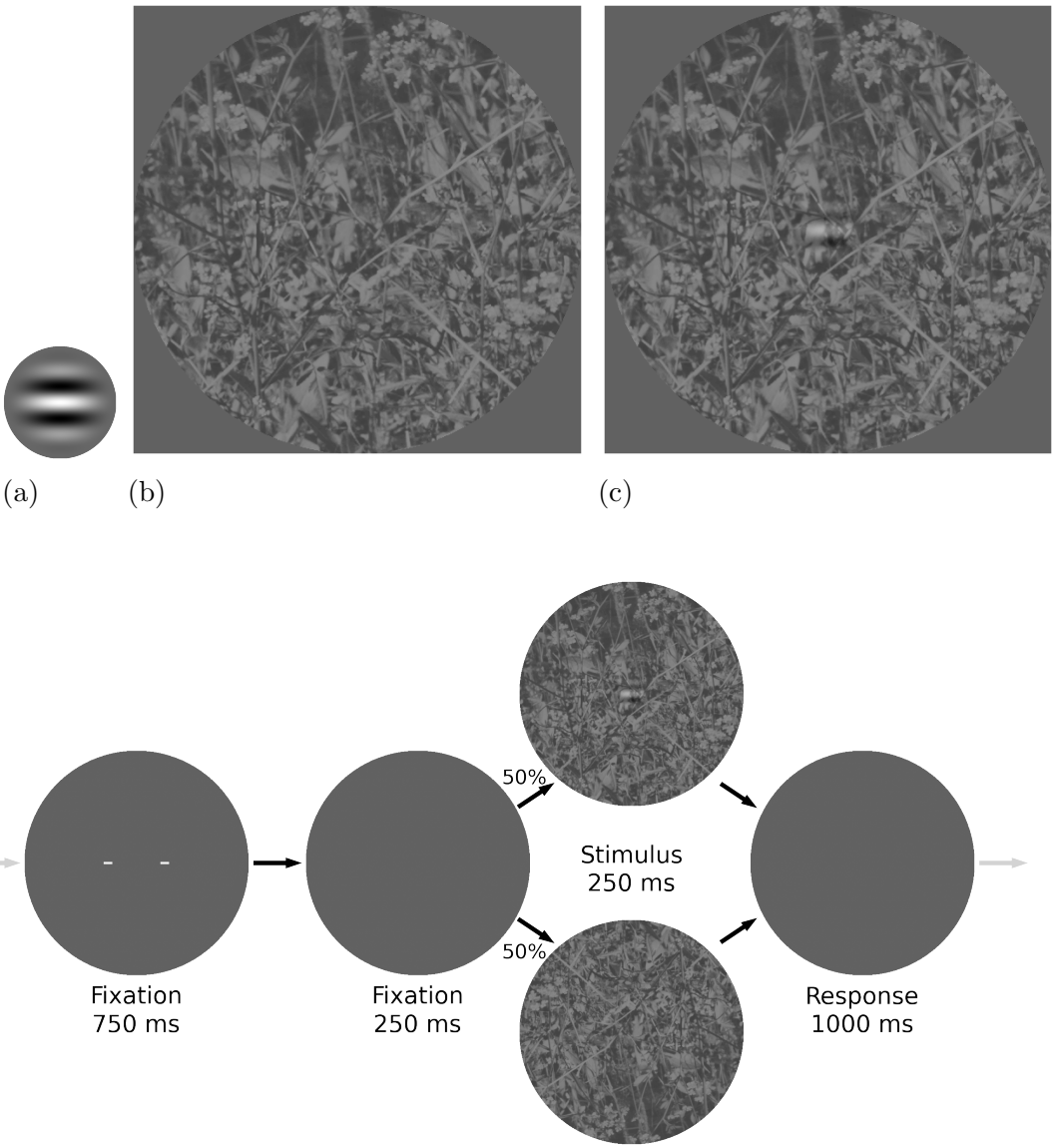


Figure 1.1: An example of detection and search. (a) Target. (b) Stimulus when the target is absent. (c) Stimulus when the target is present. (d) Timeline of stimulus presentation in a trial.

1.4 Computation in the human visual system

The HVS is a wonderfully complex, powerful, and specialized neural network for perceiving and interpreting the world visually. While the human brain uses ap-

proximately 20% of the body's energy supply on average [4], the HVS ranks among the brain's highest energy consumers. Also, more than 50% of the sensory receptors in a human body are located in the eyes [5].

The HVS consists of the eye, the optical nerve, the optic chiasm, the optic tract, the lateral geniculate body, the primary visual cortex (V1), and some other cortical areas (e.g., V2, V3, V4, the middle temporal visual area, the inferior temporal cortex, the lateral intraparietal cortex, the frontal eye field). The cortical areas that are involved in vision are highly connected [6].

Here I introduce the primary visual processing pathways in simple terms. Light from the world is registered in the eye, specifically in the retina, by the rod and cone cells. These photoreceptor cells convert the electromagnetic signal into neural signals. The optic nerve then transmit the signals to the lateral geniculate nucleus (LGN) for preliminary processing. V1 receives visual information from LGN and organizes it with a spatially mapped representation. That concludes the early visual pathway. The V2 area further prepares the information for downstream pathways. In simplified terms, the dorsal stream, involving V1, V2, V3, the middle temporal visual area MT), locates where the things are; the ventral stream, involving V1, V2, V4, the inferior temporal cortex (IT), recognizes what the things are.

My dissertation applies some of the experimentally verified computations in the early visual pathway to model human visual detection and search. I always seek to justify the model components in our computational human vision research through discoveries in experimental neuroscience and psychology, and also am open to be inspired by algorithms in engineering (e.g., image processing), artificial intelligence (AI) and computer science (e.g., convolutional neural network). The order of justification and inspiration cannot be reversed; otherwise, the research shifts into the field of

computer vision.

To start with, a receptive field [7] is a spatial region of the retina where visual stimuli affect the firing behavior of a specific neuron, such as a retinal ganglion cell (RGC), a LGN cell, or a V1 neuron. It corresponds computationally to a template, which is a predefined pattern to search for an object within an image. The simplest template has the shape of the target. Multiple regions of the image are compared/matched with the template to maximize a certain similarity measure, typically the cross-correlation (dot product).

The receptive fields of RGCs and LGC cells commonly have an on-center, off-surround structure [8], where "on" means being excited and "off" means being inhibited, in local feature dimensions such as contrast, luminance, orientation, velocity [9], spatial frequency [10]. The relationship between stimulus features and neuronal responses in the early visual pathways is modeled by linear and non-linear filters, in the language of signal processing. A tuning curve describes the neuronal response as a function of parameters in those features, while a sensitivity function describes the response of the whole HVS as a function of parameters in those features, such as the CSF.

The same object with spatially varying local features (e.g., luminance and contrast) is perceived stably, most likely through normalization and gain control in the neural circuitry [11, 12]. Local normalization of properties, such as luminance, contrast and motion, makes typical receptive field or template responses much more Gaussian-distributed [13–17].

Another fact of the HVS related to computation is the internal noise, that neural responses have inherent variability independent of visual stimuli [18]. This

noise is commonly modeled with Gaussian or Poisson distributions as a part of the overall decision variable for a specific visual task.

Due to internal noise, human observers accept target-like signals centered not exactly on the known, fixed location [19–25], which is termed intrinsic position uncertainty. I modeled this uncertainty by convolving the template with the background region with a size slightly larger than the target size and centered on the fixed target location, and then choosing the maximum or summed response.

Visual memory (working memory and long-term memory) justifies the usage of prior and Bayesian update in our decision-making framework. Given the limited resolution and capacity of memory, computations in visual processing are often heuristic rather than ideal/exact. For instance, the template may not have the exact same shape as the target.

1.5 Statistics of natural images

The behaviors and theories of human visual detection and search are tightly related to the physical and statistical properties of the most typical backgrounds—natural images. The efficient coding [26] hypothesis states that neurons encode the information efficiently, that is, with the amount of spikes minimized. Furthermore, sparse coding (compact representation) of natural images has been found in the visual system [27]. In visual detection and search, human observers represent those regularities efficiently to distinguish the target from the background, constrained by task priority and biological limitation. Therefore, measuring and understanding statistics of natural scenes [28] helps construct both optimal models [29] and biological models for the HVS.

The image perceived by the HVS resides in a high-dimensional space. Consider a digital image consisting of pixels (shorted for picture elements). Each pixel can be conceptualized as a dimension of the space the image is in. At a distance suitable to view a laptop, 1080p is the "good enough" video resolution, which refers to 2 million pixels per color channel. Though human eyes do not have uniform pixels across the visual field, the numbers of rod and cone cells per eye were estimated to be around 60 million and 3 million, respectively [30]. In other words, an image for the HVS has at least several million dimensions. For just the second-order statistics, the covariance matrix of such an image can easily possess beyond 1 trillion elements.

Interestingly, natural images occupy only a low-dimensional sub-manifold of the overall image space. The statistical regularities (weak symmetry and invariance). Imagine an image is randomly sampled with each pixel value uniformly distributed. In most cases, it will resemble much more of a white noise (static noise on an aged TV screen), than a natural scenery.

However, I have not addressed a fairly important question, that is what makes an image "natural". This question is harder to answer in our digital age. I do not intend to include arbitrary visual outputs on a monitor screen, which adventures to the whole image space. I also do not need to exclude an image of a monitor with power off sitting in an office, or a natural image rendered on a monitor screen, as the statistics of such an image is still consistent with that of the images that are indisputably natural.

Therefore, I define a natural image as a 2-dimensional projection of natural light irradiated and reflected from physical objects, or a reconstruction of such natural light with artificial light. Natural light, in contrast to artificial light, originates from sources without human intervention, such as the Sun and other stars, the moon,

fireflies, wildfire, etc. Physical objects, in our definition, can be either natural, such as water, sky, rocks, plants, animals, or artificial, such as buildings, bridges, and furniture.

By no means is this the only definition of natural images. The different definitions varies from context to context, such as visual neuroscience, material science, thermal imaging, medical imaging, remote sensing, astronomical imaging, game design, and photography.

The natural image space can be further grouped into overlapping spaces by observers in the biosphere. In my dissertation, I implicitly focus on natural images that are visible to human observers (with normal adult vision). That means X-ray, Gamma rays, infrared, and radio waves produced by natural light source is not natural for human observers to see.

Now we are well-situated to discuss the properties of natural images. It is self-evident that image features such as luminance and contrast in a natural image vary spatially. To prevent information loss of task-relevant features, the HVS can simply measure each feature at each local location, but that is quite costly. Any spatial correlation within and across features is desired, because it relaxes the sampling requirements of a visual system. Here are some empirical results.

Both local luminance and local contrast have the average auto-correlation rapidly dropped to 0.25 when 2-4 visual degrees away from the point of interest [31]. However, this result needs to be interpreted considering the range of the focal length. For a close-up shot of a uniform surface, instead of landscape images, this decorrelation scale is much larger.

As two point-wise scalars, local luminance and contrast are statistically in-

dependent or weakly dependent of each other [31, 32]. However, the local matrices of luminance and contrast seem to share a strong correlation [33]. Those seemingly inconsistent results could be explained by the possibility that luminance at a local location is partially correlated with contrasts surrounding that location. More analyses are desired on this correlation.

The best-known property of natural images is that their spectra follow a power law. Extensive literature (see Table 2.1) have shown the amplitude spectral density (ASD) of a natural image is inversely proportional to the spatial frequency, that is

$$A(f) \propto \frac{1}{f} \tag{1.1}$$

This power law implies that lower spatial frequency contents (large-scale structures and slow modulations) have higher power compared to those of higher spatial frequencies (fine details and edges) in natural images. De Valois et al. [10] show V1 neurons are selectively tuned to different spatial frequencies. In Chapter 2, I will show "whitening" in spatial frequency, that is flattening the spectrum with a weighting filter, maximizes signal-to-noise ratio and detection accuracy. I will also answer if and how much the HVS performs this whitening.

A system that displays the power law is a strong candidate for scale invariance. A random field $f(x)$ is scale-invariant of order k if $f(\lambda x) = \lambda^k f(x)$. If $f(x) = ax^{-b}$ as a power function, then $f(\lambda x) = a\lambda^{-b}x^{-b} = \lambda^{-b}f(x)$. Indeed, Ruderman [34] analyzed natural images and found the distributions of normalized local contrast, gradient, and $1/f$ spectral slope unchanging over widely varying image scales. Thomson [35] demonstrates the power law and scale invariance for a forth-order statistic. Field [36] show, for relative contrast energy to be scaling variant in a two-dimensional image, a

power law for ASD with the exact exponent of 1 is required. For this reason, Pentland [37] generates natural scenes with fractals.

In other words, the distribution of edges, textures and patterns in natural images do not change significantly when being zoomed in and out. That means conclusions based on natural image statistics at one scale, including those in my dissertation, are promisingly generalizable to most scales. A quick sufficiency test for the scale invariance of a natural image is to evaluate the $1/f$ spectral slope.

The $1/f$ power law of natural image spectra does not address whether natural images are Gaussian distributed. In fact, they have heavy-tailed, non-Gaussian, Laplacian-like distributions [34, 38]. Even by averaging pixel values over a small local region, the mean is still non-Gaussian, probably due to the abundance of very large and very small gradients [34]. The central limit theorem does not apply because those pixel values still correlate. Nevertheless, typical receptive field or template responses normalized by local luminance and contrast become nearly Gaussian [13–17]. The non-Gaussian property of natural images have been simulated by models such as dead leaves [39] and wavelet trees [40].

To summarize, I explored the correlation structure of statistics in natural images, highlighted their $1/f$ power law in the amplitude spectrum, the ensuing scale invariance, the non-Gaussian behavior, and the Gaussianization by local normalization.

For any linearly filtered Gaussian (LFG) noise, the ideal observer transforms the stimuli as close to a high-dimensional standard normal distribution as possible (called whitening), and then applies template matching (see Section 2.2 for more details). Given that the properties of natural images are somewhat similar to those

of the LFG noise, we found the ideal observer that whitens also detects the target exceedingly well in natural images.

1.6 Human behavior in visual detection and search

How does the HVS behave in visual detection and search?

Foremost, the central vision of a human has significantly higher resolution than peripheral vision. The number of ganglion cells declines by a factor of two at just 1.5 to 2 degrees of angle from the line of sight [41–43]. The cone cells, mainly responsible for daylight (photopic) vision, have considerably higher density in the fovea than in the near and far periphery [30, 44]. Different from many computer vision models, foveated vision is a default component of human vision models, especially for search in a large visual field.

Much early literature on visual search measures the time to find a target object (reaction time) among other objects (distractors)[45–49]. Reaction time typically increases as the number of distractors increases, which is termed the set size effect. The speed of increase ranges from nearly 0 ms/item to dozens of ms/item. Search is more efficient when the target is defined by a single feature rather than multiple features [45], and when the target and the distractors are more similar to each other [48]. Unsurprisingly, asymmetry in search time is found [46] when the shapes of the target and the distractor are swapped.

Our projects in this dissertation adopt a different paradigm, focusing on accuracy and statistics in the response confusion matrix rather than reaction time, and noise backgrounds rather than distractors. In overt search, we aim to account for saccade planning and selection along with spatial-temporal integration of visual in-

formation, and assume the serial processing of items is a special case when features require high cognitive efforts or their detectability decline rapidly with eccentricity. Therefore, we isolate the effect of eye movements with covert search, with presentation durations matched to fixation durations during natural overt search [50, 51]. Nevertheless, the trade-off between search accuracy and search time is theorized [52].

We treated natural images as noise field, and also generated white noise and $1/f$ noise as backgrounds. No distractor objects are placed in the stimulus; what "distracts" is the local noise regions that look similar to the target. We incorporated quantitatively the known effects in visual search, such as peripheral vision, target detectability, receptive field response, and intrinsic position uncertainty.

The external factors that affect human detection and search include background luminance, background contrast [53], similarity between target and background [54], clutter density, semantic grouping, depth and perspective.

How and how much does the foveation of human vision affect the detectability of a target? In a statistically uniform background, a target is most detectable along the temporal (horizontal meridian) direction, second along the inferior direction and in the lower visual field, and last along the superior direction and in the upper visual field [55–59]. This result is typically summarized as a detectability map in my dissertation.

Looking into the pattern of human eye movements during search, it can be roughly characterized as fixation periods separated by rapid, voluntary, gaze-shifting called saccades. Saccades enable the fovea to sample different parts of an image with high resolution. The typical duration of fixations is around 250 ms [50, 51], so we set the display duration of stimuli to be 250 ms in all our detection and covert search experiments, as shown in Figure 1.1d.

Visual detection and search employs visual attention. The term generally describes the filtering, selection and weighting of information by object, location, or feature in the visual field. Bottom-up attention is stimulus-driven, involuntary, exogenous, and unselected; top-down attention is goal-oriented, voluntary, endogenous, and selected.

Though the research in my dissertation only treats visual attention as a computational component in the detection and search models, the neural implementation of visual attention has been extensively studied [60]. Moran and Desimone [61] show when monkeys are trained to focus on the stimuli at one location and ignore stimuli at another, the response of some neurons in V4 to the unattended stimulus was significantly reduced. Fries et al. [62] measured the synchronization frequency in V4 neurons. When the stimuli are attended, frequencies from 35 to 90 Hz increase, and those less than 17 Hz decrease. Bisley and Goldberg [63] found some neurons in Lateral Intraparietal Area (LIP) are involved in attention, with their ensemble activity matching the spatial and temporal dynamics of attention.

Inattentional blindness is the phenomenon where individuals do not notice salient but task-irrelevant objects. The famous experiment by Simons and Chabris [64] show when the 192 naive observers were asked to count basketball passes in a video, almost half of them failed to notice a gorilla or a woman carrying an umbrella passing by. Gorilla images in chest CT scans [65] and in fingerprints [66] were not noticed by even expert radiologists and fingerprint analysts. Overall, these experiments indicate the crucial component of information selection in the HVS, that the top-down, goal-oriented attention can override the bottom-up, stimulus-driven attention.

In Chapter 4, I will describe and model another type of "blindness" with gain

control. Individuals were much less sensitive to targets right at fovea in a large visual field, compared to in a small visual field [59]. They distributed limited attentional resources to focus on peripheral regions that have more potential target locations, so search accuracy was increased. This phenomenon is coined "foveal neglect". Inattentional blindness is not seeing something right in front of you when you are not looking for it; foveal neglect is not seeing something right in front of you when you are looking for it.

At the end of this section, I briefly introduce two efficient strategies found in human overt search. Inhibition of Return refers to the phenomenon where individuals are less likely to fixate back to the regions they recently examined [67], favoring the exploration of new areas [68]. Another phenomenon is that fixation locations are concentrated in a donut-shape region, a few visual degrees from the center of the display [58, 69]. Interestingly, the upper and lower periphery have the highest count, consistent with the models taking advantage of the detectability structure that a target in a uniform background is most detectable along the temporal direction. Nevertheless, a recent research suggests the fixation locations are distributed more uniformly [70].

1.7 Psychophysics: Measuring visual behavior

Psychophysics is a field of psychology that studies the quantitative relationship between physical stimuli and the behavior (either observed or communicated) they affect. In my dissertation, the scope of psychophysics is limited to tasks with objective standards (e.g., detection accuracy) instead of subjective preferences (e.g., cuteness of dogs), assuming beyond reasonable doubt that all human participants sincerely reported what they perceived to be most likely true.

Psychometric function describes this relationship using quantitative variables. A typical case is shown in Figure 1.2a, where the behavioral performance is fitted as a smooth function of a physical stimulus. When the physical quantity is extremely small or large, behavior response plateaus; fast change of behavioral performance occurs when the physical quantity changes from a moderate level. For instance, a target is totally undetectable with an extremely small amplitude, and totally detectable with an extremely large amplitude. Nevertheless, the shape of a psychometric function can vary diversely, even with the assumption of smoothness, monotonicity and asymptotic saturation (Figure 1.2b-d).

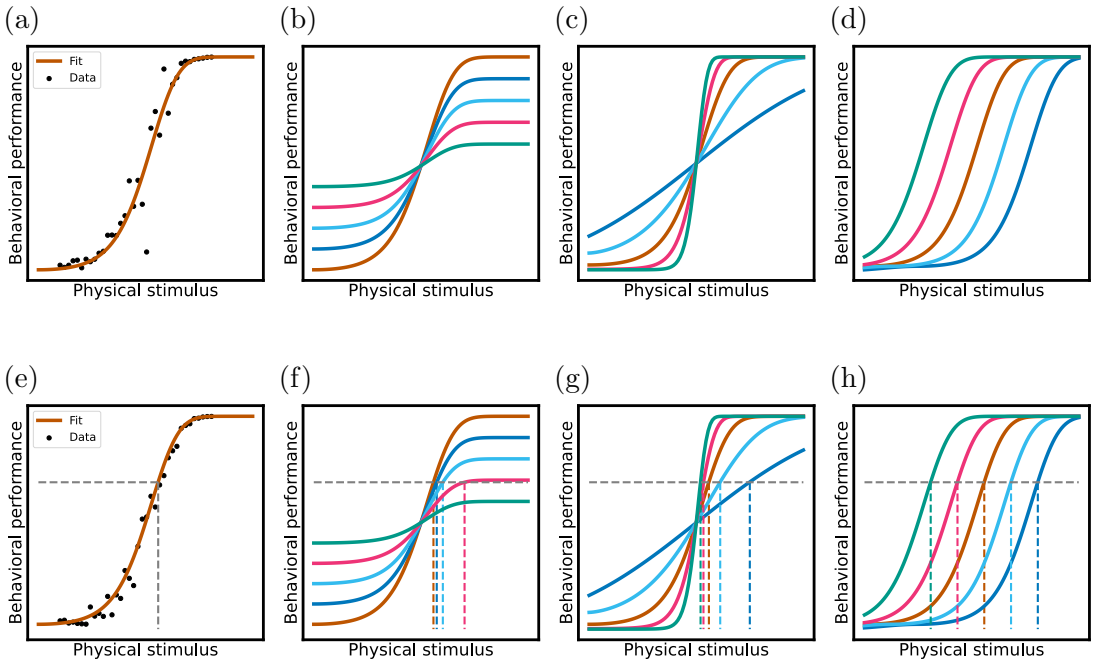


Figure 1.2: Psychometric function. (a) Fitting from mock data. (b) Change in "lapse". (c) Change in "slope". (d) Change in "location". (e-h) Thresholds.

In the context of visual detection and search, quantities for physical stimuli include target amplitude, background luminance, background contrast, similarity

between target and background; quantities for behavioral performance include proportion correct, hit rate, and correct rejection rate.

Threshold is a fundamental concept in psychophysics and a major performance metric in my dissertation. It refers to the value of a physical stimulus when the behavioral performance reaches a certain value. For example, we define the detection threshold as the level of target amplitude when the overall detection accuracy just reaches $\Phi(1/2) \approx 69\%$. Figure 1.2e-h) visualizes the process of obtaining a threshold. Though the same threshold value does not necessitate the same psychometric function, visual behavior is comparable through threshold values in slightly different conditions.

1.8 Theories of human visual detection and search in natural images

What theories are applicable to human visual search in natural backgrounds? Treisman and Gormican [47] proposed the Feature Integration Theory that has a two-stage search process. On the pre-attentive stage, basic visual features such as color and orientation are processed in parallel. On the focused attention stage, a target as a conjunction of features is searched in space serially. Bundesen [71] (also [72]) proposed a theory of visual attention, with an attention mechanism selecting elements/items (filtering) and another selecting categories (pigeonholing).

It is worth pointing out Wolfe et al. [49] proposed visual search is guided by both bottom-up and top-down mechanisms. In their most recent theory [73], prior history, reward, and scene syntax and semantics are also involved in guiding attention.

Green and Swets [2] summarized the SDT and hence provided a quantitative framework to describe the process of making decisions based on observation. It has been successfully applied to the visual detection and search tasks [74–78]. The

Bayesian detection and search theory I am introducing belongs to this family of theory. Here I start with the core concepts in SDT.

The simplest signal detection model in noise is the equal-variance Gaussian model (Figure 1.3a). A decision variable D, as a statistic of a system, follows a Gaussian distribution for a certain state of the system $N(\mu_a, \sigma^2)$ (a indicates the target is absent in the stimulus), and another Gaussian distribution with the same variance for another state $N(\mu_b, \sigma^2)$ (b indicates the target is present in the stimulus). The state of the system S is either a or b. For simplicity, $\mu_a \leq \mu_b$. A single measurement/observation in a trial, noted as d, is a random sample from the distributions based on the actual state of the system.

Now I define a simple decision rule for the response (estimation of the state) as

$$\hat{S} = \begin{cases} a, & d < \gamma \\ b, & d > \gamma \end{cases} \tag{1.2}$$

where γ is the decision criterion.

Over infinite trials with equal numbers of target-present and target-absent stimuli, the accuracy

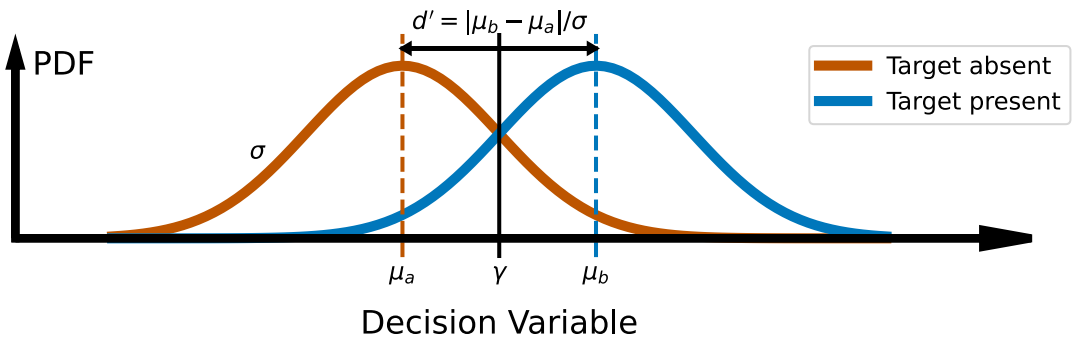
$$A(\gamma) = \frac{1}{2}\left[\Phi(\frac{\gamma - \mu_a}{\sigma}) + 1 - \Phi(\frac{\mu_b - \gamma}{\sigma})\right] \tag{1.3}$$

To find the optimal criterion where the accuracy is maximized, we obtain

$$\frac{dA}{d\gamma} = \frac{1}{2\sigma}\left[\phi(\frac{\gamma - \mu_a}{\sigma}) - \phi(\frac{\mu_b - \gamma}{\sigma})\right] \tag{1.4}$$

That means, the optimal criterion is the value where the probability densities of the two distributions are equal, that is $\gamma_o = (\mu_a + \mu_b)/2$.

(a)



(b)

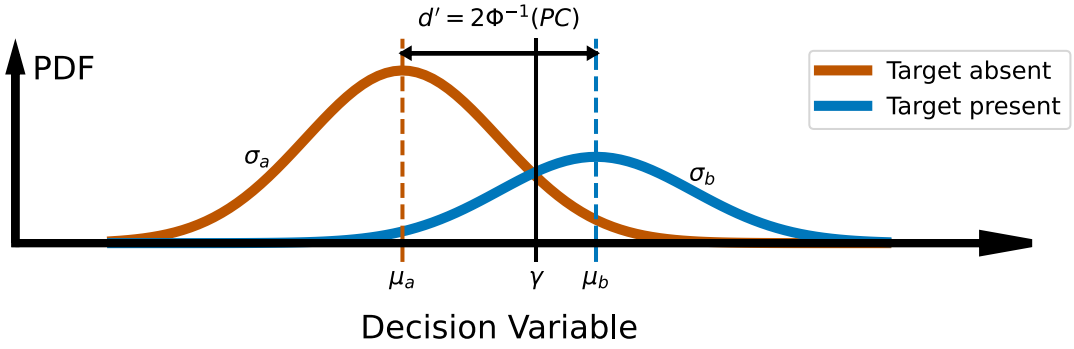


Figure 1.3: Signal Detection Theory. (a) Equal-variance Gaussian model. (b)Unequal-variance Gaussian model.

The detectability index, also called sensitivity index or discriminability index, is defined as

$$d' = \frac{\mu_b - \mu_a}{\sigma} \tag{1.5}$$

Detectability characterizes how good the decision variable is to help distinguish between the two states of the system. A large d' indicates the signal and noise distributions are well-separated. A small d' indicates the distributions overlap. More difference in the means and less variance increase d'.

The two-state confusion matrix categorizes trials into the following types:

- hit (true positive) rate: $P(\hat{S}=b|S=b)=\Phi(\frac{d'}{2}-\gamma_o)$
- correct rejection (true negative) rate: $P(\hat{S}=a|S=a)=\Phi(\frac{d'}{2}+\gamma_o)$
- false alarm (false positive) rate: $P(\hat{S}=b|S=a)=1-P(\hat{S}=a|S=a)=\Phi(-\frac{d'}{2}-\gamma_0)$
- miss (false negative) rate: $P(\hat{S}=a|S=b)=1-P(\hat{S}=b|S=b)=\Phi(-\frac{d'}{2}+\gamma_0)$

We cannot measure the decision variable by human behavior in psychophysics alone, but the empirical hit rate p_h and correct rejection rate p_{cr} can be obtained. Then the corresponding detectability and optimal criterion can be solved reversely as

$$d'=\Phi^{-1}(p_h)+\Phi^{-1}(p_{cr}) \qquad \gamma_o=\frac{1}{2}\left[\Phi^{-1}(p_{cr})-\Phi^{-1}(p_h)\right] \tag{1.6}$$

The calculation of detectability and criterion in our simulations and analyses also consider the case where the decision variable still follows Gaussian distributions but with different variance (Figure 1.3b). We have

$$\frac{dA}{d\gamma}=\frac{1}{2\sigma_a}\phi(\frac{\gamma-\mu_a}{\sigma_a})-\frac{1}{2\sigma_b}\phi(\frac{\mu_b-\gamma}{\sigma_b}) \tag{1.7}$$

The optimal decision rule in this case uses two criteria, but the error from using a single criterion is usually small. Assume we only apply a single decision criterion, its optimal value is (still with the convention that $\mu_a<\mu_b$)

$$\gamma_0=\frac{\mu_a\sigma_b^2-\mu_b\sigma_a^2+\sigma_a\sigma_b\sqrt{(\mu_b-\mu_a)^2+2(\sigma_b^2-\sigma_a^2)\ln\frac{\sigma_b}{\sigma_a}}}{\sigma_b^2-\sigma_a^2} \tag{1.8}$$

The maximal accuracy (proportion correct) $PC=\frac{1}{2}[\Phi(\frac{\gamma_0-\mu_a}{\sigma_a})+\Phi(\frac{\mu_b-\gamma_0}{\sigma_b})]$, and the (generalized) detectability $d'=2\Phi^{-1}(PC)$.

1.9 Bayesian decision-making in visual detection and search

In this section, I present our Bayesian visual detection and search theory. It serves as a normative benchmark, the upper limit of human performance, an observer skeleton to incorporate and discover the effects of biological factors on detection and search, and also a compact yet intuitive language to understand and explain how humans process visual information to make decisions in those tasks.

Consider the target location space $\mathbb{X} = \{0, 1, \cdots, n\}$, where "0" is the location label for target absence, and "1" is for the first target location, and so on. We define $\mathbb{Y} = \{1, \cdots, n\}$ as the location space without target absence. Therefore, x is the actual location of the target in a trial and y is the present-location variable, and the trial response is the estimated location $\hat{x}$.

Suppose we have a single deterministic target with an amplitude of a and the corresponding template T ($||T|| = 1$). The target is additive to the background; in other words, the stimulus at location

$$\forall x \in \mathbb{X}, \forall y \in \mathbb{Y}, S_y = \begin{cases} B_y & y \neq x \\ aT + B_y & y = x \end{cases} \tag{1.9}$$

Let $\mathbb{V}$ be the space for all pixel coordinates v in a target-size matrix. In the case where background at each location is independent white noise with uniform variance,

$$\forall y \in \mathbb{Y}, \forall v \in \mathbb{V}, B_y(v) \sim N(\overline{B_y}, \sigma_y^2) \tag{1.10}$$

where $\overline{B_y}$ is the background luminance at location y, and σ_y^2 is the background variance at location y.

Assume an observer is able to normalize the local luminance and obtain $D_y = S_y - \overline{B_y}$, then

$$\forall x \in \mathbb{X}, \forall y \in \mathbb{Y}, \forall v \in \mathbb{V}, D_y(v) \sim \begin{cases} N(0, \sigma_y^2) & y \neq x \\ N(aT(v), \sigma_y^2) & y = x \end{cases} \tag{1.11}$$

The likelihood at each target location is

$$\forall y \in \mathbb{Y}, p(D_y|x \neq y) = \prod_v \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{D_y^2(v)}{2\sigma_y^2}\right] \tag{1.12}$$

$$p(D_y|x = y) = \prod_v \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{[D_y(v) - aT(v)]^2}{2\sigma_y^2}\right\} \tag{1.13}$$

Note that the likelihood when target is absent ($x = 0$) is included in Equation (1.12).

The log-likelihood ratio (LLR) of the target at location y versus being absent is

$$ll_y = \ln \frac{p(D_y|x = y)}{p(D_y|x = 0)} = \frac{1}{2\sigma_y^2}\left(2aD_y \cdot T - a^2 T \cdot T\right) = \frac{a}{\sigma_y^2}\left(D_y \cdot T - \frac{a}{2}\right) \tag{1.14}$$

By definition, the target absent location has an LLR of 0, that is $ll_0 = 0$.

What distribution does the LLR follow? For mutually independent variables $X_i \sim N(\mu_i, \sigma_i^2)$,

$$\forall a_i, b_i, \sum_{i=1}^{n}(a_i X_i + b_i) \sim N\left(\sum_{i=1}^{n}(a_i\mu_i + b_i), \sum_{i=1}^{n} a_i^2\sigma_i^2\right) \tag{1.15}$$

By combining Equations 1.11, 1.14 and 1.15, we have

$$
\begin{aligned}
ll_y &\sim \begin{cases} \frac{a}{\sigma_y^2}\left[N(0, T\cdot T\sigma_y^2) - \frac{a}{2}\right] & y \neq x \\ \frac{a}{\sigma_y^2}\left[N(aT\cdot T, T\cdot T\sigma_y^2) - \frac{a}{2}\right] & y = x \end{cases} \\
&\sim \begin{cases} N\left(-\frac{1}{2}(\frac{a}{\sigma_y})^2, (\frac{a}{\sigma_y})^2\right) & y \neq x \\ N\left(\frac{1}{2}(\frac{a}{\sigma_y})^2, (\frac{a}{\sigma_y})^2\right) & y = x \end{cases}
\end{aligned} \tag{1.16}
$$

Now I introduce the prior probability of each location (including the target absent location) as p_x. According to Bayes' theorem, the log-posterior ratio at location x is

$$
lp_x = lr_x + ll_x \tag{1.17}
$$

where $lr_x = \ln \frac{p_x}{p_0}$ is the log-prior ratio, and $lr_0 = lp_0 = 0$.

If $p_0 = 0$, the log-prior ratio becomes a singularity. Because we only need this term for maximum a posteriori (MAP) response, it can simply be replaced by the log prior $\ln p_x$, and allows $\ln p_0 = -\infty$, so no ideal observer will ever respond target-absent in a search task if the target is always present.

The maximum likelihood response in a search task is

$$
\hat{x} = \underset{x\in\mathbb{X}}{arg\,max} \prod_{y\in\mathbb{Y}} p(D_y|x=x) = \underset{x\in\mathbb{X}}{arg\,max} \left[\frac{\prod_{y\in\mathbb{Y}} p(D_y|x=x)}{\prod_{y\in\mathbb{Y}} p(D_y|x=0)}\right] \tag{1.18}
$$

Equation 1.11 shows $\forall x \neq y, p(D_y|x=x) = p(D_y|x=0)$, therefore

$$
\begin{aligned}
\hat{x} &= \underset{x\in\mathbb{X}}{arg\,max} \left[\frac{p(D_y|x=y)}{p(D_y|x=0)} \text{ if } x \neq 0, 1 \text{ if } x = 0\right] \\
&= \underset{x\in\mathbb{X}}{arg\,max}\, l_x = \underset{x\in\mathbb{X}}{arg\,max}\, ll_x
\end{aligned} \tag{1.19}
$$

where $l_x = \frac{p(D_y|x=x)}{p(D_y|x=0)}$ is the likelihood ratio at location x, with $l_0 = 1$.

Similarly, the MAP response in a search task is

$$\hat{x} = \underset{x \in \mathbb{X}}{arg\,max}\, lp_x \quad \text{or} \quad \hat{x} = \underset{x \in \mathbb{X}}{arg\,max}\, [\ln p_x + ll_x] \tag{1.20}$$

For mutually exclusive events, the posterior probability that one of the mutually exclusive events happens is the sum of posterior probabilities that each mutually exclusive event happens. However, that is not necessarily the case for likelihood. The likelihood of target presence is undefined, so there is no maximum likelihood response in a detection task without further assumptions.

The present-absent posterior ratio

$$LP = \ln \frac{\sum\limits_{y \in \mathbb{Y}} \left[p_y \prod\limits_{y' \in \mathbb{Y}} p(D_{y'}|x = y) \right]}{p_0 \prod\limits_{y' \in \mathbb{Y}} p(D_{y'}|x = 0)} = \ln \sum_{y \in \mathbb{Y}} r_y l_y \tag{1.21}$$

where $r_y = \frac{p_y}{p_0}$ is the prior ratio at location y.

Therefore, the MAP response in a detection task

$$\hat{S} = \begin{cases} a & LP < 0 \\ b & LP > 0 \end{cases} \tag{1.22}$$

where a is target-absent, b is target-present, consistent with Equation 1.2.

Clearly, if $p_0 = 1$, $r_y = 0$, $LP = -\infty$, the response is always target-absent; if $p_0 = 0$, $r_y = \infty$, $LP = \infty$, the response is always target-present.

So far, I have derived the maximum likelihood and MAP responses in detection and search tasks for a single, deterministic target in white noise. When there is only one target-present location, the decision rules for detection and search tasks converge.

However, we as humans do not live in a world of white noise, but of natural images. How practical can these ideal observers be to understand human visual detection and search? That depends on how far the framework can be extended and stretched given our assumptions. Indeed, natural images are well-behaved enough for these models to achieve high performance and explain human behavior [17, 25, 79, 80].

Notice that the LLR is distributed normally (Equation 1.16). Receptive field response to natural images is nearly Gaussian after proper normalization [15–17]. We factor out from the LLR a standard normal random variable, $Z \sim N(0,1)$, and interestingly $d'_y = \frac{a}{\sigma_y}$ is the only other variable left (cf. Equation 1.5). Therefore,

$$ll_y = \begin{cases} d'_y Z - \frac{{d'_y}^2}{2} & y \neq x \\ d'_y Z + \frac{{d'_y}^2}{2} & y = x \end{cases} \tag{1.23}$$

By defining the normalized receptive field response

$$R'_y \sim \begin{cases} N(0,1) & y \neq x \\ N(d'_y, 1) & y = x \end{cases} \tag{1.24}$$

The LLR turns into

$$ll_y = d'_y(R'_y - d'_y/2) \tag{1.25}$$

In other words, as long as a decision variable follows Gaussian distributions with equal variance (probably through normalization) at each location, a prior map and a d' map are all we need to apply the optimal detect and search rules. For example, the ideal searcher in Chapter 4 uses the rule $\hat{x} = \underset{x \in \mathbb{X}}{arg\,max} \left[\ln p_x + d'_x(R'_x - d'_x/2)\right]$. The d' map can be measured directly through single-location detection tasks.

I derived the optimal decision rules for more complex detection and search tasks. Given no experiments in my dissertation were designed under those conditions,

I recorded my results in the appendix. See Appendix A for a detection or search task with an objective function as a linear combination of the confusion matrix statistics. See Appendix B for a detection or search task where multiple targets can be present (but not at the same location). See Appendix C for a detection or search task where receptive field responses correlate temporally.

Nevertheless, human observers may use heuristic rules in the decision process to simplify visual processing. Results in Chapter 4 show many extremely simple heuristics are sufficient to achieve near-optimal search performance. Systematic analysis in Chapter 5 is a further investigation on how variation and combination of heuristics affect search performance.

To summarize, I first introduced my motivation and the values of studying human visual detection and search in natural images. Then I paved the road for the rest of my dissertation by explaining what are visual detection and search tasks, what neural computation mechanisms have been discovered in the HVS, what statistical regularities natural images conform to, what I already know about human detection and search behaviors, what theories have been proposed to understand those behaviors, and how our Bayesian detection and search theory stands out among those theories.

Chapter 2: Detection: Double Whitening

Abstract

We measured human detection performance for various target shapes presented in Gaussian $1/f$ noise backgrounds with and without uniform contrast over space. We found that the pattern of human thresholds is not consistent with the ideal observer that whitens in both space and spatial frequency, but is consistent with a sub-optimal observer that whitens fully in space and partially in spatial frequency, with a small level of intrinsic position uncertainty.

2.1 Introduction

In Section 1.5, I explained the spectral and spatial statistics of natural images. Specifically, natural images have varying local luminance and contrast, and a power-law spectrum. In this chapter, I will answer the question of how well the human visual system (HVS) makes use of those statistical regularities to detect targets in natural images. Much content in this chapter is included in this peer-reviewed article [80].

Our detection task was specified in the following ways from the definition of visual detection in Section 1.3. First, each experimental condition had only a single, deterministic target per trial. In other words, the observer did not need to memorize multiple target shapes and detect them simultaneously. Second, there was only a single possible location that the target could be present at. Third, the observer was asked to focus at the center of the stimulus display and make no saccade. The

presentation duration was short enough to allow only one central fixation before response.

We used contrast-modulated $1/f$ noise to approximate natural images as background in detection. As specified in Equation 1.1, $1/f$ noise has an amplitude spectral density inversely proportional to spatial frequency. Table 2.1 shows this relationship is valid for natural images. Most literature in the table focuses on the value of β, where the one-dimensional, cross-sectional radial power spectral density $P(f) \propto f^{-\beta}$. I calculated the values and ranges of β when they were indirectly reported. A β of 2 indicates $1/f$ noise (see Appendix D for more clarification).

Ref.	Year	#Img	Image content	$\beta(\pm\sigma_\beta)$	Main equipments	Image format
[81]	1987	19	Sparsely wooded, rolling grassland, always centered on a vehicle, only during clear or overcast cloud conditions	1.76±0.04 (row-wise), 2.44±0.04 (column-wise)	Nikon FE camera fitted with a Nikon ED f/5.6, 600-mm lens, Kodak Ektachrome 200 Professional 35-mm color reversal film	128 x 128 sq px, 8-bit
[36]	1987	6	Trees, rocks, bushes, water around England and Greece	≈ 2	Keystone 3572 camera (35 mm), XP1 Kodak monochrome film	256 x 256 sq px, 8-bit
[82]	1992	117	Published photographs of trees, gardens, animals, mountain scenery, people, urban scenery	2.13±0.36	A Philips 56470 CCD-camera module, Data Translation DT2861 frame grabber	128 x 128 sq px

[83]	1992	135	Animals, plants, the English countryside, buildings, vehicles, and laboratory equipment	2.40±0.26 (images), 2.36±0.30 (segments)	Kodak Tmax-100 35 mm film	256 x 256 sq px (images), 128 x 128 sq px (segments)
[34]	1994	45	Woods, trees, scrub, rocks, a stream, central New Jersey, during springtime.	1.81±0.01	Sony Mavica MVC-5500 still video CCD camera with a 9.5–123.5 mm zoom lens	256 x 256 sq px, 8-bit
[84]	1995	320	Videotapes made by authors, subjectively natural to authors	2.3	Sony Handycam CCD-FX710 camera	64 x 64 sq px, 8-bit
[85]	1996	276	Woods, fields, parks, residential areas, with varying distances, elevations, times of the day, types of weather, seasons	1.88 ± 0.43, 1.88±0.51 (per orientation)	77RS CCD camera by PCO Computer Optics with a 16 mm Sony TV lens	541 x 512 sq px, 8-bit
[86]	1997	20	Around New York, Canada and Alaska, no man-made structures, including sky and water	2.20±0.28	A 35 mm camera with Ilford XP1 film	512 x 512 sq px, 12-bit
[87]	1997	82	Trees, plants, roads, Brodatz collection of natural textures	2.38	A SLR camera with a 50-mm lens and a 135-mm lens	512 x 512 sq px, 8-bit

[88]	1997	16/group	3 groups: meadows, forest, close-ups. In the Sierra mountains near Reno, Nevada and at Tahoe Meadows	2.418 ± 0.160 (meadows), 2.150 ± 0.110 (forest), 2.228 ± 0.422 (close-ups)	Kodak DCS420IR monochrome digital camera with added IR filter	256 x 256 sq px, 8-bit
[89]	1998	29	Plants, flowers, trunks, branches, grass, leaves, trees, bushes, rocks, and sky	2.22 ± 0.26	A hyperspectral camera with a Fuji CF25B 25-mm f/1.4 lens	256 x 256 sq px, 8-bit
[90]	2001	133	Wood, field, close-ups, urban areas	1.88 ± 0.42	From the Hateren database (Kodak DCS420 digital camera with a 28-mm lens)	512 x 512 sq px, 8-bit
[91]	2003	6,000	River, waterfall, forest, field, mountain, beach, coast	1.98 ± 0.58 (horizontal), 2.02 ± 0.53 (oblique), 2.22 ± 0.55 (vertical)	Corel stock photo library	256 x 256 sq px
[92]	2004	95	MIT campus	2.29 (spherical harmonic)	Data from MIT City Scanning Project	Spherically tiled illumination maps

Table 2.1: Power spectra of natural images.

I will explain the specific computation of the ideal observer in the next section. Conceptually, the ideal observer extracts spatial information from contrast-modulated

$1/f$ noise by reliability weighting (i.e., whitening in space), and frequency information by whitening. We found that the detection pattern of human observers is consistent with the model observer that fully whitens in space but partially whitens in spatial frequency, and with a small level of intrinsic position uncertainty. For the partial whitening in spatial frequency, we reached the same conclusion with multiple previous studies [93–97].

2.2 Ideal observer in linearly filtered Gaussian noise

A question you might ask is why we did not use natural images directly, rather than a proxy family of images. The reason is that we sought to separate the spatial contrast and frequency information from other information in the natural images in order to investigate how well human observers can make good use of them. Statistical optimality of the detection rule is undefined in natural images because they are not strictly statistically uniform (wide-sense stationary). However, we obtained the ideal observer in contrast-modulated $1/f$ noise. The difference in performance pattern between the ideal observer and human observers provides insight on how close the HVS is to optimally extract spatial and frequency information for detection.

In fact, this ideal observer is optimal for any image combining a known luminance profile and a linearly filtered Gaussian (LFG) noise. LFG noise is a generalized case of contrast-modulated $1/f$ noise, that is any Gaussian noise with linear filters applied to the spatial and spatial frequency domains, or

$$N = f_1 \mathbb{F}^{-1}\{f_2 \mathbb{F}\{N_0\}\} \tag{2.1}$$

where $N_0 \overset{i.i.d.}{\sim} N(0,1)$ is the Gaussian white noise, $\mathbb{F}$ is the Fourier transform,

f_1 and f_2 are linear filters in space and spatial frequency, respectively.

The LFG noise space is enormous. Figure 2.1 shows a few examples. Amplitude spectrum can be filtered in any way, such as with power laws, band-passing, and band-stopping. Local contrast can be modulated in regions separated arbitrarily, such as in a chessboard pattern and through dynamic thresholding.

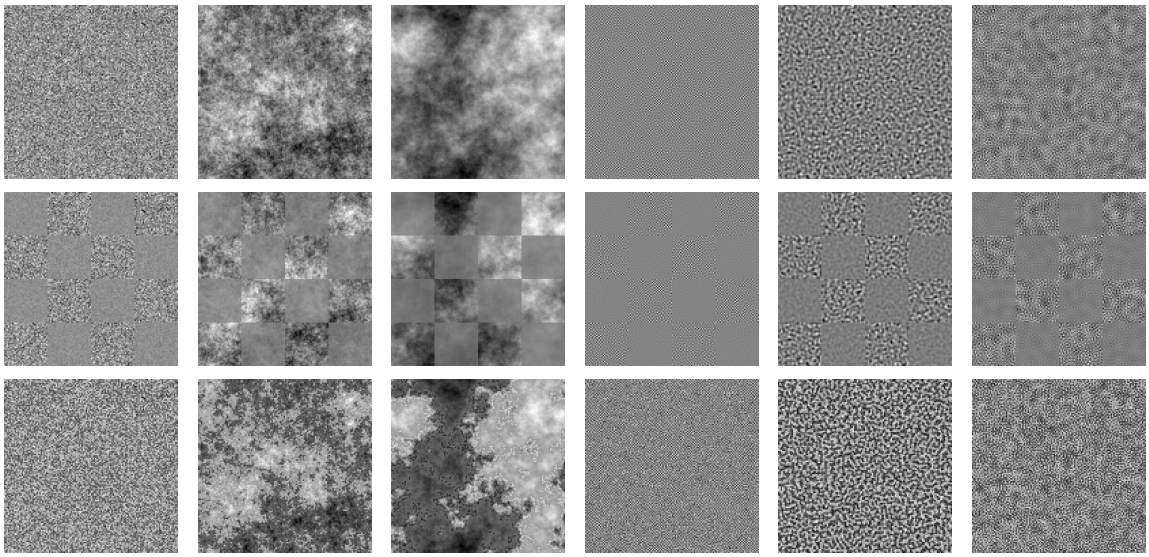


Figure 2.1: Examples of linearly filtered Gaussian image. Rows from top to bottom: no modulation of local contrasts; local contrasts modulated with a chessboard pattern; local contrasts modulated by regions based on thresholding of local luminance. Columns from left to right: white noise, $1/f$ noise, $1/f^{1.5}$ noise, f^5 noise, band-pass noise, band-stop noise.

When both f_1 and f_2 are identity filters, no whitening is needed to obtain the optimal [2, 98] decision variable

$$R = N \cdot T \tag{2.2}$$

where N is the initial LFG noise (uniform white noise for optimality), T is the template of the target. By convention, its Euclidean norm $||T|| = 1$. One obtains

the initial LFG noise by subtracting the known mean luminance image from the background.

This decision variable is a simplified version of the log-likelihood ratio in Equation 1.14. Because the amplitude of the target and the variance of the noise are constant, they can be absorbed into the decision criterion. This simple template matching variable is a sufficient statistic for the maximum a posteriori decision rule.

When f_1 is an identity filter, then whitening in spatial frequency produces the optimal decision variable R_w [99, 100], that is

$$R_w = N_w \cdot T_w \tag{2.3}$$

$$N_w = \mathbb{F}^{-1}\{f_2^{-1}\mathbb{F}\{N\}\} \tag{2.4}$$

$$T_w = \mathbb{F}^{-1}\{f_2^{-1}\mathbb{F}\{T\}\} \tag{2.5}$$

where N_w is the LFG noise whitened by the linear filter f_2^{-1}, and T_w is the template whitened by the same whitening filter.

The intuition to also whiten the template is that if the target is added to the background, whitening of the noise will also act on the target, so the template needs to track this "new" target in the whitened noise.

When f_2 is an identity filter, then whitening in space, sometimes called reliability weighting, produces the optimal decision variable R_r [79], that is

$$R_r = N_r \cdot T_r \tag{2.6}$$

$$N_r = f_1^{-1} N \tag{2.7}$$

$$T_r = f_1^{-1} T \tag{2.8}$$

where N_r is the white noise with contrast varying spatially weighted by the

linear filter f_2^{-1} to regain spatially uniform contrast, and T_w is the template weighted by the same filter.

Similarly, the intuition to also weight the template is that if the target is added to the background, reliability weighting of the noise will also act on the target, so the template needs to track this "new" target in the weighted noise.

As you may have guessed, the ideal observer for detecting a target in an arbitrary LFG noise whitens in both space and spatial frequency, or

$$R_{wr} = N_{wr} \cdot T_{wr} \tag{2.9}$$

$$N_{wr} = \mathbb{F}^{-1}\{f_2^{-1}\mathbb{F}\{f_1^{-1}N\}\} \tag{2.10}$$

$$T_{wr} = \mathbb{F}^{-1}\{f_2^{-1}\mathbb{F}\{f_1^{-1}T\}\} \tag{2.11}$$

Figure 2.2 gives a demonstration of the whitening procedure. If the initial background (upper right) is contrast-modulated (left vs. right) $1/f$ noise, then whitening it in space produces uniform $1/f$ noise, and whitening it in spatial frequency produces contrast-modulated white noise. The ideal observer further performs the complimentary filtering and fully whitens the background. Pay attention to the change in target shape (template) in the lower right corner of each background. Reliability weighting modulates the local contrast of the template, while whitening in spatial frequency sharpens the template (boost in high frequency content).

Transforming a noise as close to a high-dimensional standard normal distribution as possible maximizes the signal-to-noise ratio for template matching.

In the following discussion, I regard the model with Equation 2.2 as the simple template matching (TM) observer, the model with Equation 2.3 as the whitened

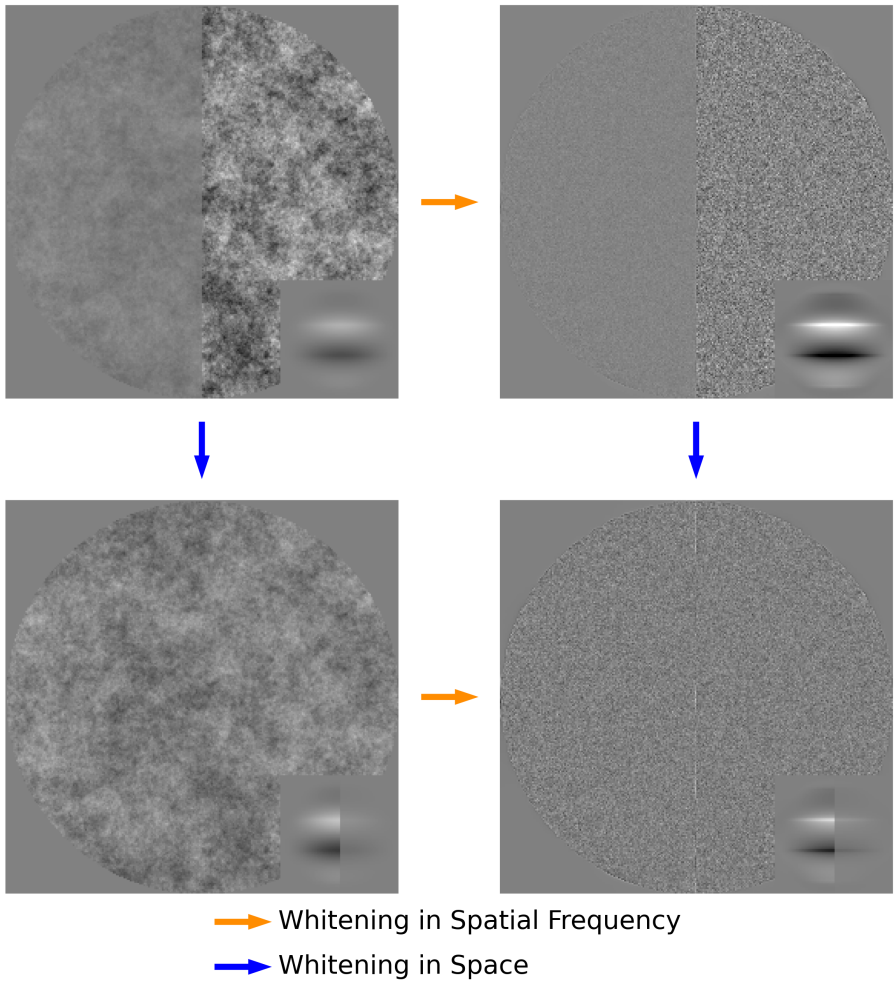


Figure 2.2: Whitening in space and in spatial frequency.

template matching (WTM) observer, the model with Equation 2.6 as the reliability weighting (RTM) observer, and the model with Equation 2.9 as the whitened reliability weighting (RWTM) observer.

2.3 Methodology and experiments

All experimental procedures below were approved by the University of Texas Institutional Review Board (IRB). Informed consent was obtained from all partici-

pants. The study included three male participants, aged 22–26. They all had normal or corrected-to-normal acuity. In a trial, the observer's head was stabilized with a chin and head rest.

The stimuli in the experiments were generated with MATLAB 2020a and the Psychophysics Toolbox [101, 102]. The stimuli were displayed with a resolution of 60 pixels per visual degree on a well calibrated Sony GDM-FW900 cathode-ray-tube (CRT) monitor. The monitor had a display size of 1920 x 1200 pixels, a refresh rate of 85 Hz, and a bit depth of 8. Prior to display on the screen, the stimuli were clipped ($< 0.02\%$ pixels), gamma-compressed, and quantized to gray levels in the range of 0-255.

Experiment 1

Psychometric functions were measured in a detection task for 10 conditions: two types of backgrounds and five target shapes (Figure 2.3). Each condition included 600 trials (10 amplitude levels $\times$ 30 trials $\times$ 2 sessions) per participant. Amplitude thresholds were calculated by fitting with maximum likelihood estimation to generalized cumulative Gaussian functions to hits and false alarms, which are

$$p_h(a|\alpha,\beta,\gamma) = \Phi\left[\frac{1}{2}(\frac{a}{\alpha})^\beta - \gamma\right] \tag{2.12}$$

$$p_{fa}(a|\alpha,\beta,\gamma) = \Phi\left[-\frac{1}{2}(\frac{a}{\alpha})^\beta - \gamma\right] \tag{2.13}$$

where p_h is the hit rate, p_{fa} is the false alarm rate, a is the target amplitude, α, β, γ are the slope, shape, and criterion parameters of the generalized cumulative Gaussian function, $\Phi[\cdot]$ is the standard normal cumulative distribution function.

The detection threshold was defined to be the target amplitude giving $d' = 1.0$, or 69% proportion correct with the optimal (unbiased) criterion. Based on Equations 2.12 and 2.13, the threshold was equal to α. Its confidence interval was obtained by bootstrapping.

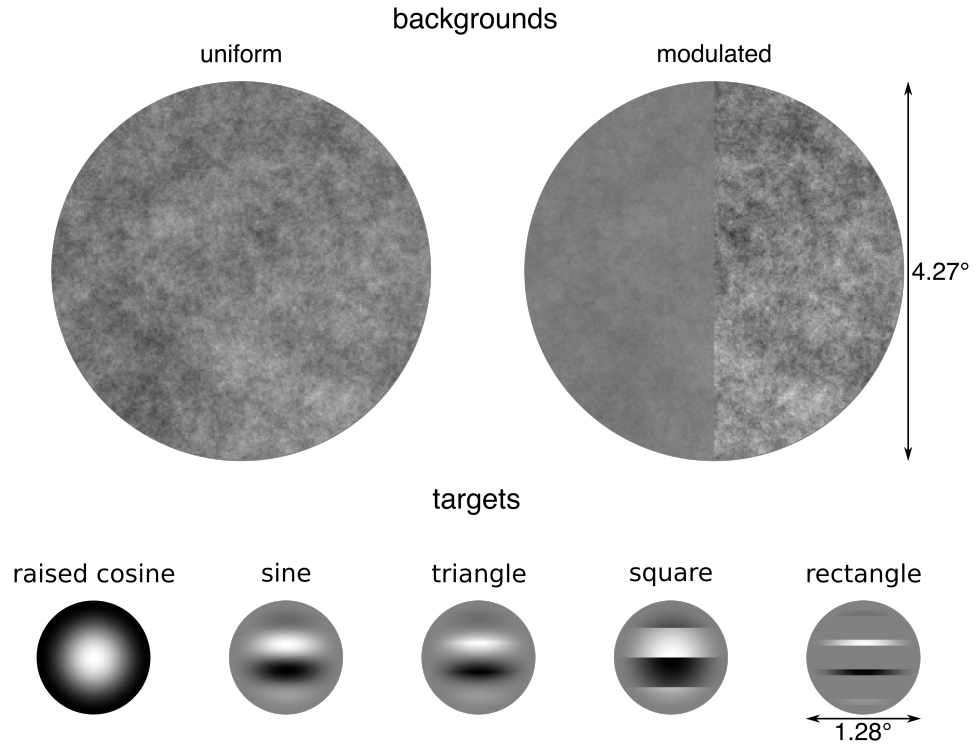


Figure 2.3: Backgrounds and targets in Experiment 1.

The uniform background shared the same total contrast power with the contrast modulated background, and had a root-mean-square (RMS) contrast of 20.4%. The contrast-modulated background had one low-contrast (RMS=7.0%) and one high-contrast (RMS=28.0%) region; those two regions had a contrast ratio of 4.0, and randomly alternated between the left and right halves of the display per trial to reduce effects of contrast adaptation. Both the uniform and contrast-modulated

background had a diameter of 4.27 visual degrees, or 256 pixels. The mean luminance of circular background patches was always 46 cd/m^2, which was equal to the luminance outside the patch on the screen. Each background patch was sampled randomly from a 4096 $\times$ 4096 squared-pixel field of $1/f$ noise. The spatial frequencies below 16 cycles per image were removed from the field before sampling. This ensured the lowest frequency in the background patch is 1 cycle per patch image.

The targets were a raised-cosine blob, and raised-cosine-windowed sine, triangle, square, and rectangle waves with a duty cycle of 10%. All targets had a diameter of 1.28 visual degrees, or 77 pixels, and the same level of total contrast power. Except for the blob target, they all had a mean of zero and a spatial frequency of 1.5 cycles per visual degree (cpd). The amplitude of the target was defined as the square root of the sum of the squared pixel values (the square root of the target energy).

In each trial, a central fixation cue was given for 750 ms and then extinguished for 250 ms (Figure 2.4). Then a stimulus was displayed for 250 ms, that is the typical fixation duration during natural overt search [50, 51]. The target was present for half of the trials, and if present, always at the very center of the display. A human observer was asked to press the left arrow key to respond "target-absent", and the right arrow key to respond "target-present". Auditory feedback was given at the end of each trial on whether the response was correct.

Compared to previous similar studies of detection in LFG noises [103, 104], we varied not only the properties of the target, but also the properties of the background.

Experiment 2

The differences between Experiment 1 and this experiment include: (1) The raised cosine target was not included; (2) All wave gratings underlying the raised-

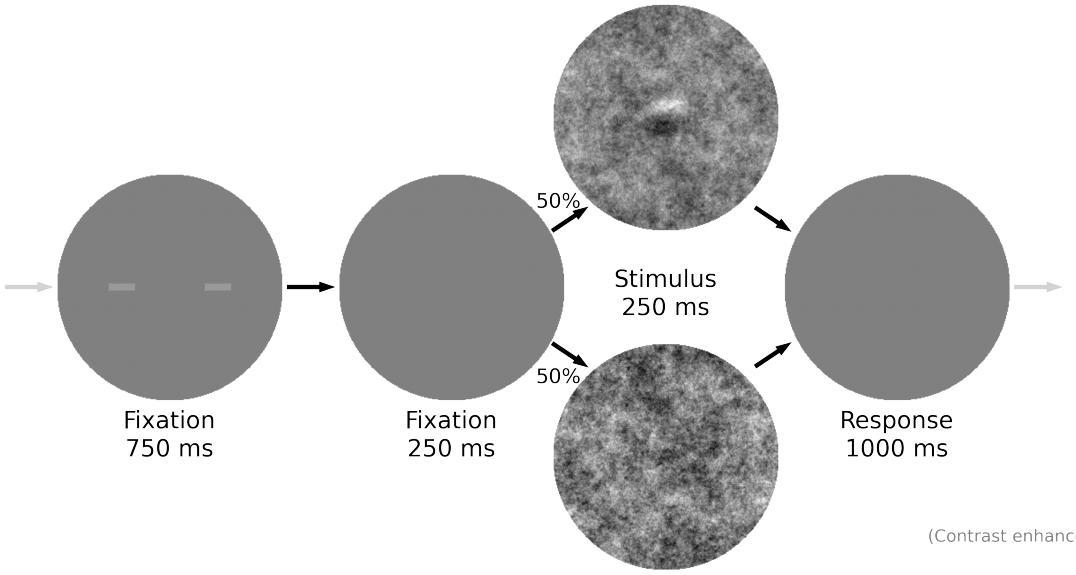


Figure 2.4: Timeline of a trial in Experiment 1 (also in Experiment 2). The example condition includes the triangle wave target and the uniform $1/f$ background.

cosine window had a spatial frequency of 3.0 cycles per visual degree, instead of 1.5; (3) All targets had a diameter of 0.64 visual degrees, or 38 pixels; (4) Two of the three participants in Experiment 1 completed this experiment; (5) This experiment was run before Experiment 1.

2.4 Human detection performance in 1 / f noise

I will first summarize the pattern of detection performance measured from the human observers. Then I will describe the detection performance of varying model observers and compare the average human thresholds with the predictions from those model observers.

Figure 2.5 provides an overview of human detection performance. As expected, when the target amplitude increased, the detection accuracy increased for each participant in each condition. The range of the overall accuracy typically went from 55%

to 95%, so the threshold ($\sim 69\%$) was well captured.

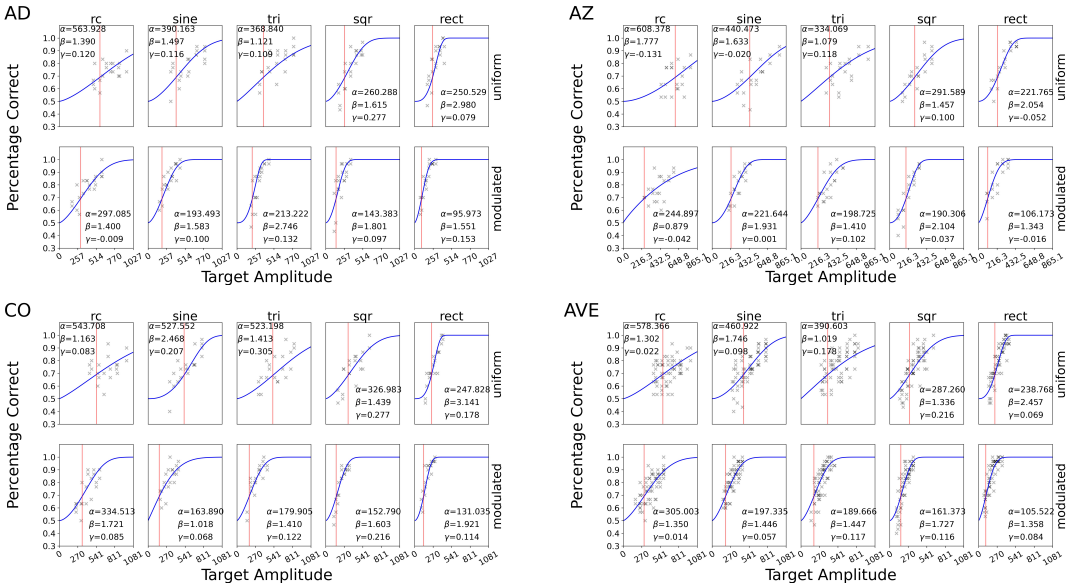


Figure 2.5: Psychometric functions of Experiment 1. Each subplot corresponds to one of the human observers or the average human observer. In each subplot, there were ten combinations of the target and background conditions. Gray cross-check: human data; blue curve: psychometric fit; red line: threshold.

For quantitative comparison, thresholds were plotted in Figure 2.6 after being calculated with

$$t = 20 \log_{10}(\alpha) \tag{2.14}$$

where t is the amplitude threshold in decibel scale. The multiplier is 20 instead of 10 because decibels are traditionally used for power, which is proportional to the square of amplitude. Note that a difference of 6 dB corresponds to a factor of 2 in threshold change.

Averaged across all five targets, the detection threshold in the modulated background was 6.1 dB lower than that in the uniform background. In other words,

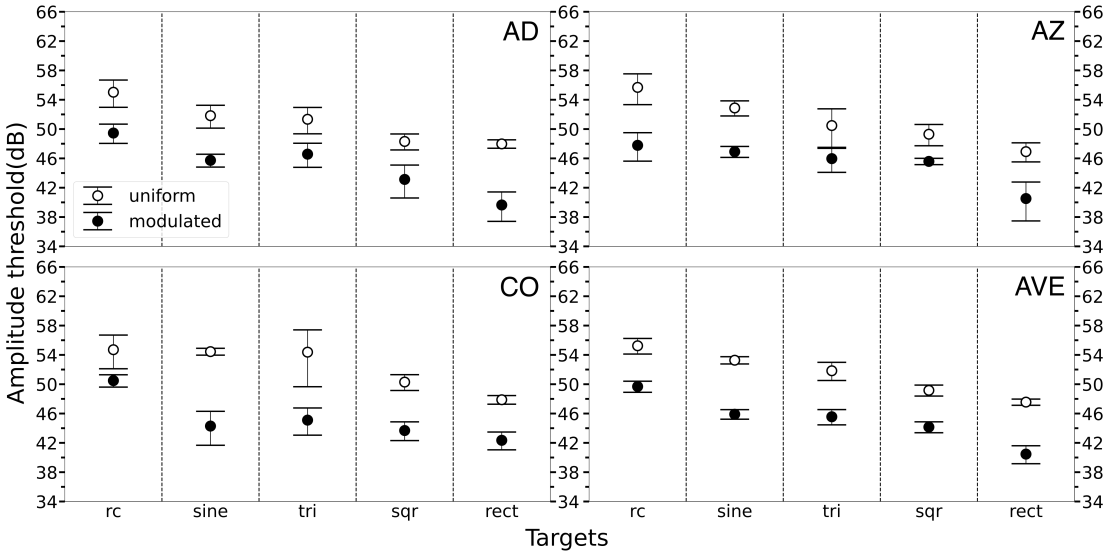


Figure 2.6: Thresholds of human observers and the average human observer in Experiment 1. Solid circle: uniform-contrast background; open circle: contrast-modulated background.

targets are a little more than twice detectable in the modulated background as in the uniform background.

In both background conditions, the detection threshold decreased mostly monotonically as the mean amplitude of spatial frequencies of the target increased (rc, sine, tri, sqr, rect). As shown in Figure 2.7, edges, especially sharper and thinner ones, generate high frequency components. In $1/f$ noise background, those high frequencies have higher signal-to-noise ratios than low frequencies.

Psychometric functions in Experiment 2 (Figure 2.8) show consistent relationship between the target amplitude and the detection accuracy. Detection accuracy gradually increased as target amplitude increased.

Detection thresholds across background conditions in Experiment 2 (Figure 2.9) have a trend also similar to that in Experiment 1. The threshold in the mod-

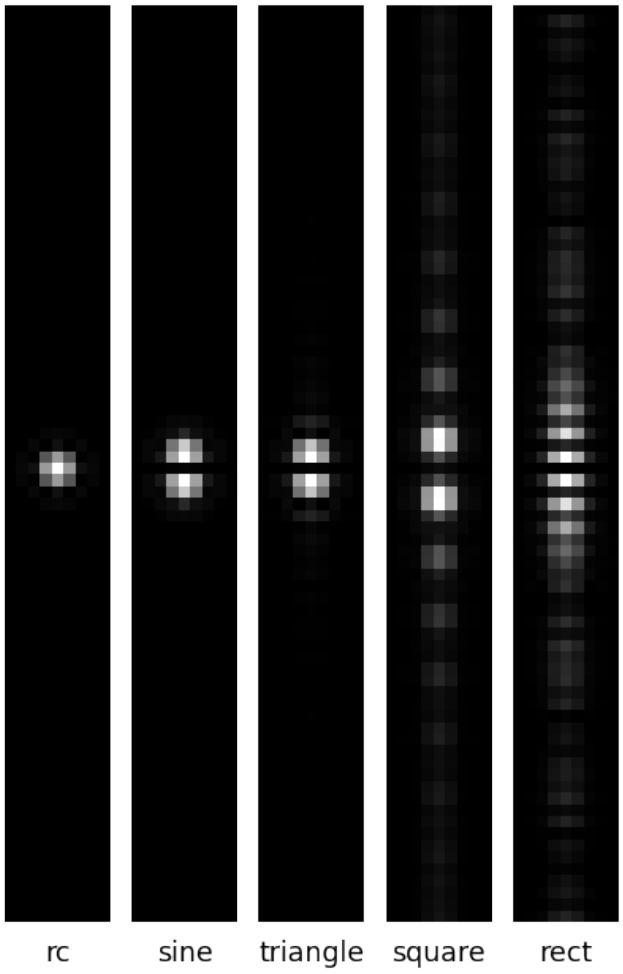


Figure 2.7: Amplitude spectra of the targets in Experiment 1.

ulated background was 3.8 dB lower than that in the uniform background. Targets were more detectable in the modulated background than in the uniform background. Nevertheless, thresholds varied less across targets.

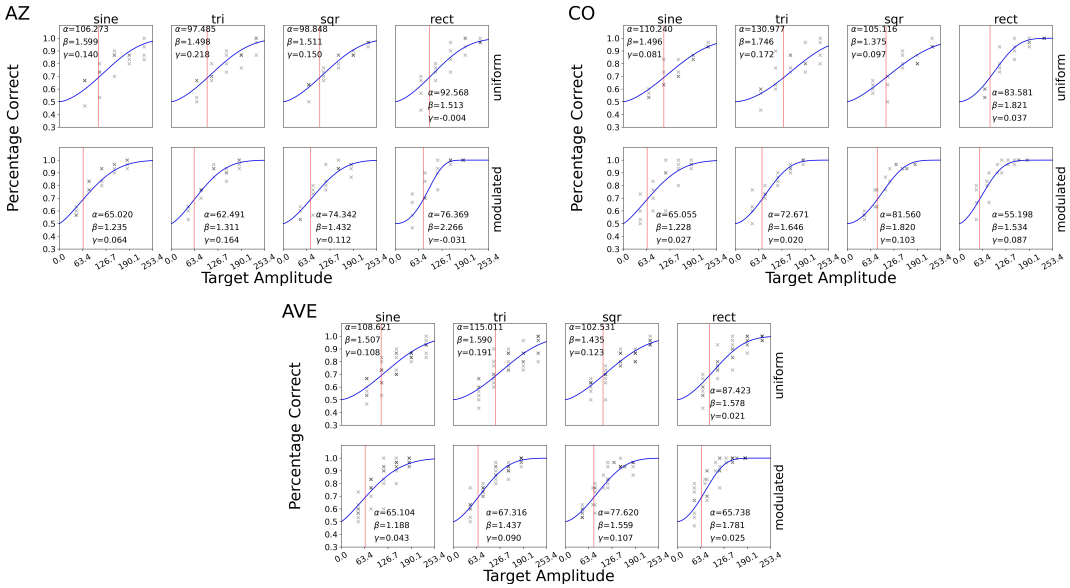


Figure 2.8: Psychometric functions of Experiment 2. Each subplot corresponds to one of the human observers or the average human observer. In each subplot, there were ten combinations of the target and background conditions. Gray cross-check: human data; blue curve: psychometric fit; red line: threshold.

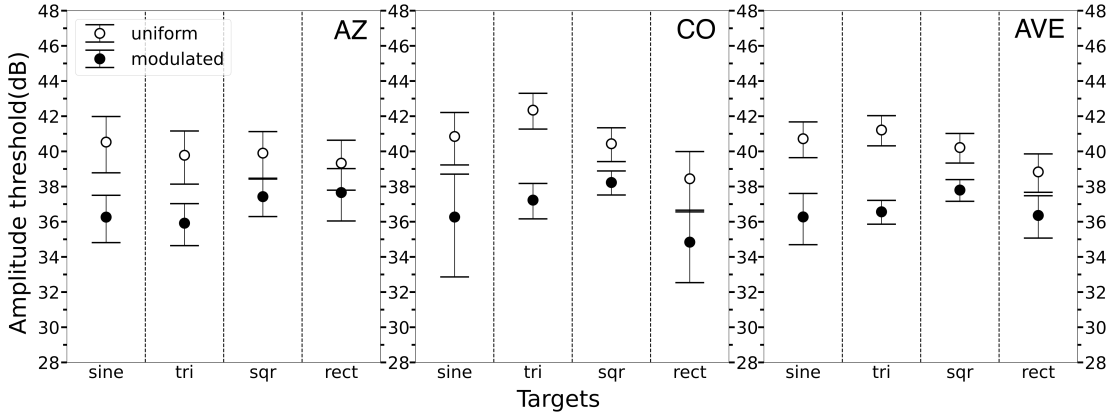


Figure 2.9: Thresholds of human observers and the average human observer in Experiment 2. Solid circle: uniform-contrast background; open circle: contrast-modulated background.

2.5 Template matching models for human detection

Models without biological constraints

Figure 2.10 demonstrates the computation of ideal template matching models in our experiments. In $1/f$ noise, whitening boosts high frequencies and suppresses low frequencies, so the target looks sharper while the background turns into white noise. Reliability weighting increases contrast in the low-contrast (left) region, and decreases contrast in the high-contrast (right) region, so the left side of the target gains higher contrast while the background becomes uniform in contrast.

Which model observer does the HVS behave most similarly to? To answer this question, we compared the absolute values of detection thresholds and their pattern of change across conditions. We fitted model observers to the average human observer with only one free parameter—an overall efficiency scale factor. This factor shifts thresholds in decibel vertically.

The TM observer displays a similar pattern of thresholds as the human observer across targets, as can be seen in Figure 2.11. Thresholds decreased monotonically as the mean amplitude of spatial frequencies of the target increased. The rect target is 11 dB easier to be detected than the rc target. Our explanation is that high-frequency components have high signal-to-noise ratio in $1/f$ noise than low-frequency components. However, TM has the same thresholds in uniform and modulated backgrounds for each target, unlike the human observers. If the background was white noise, the TM observer would have identical thresholds in all ten conditions [79], because the targets and the backgrounds were designed to have the same amount of total contrast power.

The WTM observer has a steeper decline in thresholds across targets than that for the TM observer and human observers. The steeper decline occurs because the

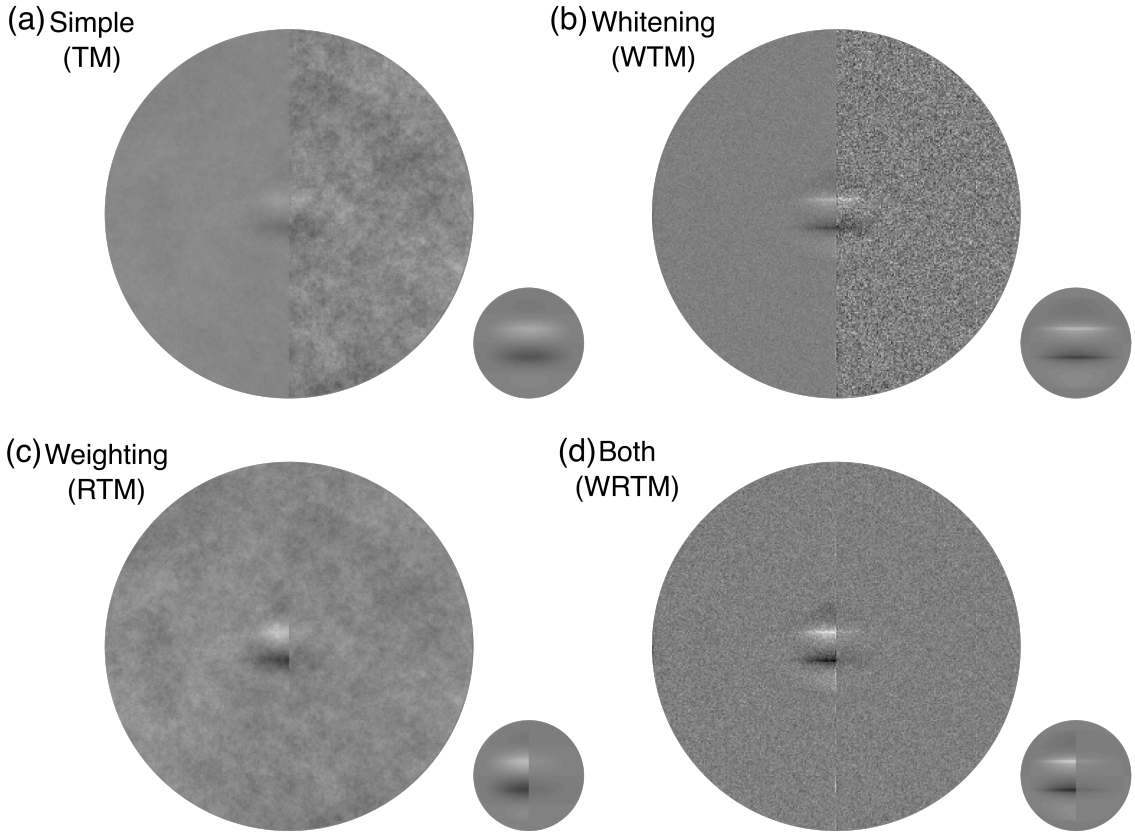


Figure 2.10: Computation of four template matching models without biological constraints. (a) The large patch is the actual stimulus with target present. The small patch is the template of the triangle target, used by the simple template matching (TM, Equation 2.2). (b) Whitened template matching (WTM, Equation 2.3). The large patch is the whitened stimulus with target present. The small patch is the whitened template. (c) Reliability-weighting template matching (RTM, Equation 2.6). The large patch is the weighted stimulus with target present. The small patch is the weighted template. (d) Whitened, reliability-weighting template matching (RWTM, Equation 2.9). The large patch is the whitened and weighted stimulus with target present. The small patch is the whitened and weighted template.

WTM observer perfectly exploits the high spatial frequency content in the targets. Besides the steeper decline, the WTM observer still fails to predict human performance because it has the same thresholds in uniform and modulated backgrounds for

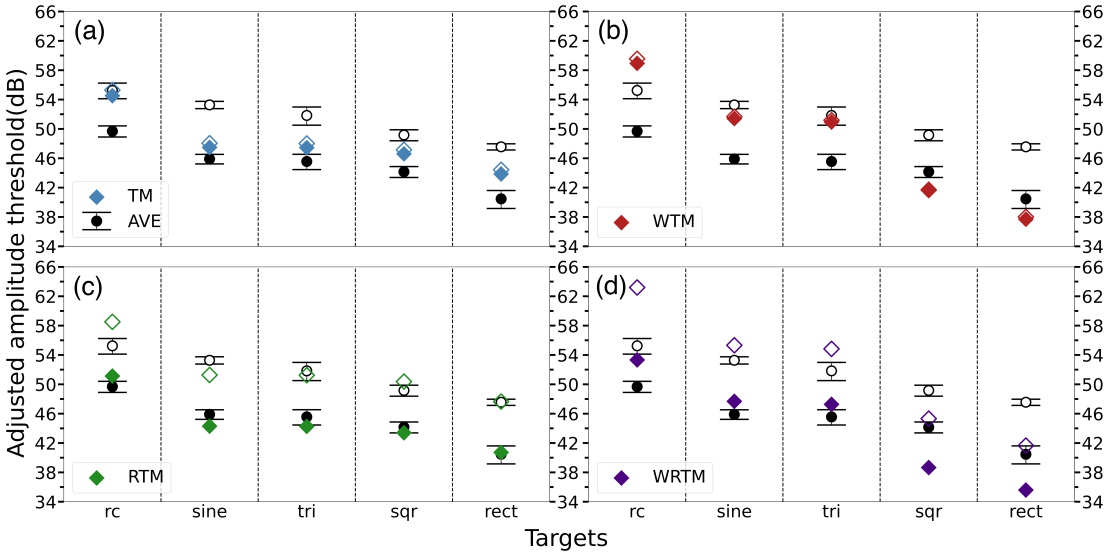


Figure 2.11: Comparison of model and human thresholds in Experiment 1. Open symbols: uniform background; solid symbols: contrast-modulated background. Black circles: average human thresholds from Figure 2.6. Colored diamonds: adjusted thresholds of four model observers that have no biological constraints: (a) TM, (b) WTM, (c) RTM, (d) WRTM. Thresholds of each model were adjusted with a single scalar (0.884, 0.459, 0.610, 0.302 for TM, WTM, RTM, WRTM, respectively) to best match the human thresholds. The RMS errors in decibels are 3.204, 5.717, 1.532, 4.465, for TM, WTM, RTM, WRTM, respectively.

each target.

The RTM observer behaves most consistently to the human observers among the four models, evidenced by the minimum RMS error. The remaining discrepancy is that the RTM observer predicts too small of a difference in thresholds between the raised-cosine and the sine wave targets. The WRTM observer, as the ideal observer for all ten conditions, predicts too steep decline across targets.

Next, to compare the absolute threshold values, we plotted the difference of the model and human thresholds in Figure 2.12. Typically, models perform better than human due to the resolution and memory advantages. If a model has a threshold

higher than that for a human observer (i.e., human efficiency is greater than 1), then the model must miss important information processing that the HVS has. Also, if all the threshold differences in a panel fall on a single horizontal line, then a single efficiency scale factor is sufficient to align the detection pattern of human and model observers.

In the contrast modulated background, thresholds of the TM and WTM observers for certain targets are higher than human thresholds, while thresholds of the RTM and WRTM observers for all targets are lower than human thresholds. Given the limited types of visual information in our stimulus, this result indicates the existence of reliability weighting in the HVS.

Comparing with the ideal WRTM observer, humans are most efficient with the raised-cosine target and least efficient with the rectangular grating target. Specifically, human thresholds for the raised-cosine target are 2–7 dB higher than those of the ideal observer, and 16 dB higher for the rectangular grating target.

The performance comparison between human and model observers in Experiment 2 has similar trends (Figure 2.13). The TM and WTM observers do not take into account of the contrast modulation as human observers exhibited. Full whitening (in spatial frequency) generated a much steeper decline in thresholds across targets. The ideal observer is 8–16 dB better than the human observers, resulting in an efficiency of 24.7% (Figure 2.14).

Overall, our interim conclusion is that for $1/f$ noise background, human observers showed reliability weighting, but not full whitening in spatial frequency.

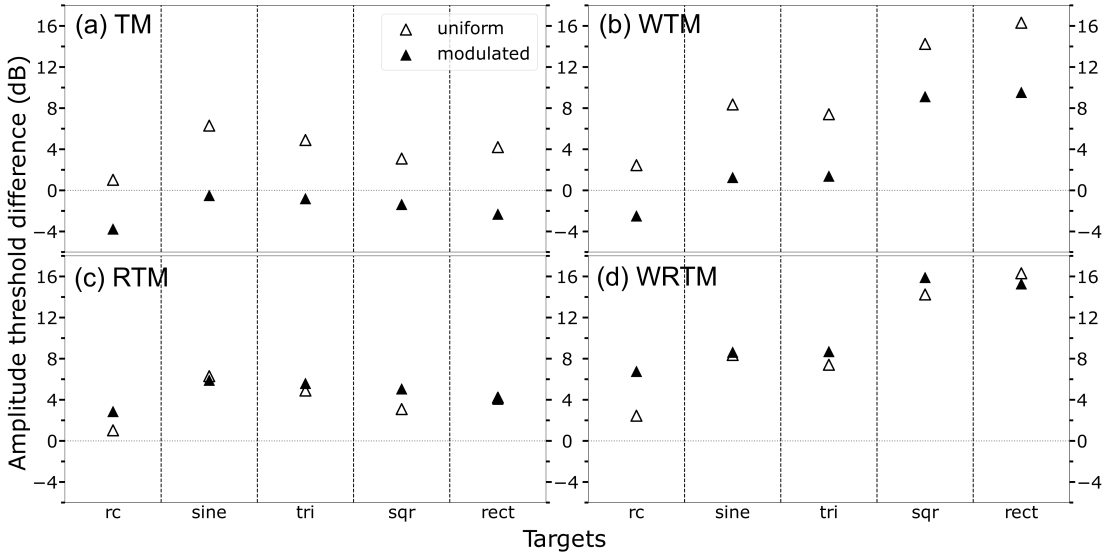


Figure 2.12: Absolute difference of the model and human thresholds in Experiment 1. Each panel plots the average human thresholds subtracted by the thresholds of a model observer. (a) TM, (b) WTM, (c) RTM, (d) WRTM.

Models with biological constraints

We incorporated two properties of the HVS relevant to this detection task: the contrast sensitivity filtering and the intrinsic position uncertainty.

The early visual pathway filters the input light signal with the optical transfer function (OTF) of the eye to obtain the retinal image, which is then transmitted through retinal ganglion cells, bipolar cells, lateral ganglion nucleus to the primary visual cortex.

There is evidence that optical and retinal factors are primarily responsible for detectability of foveal targets in uniform-luminance backgrounds. For example, Bradley et al. [105] find that the optics of the eye and the sampling and filtering of the midget retinal ganglion cells predict the measured foveal detectability for a wide variety of targets reported in the ModelFest Dataset [106]. These measurements (from

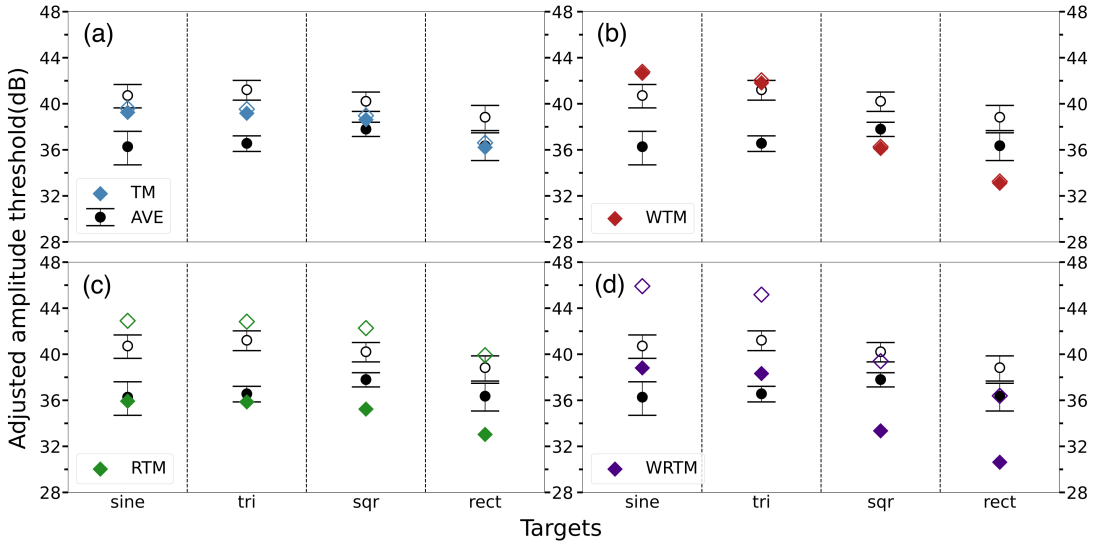


Figure 2.13: Comparison of model and human thresholds in Experiment 2. Open symbols: uniform background; solid symbols: contrast-modulated background. Black circles: average human thresholds from Figure 2.9. Colored diamonds: adjust thresholds of four model observers that have no biological constraints: (a) TM, (b) WTM, (c) RTM, (d) WRTM. Thresholds of each model were adjusted with a single scalar (0.676, 0.354, 0.462, 0.247 for, TM, WTM, RTM, WRTM, respectively) to best match the human thresholds. The RMS errors in decibels are 1.835, 4.082, 1.969, 3.735, for TM, WTM, RTM, WRTM, respectively).

16 human observers in 10 labs) include targets of similar size with similar presentation duration to those in the current study.

To model the foveal amplitude transfer function (ATF) of the early visual system, we fitted the following function to the ModelFest CSF and normalized it to a peak of 1.0:

$$E(f) = k \cdot OTF(f) \cdot \exp(-\alpha f)[1 - \gamma \exp(-\beta f^2)] \tag{2.15}$$

where $OTF(f)$ is the average human optical transfer function for a 4 mm diameter pupil at a wavelength of 555 nm as in Watson [107], α, β, γ are fitting

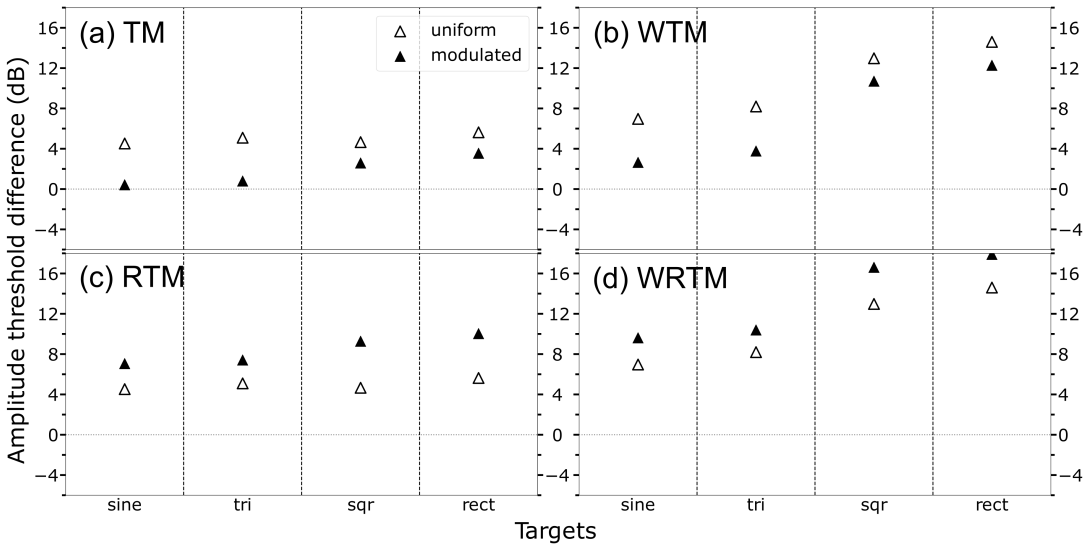


Figure 2.14: Absolute difference of the model and human thresholds in Experiment 2. Each panel plots the average human thresholds subtracted by the thresholds of a model observer. (a) TM, (b) WTM, (c) RTM, (d) WRTM.

parameters, with $\alpha = 0.856$, $\beta = 0.152$, and $\gamma = 0.065$, and k is the normalizing constant to ensure $E_{max} = 1$.

This eye filter (Figure 2.15) partially suppresses low spatial frequencies and hence performs a partial whitening operation. As we replaced the full whitening filter f_2^{-1} with this eye filter in Equations 2.3 and 2.9, we obtained the eye-filtered template matching (ETM) and the eye-filtered, reliability-weighted template matching (ERTM) observers. The ETM observer shares this critical eye filtering component as the non-prewhitening eye filter (NPWE) observer, a popular model in the medical imaging literature [93, 94, 108]. The decision variable for the ERTM observer is

$$R_{er} = N_{er} \cdot T_{er} \tag{2.16}$$

$$N_{er} = \mathbb{F}^{-1}\{E\,\mathbb{F}\{f_1^{-1}N\}\} \tag{2.17}$$

$$T_{er} = \mathbb{F}^{-1}\{E\,\mathbb{F}\{f_1^{-1}T\}\} \tag{2.18}$$

where E is the eye filter operator.

Because the evidence above have suggested reliability weighting as an information processing component of the HVS in this detection task, we will focus on the ERTM observer in later discussion.

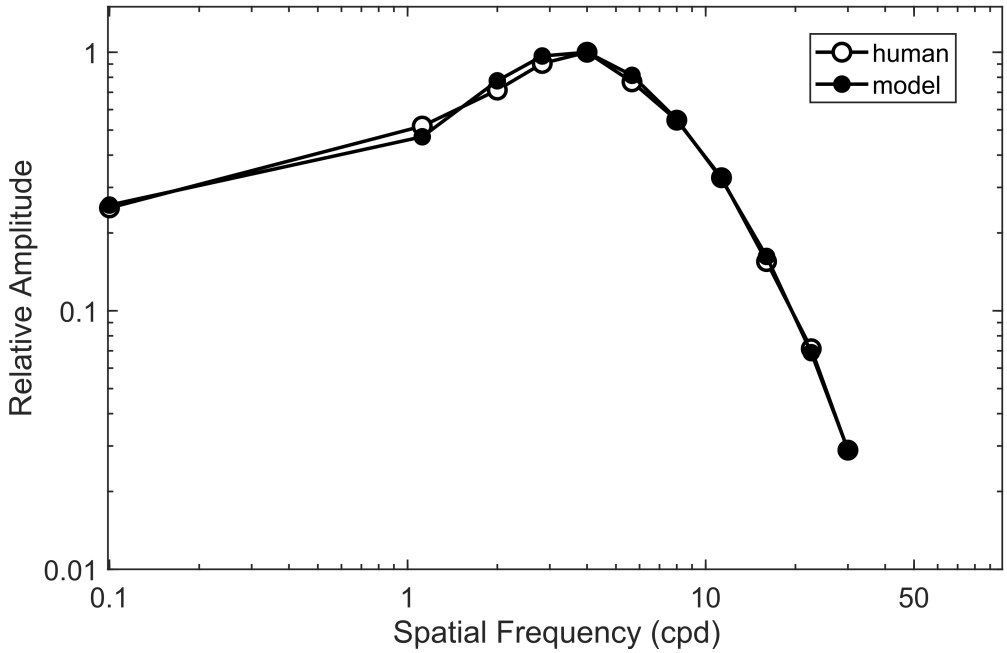


Figure 2.15: Eye filter. Open circles: the average human contrast sensitivity function normalized to a peak of 1.0. Solid circles: the fit of the eye filter equation (Equation 2.15).

Intrinsic position uncertainty is another internal factor that may contribute to the differences in thresholds across different types of targets. Intrinsic position

uncertainty refers to the phenomenon that, even if a target is always presented at exactly the same physical location on the display screen, and this condition is perfectly understood by a human observer, the human observer will still accept, intentionally or not, target-like features in (a small region of) the surrounding background [19–25], because the observer has and acknowledges the unavoidable internal noise of knowing the exact target location. In other words, slightly displaced "features" in the noise background that correspond to the target are considered as evidence of the target.

We implemented this intrinsic uncertainty into our models by applying the template over a small region centered on the actual target location, and then probabilistically selecting the maximum of the template responses as the final decision variable. For rough estimate, we assumed the specific parameters of uncertainty were the same in Michel and Geisler [23]. For an 1-octave, 6-cpd Gabor target, they estimated the standard deviation of the intrinsic position uncertainty in the fovea under a Gaussian assumption was $\sigma_U = 0.083\,\text{deg}$, about a width of 0.7 mm at an arm's length of 50 cm. Though we did not directly measure the position uncertainty in this chapter, we later designed an experimental procedure for its measurement (see Section 3.3).

In summary, the uncertain, eye-filtered, reliability-weighted template matching (UERTM) observer is

$$R_{uer}(\vec{x}) = N_{er}(\vec{x}) \cdot T_{er}(\vec{x}) \tag{2.19}$$

$$N_{er}(\vec{x}) = \mathbb{F}^{-1}\{E\,\mathbb{F}\{f_1^{-1}(\vec{x})N(\vec{x})\}\} \tag{2.20}$$

$$T_{er}(\vec{x}) = \mathbb{F}^{-1}\{E\,\mathbb{F}\{f_1^{-1}(\vec{x})T\}\} \tag{2.21}$$

$$R_{uer} = \max_{\vec{x}\in\mathbb{U},\propto p_U(\vec{x})} [R_{uer}(\vec{x})] \tag{2.22}$$

$$p_U(\vec{x}) = \frac{1}{2\pi\sigma_U^2} \exp\left(-\frac{||\vec{x}||^2}{2\sigma_U^2}\right) \tag{2.23}$$

where $\mathbb{U}$ is the set of all uncertain locations (vectors starting from the actual target location), $\vec{x}$ is a specific uncertain location, and $p_U(\vec{x})$ is the uncertainty distribution. After applying the eye filtering, we created a random binary map of ones and zeros where the probability of "1" at each pixel location was given by $p_U(\vec{x})$ normalized with a peak of 0.5. Then the maximum response among the "1" locations becomes the final decision variable. Notice the uncertain stimulus and the template were reliability weighted with a filter with values bifurcated perfectly along the actual contrast modulation boundary.

The ERTM observer predicts a slower decline in thresholds across targets than the WRTM observer (Figure 2.16a-b, Figure 2.17a-b), because the eye filter only whitens partially. This less steep change is more consistent with the measured human detection pattern.

Nevertheless, the UERTM observer is a better predictor of human performance than the ERTM observer (Figure 2.16c-d, Figure 2.17c-d). For example, the threshold difference between the raised-cosine and sine wave targets is smaller, as for human observers. That is because template responses are more correlated for targets with more low spatial frequencies. Therefore, the same degree of intrinsic position uncertainty

leads to maximum responses less different than the response without uncertainty (at the actual target location) for such targets, and thus detection thresholds increase less.

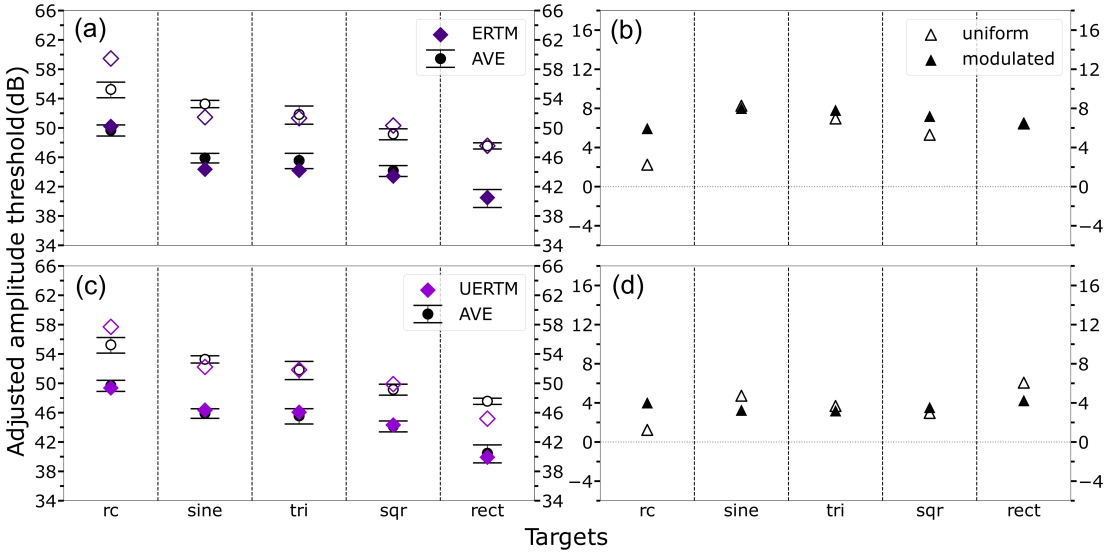


Figure 2.16: Comparison of model and human thresholds in Experiment 1. Open symbols: uniform background; solid symbols: contrast-modulated background. Black circles: average human thresholds from Figure 2.6. Colored diamonds: adjusted thresholds of the ERTM (a) and UERTM (c) model observers. Thresholds of each model were adjusted with a single scalar (0.475 and 0.654 for ERTM and UERTM respectively) to best match the human thresholds. The RMS errors in decibels are 1.663 and 1.189 for ERTM and UERTM respectively. Triangles: absolute difference of the model and human thresholds in Experiment 1, for ERTM (b) and UERTM (d) model observers.

In fact, the UERTM observer is the best model among all template matching models, with the smallest RMS error over the 18 conditions in both experiments (Figure 2.18a). The average error was about merely 1.0 dB. Notice that there is only one free parameter for each model observer (the efficiency scaling factor), so the goodness of fit varies not due to different number of parameters.

The average human observer had an efficiency factor to the UERTM observer

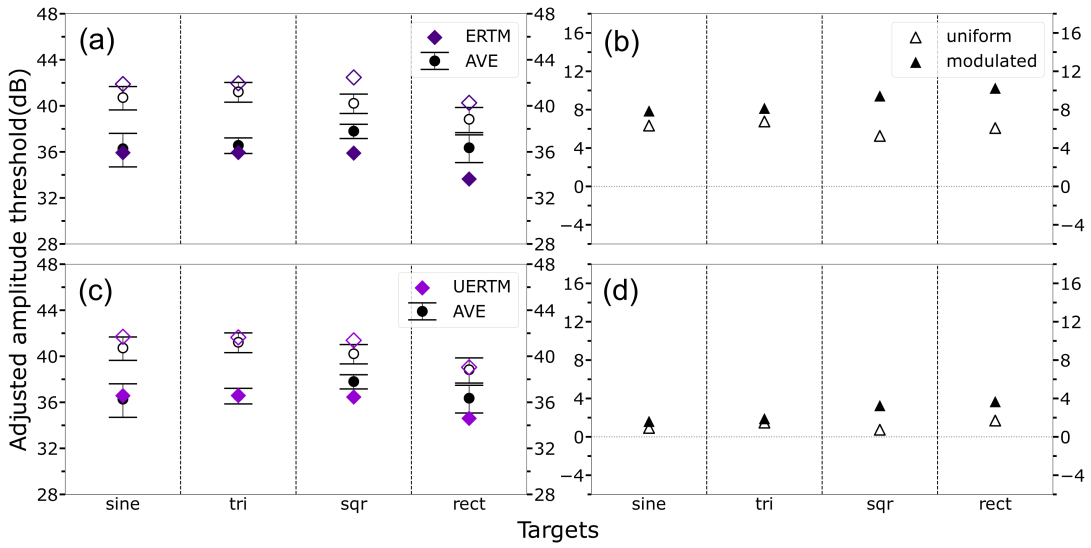


Figure 2.17: Comparison of model and human thresholds in Experiment 2. Open symbols: uniform background; solid symbols: contrast-modulated background. Black circles: average human thresholds from Figure 2.9. Colored diamonds: adjusted thresholds of the ERTM (a) and UERTM (c) model observers. Thresholds of each model were adjusted with a single scalar (0.421 and 0.803 for ERTM and UERTM respectively) to best match the human thresholds. The RMS errors in decibels are 1.603 and 0.968 for ERTM and UERTM respectively. Triangles: absolute difference of the model and human thresholds in Experiment 2, for ERTM (b) and UERTM (d) model observers.

almost as high as the TM observer (Figure 2.18b). The decreasing detectability of the UERTM model results from the intrinsic position uncertainty, with which responses pick up in the surrounding more noise but not more signal. The detection inefficiency of human observers (~ 0.3 to the WRTM observer) can be partially explained by the eye filtering and intrinsic position uncertainty.

It is out of the scope of this project to measure human detection performance in natural images and compare them with template matching models. However, we compared the performance of model observers in $1/f$ and natural background (Figure 2.19). The pattern of threshold differences in those two types of backgrounds

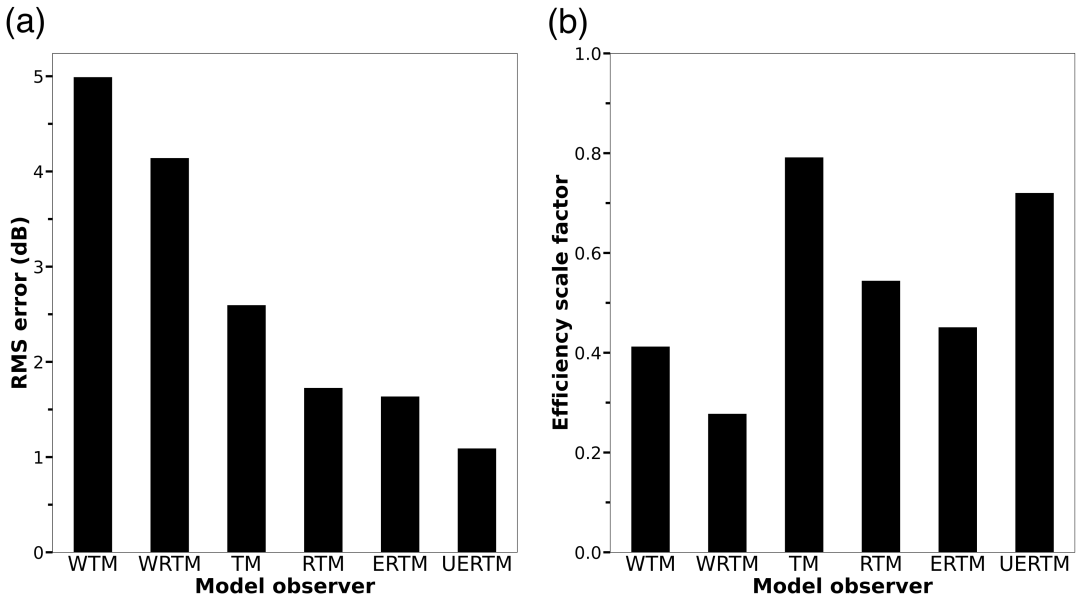


Figure 2.18: RMSE and efficiency scaling of all models to human observers. (a) RMS error averaged across all 18 conditions in both experiments. The models are ordered from the highest error to the lowest. (b) Efficiency scale factor for each model observer, with the ordering of the models the same as in (a). The further the scale factor below 1.0, the lower the model observer's threshold relative to the human threshold. Akaike information criteria (AIC) for model observers based on all conditions: WTM, 1345.47; WRTM, 701.14; TM, 478.93; ERTM, 259.78; RTM, 263.42; UERTM, 232.34.

are similar. It should be kept in mind that the WRTM observer is no longer the ideal observer for detection in natural background. Unsurprisingly, the TM observer has the highest thresholds, and the WRTM observer detects the best. The ERTM observer is about 2.0 dB better than the RTM observer, showing a modest but real benefit from the partial whitening of the contrast sensitivity function for detection in $1/f$ noise, at least for our targets. Interestingly, the partial whitening of the eye filter is even more beneficial for detection in natural background. Lastly, even the UERTM observer is limited by the intrinsic position uncertainty, it still performs slightly better than the TM observer.

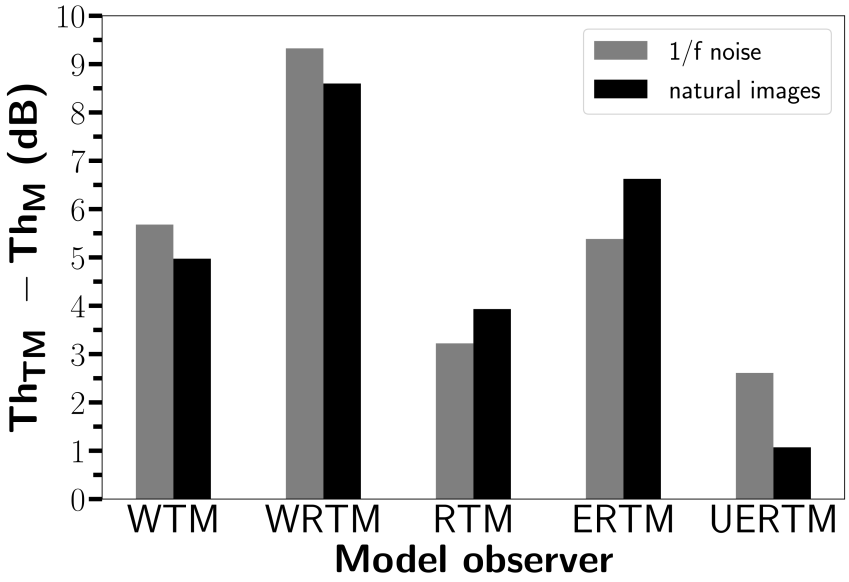


Figure 2.19: Thresholds of models in $1/f$ noise and natural images. The thresholds of all models were averaged over the ten conditions in Experiment 1. The bars are the threshold of the TM observer subtracted by the threshold of one of the five more sophisticated model observer. Light bars: in $1/f$ noise background; dark bars: in natural images. Natural images were high-resolution, calibrated, and sampled from this dataset [1].

2.6 Discussion

The ideal observer in linearly filtered Gaussian noise whitens in both space and spatial frequency. On each trial, it applies a spatial-frequency whitening filter and a reliability-weighting filter to the mean-subtracted input image, and applies both filters also to the template that has the shape of the target, and then calculates the decision variable as the inner product between the stimulus and the template that have both been whitened in space and spatial frequency.

Detection thresholds were measured for a range of targets in uniform and contrast-modulated $1/f$ noise backgrounds that have the amplitude spectrum of natural images. Human thresholds were compared to those of the ideal observer and five

suboptimal observers, where the only free parameter used to fit each observer was an overall efficiency scalar. For the model observers without biological constraints, reliability-weighting without whitening in spatial frequency leads to the detection pattern most consistent with that of human observers. Further including both the human CSF and the intrinsic position uncertainty gives rise to the best match of human performance among all models, with an average RMS error of about 1 dB.

Full whitening in spatial frequency was not found from our experiments, but the improvement of fit by including the human eye filtering is evidence for partial whitening, as reported in many previous studies [93–97]. Furthermore, the eye filter actually improves detectability in $1/f$ noise and natural backgrounds for the present range of targets.

Reliability weighting considers the contrast modulation in the background and predicts a substantial drop in thresholds. The magnitude of the drop is simulated to be approximately the same for all targets in our experiments. Humans showed a similar behavior in our experiments (also see the partial-masking effect in [79]).

The effects of non-stationary structure in backgrounds have been measured and modeled at a larger spatial scale, where the background is stationary only within the target region [109]. Hotelling and channelized Hotelling observers have been used for tasks with non-stationary backgrounds [110, 111] because those observers can take into account the covariance structure in varying regions of the background. In principle, the usage of the diagonal terms of a full-size covariance matrix in Hotelling observers can capture local reliability weighting across pixels within the target region [112].

The implementation of the intrinsic position uncertainty could have been more elegant. In Chapter 3, we incorporated the uncertainty just as a log-prior term, as

in Equation 3.4. The spatial variation in uncertainty was represented within the log-prior, not by a sampling acceptance distribution. However, even though we know the intrinsic position uncertainty must exist in the HVS, it may be more complex and nuanced than either of the modeling suggestions. For example, it can vary by target, background, actual target location, direction of shift, and the idiosyncrasy in a particular visual system.

Medical images have a lower degree of power-law regularities than natural images. For X-ray imaging of breast tissue (mammography), the power exponent β is around 3.0 [113–121]. For computed tomography of the breast (bCT), β is around 2.0 [117, 121]. This implies the methods and results of evaluating the two whitening mechanisms in this chapter are promisingly transferable to human visual detection and search in medical images.

For future research, one could use this experimental paradigm to measure and analyze human detection pattern in natural and medical backgrounds. They have a less random phase structure than the linearly filtered Gaussian noise. More relevant targets (e.g., mass and microcalcification) and modulation of local contrast (e.g., by semantic regions) could be used to ensure the conclusions are practical.

Chapter 3: Detection: Phase Similarity

Abstract

Visual detection in natural images is greatly affected by background luminance, RMS contrast, amplitude-spectrum similarity, and partial masking factor. In this chapter, we measured and analyzed the effect of similarity in phase, an essential fifth dimension, on human detection performance. We discovered the target was more detectable when it was less similar to background in spectral amplitude, and surprising more detectable when it was more similar to background in phase. When we incorporated into template matching models a small level of intrinsic position uncertainty directly measured from a position-discrimination task, this pattern of phase-dependent asymmetry emerged. The similarity in phase modulates the effect of intrinsic position uncertainty with the most likely response location attracted to and repulsed from the actual target location.

3.1 Introduction

In this chapter, I will answer how well the human visual system (HVS) makes use of the similarity of target and background in phase structure for visual detection, and what effect the similarity in phase between target and background has on human visual detection in natural background. Much content in this chapter is included in this peer-reviewed article [25].

Neural computations for detection and search in the HVS are efficient and robust in natural backgrounds. As pointed out in Section 1.5, human observers distinguish a target from a background by recognizing and exploiting the patterns, or the physical and statistical regularities in the background. To understand and predict human behavior in visual detection and search, it is essential to measure and analyze those properties. Furthermore, by simulating naturalistic or artificial images with specific statistics, one could test principled hypotheses on how the HVS processes this information in a biologically constrained and task-relevant manner.

However, most of the early studies are much limited in applying the natural scene statistics to the detection task. For example, the statistical properties of natural images were altered [105]. In other cases, only a small number of natural images were tested [122–124]. Furthermore, often the multiple-interval forced choice procedures were used [122, 124, 125], which are not the most representative of detection under natural conditions where one typically does not have the opportunity to compare the exact same image with and without the target.

Recently, Sebastian et al. [17, 79] used a constrained sampling approach to measure how various properties of natural background affect performance in a simple yes/no detection task. Specifically, Sebastian et al. [17] binned millions of patches of natural background (with the size of the target) into joint histograms along the dimensions of luminance (L), RMS contrast (C), and cosine similarity of the target and background amplitude spectra (amplitude-spectrum similarity, S_A). They then measured detection thresholds for backgrounds sampled from a sparse subset of bins across the whole space. For windowed sine wave and plaid targets added to the background, they found that amplitude threshold a_t increases linearly along all three dimensions, akin to a separable Weber's law: $a_t \propto L \cdot C \cdot S_A$ (see also [126]). Sebastian

et al. [79] found another factor, the partial masking factor $||T_p||$, so that $a_t \propto L \cdot C \cdot S_A/||T_p||$. The partial-masking factor entails that for natural backgrounds with the same luminance, contrast, and amplitude-spectrum similarity, the more the luminance and contrast varies within the target region, the more detectable the targets are (see also [80]).

Nevertheless, amplitude-spectrum similarity S_A is a phase-invariant similarity measure. While an amplitude spectrum determines the absolute and relative amplitudes of sine and cosine wave components in an image, the phase structure of an image determines how those wave components align and misalign with each other, corresponding to lines and edges in the spatial domain of the image.

Rideaux et al. [127] measured the effect of similarity in the spatial domain on the detectability of derivative-of-Gaussian targets that were added in different phases with respect to contours located within natural images. They found that thresholds were lowest when the target was in phase with the contour and highest when completely out of phase, opposite to the effect of S_A. They also found that the effects of phase-dependent similarity dominate the effects of phase-invariant similarity. However, their task was not a simple detection task, but covert visual search. The target could appear anywhere within the 2-degree backgrounds, which was a substantial level of extrinsic position uncertainty. Target locations were artificially correlated with RMS contrast through a selection procedure based on convolution response. Furthermore, their natural backgrounds had more indoor objects than outdoor scenes.

Here, we measured the effect of phase-dependent similarity for simple detection in binned natural backgrounds. Similar to Rideaux et al. [127], we found the effect of phase-dependent similarity is highly symmetric. Differently, we found both phase-independent and phase-dependent similarities are major factors of detection accuracy

in natural backgrounds.

As we applied the normalized template matching observer [17, 79] to our stimuli, we found that it predicts symmetric thresholds as a function of phase-dependent similarity, reaching the minimum when the target and background are orthogonal in phase. This prediction stays the same independent of the level of phase independent similarity. Surprisingly, when we incorporated the intrinsic position uncertainty in the HVS into the template matching observer, it predicts an asymmetric pattern of thresholds similar to that of the human observers. The uncertainty level was measured directly for the same human observers in a separate position-discrimination task for the same target in a gray background with the same diameter as the natural background in the detection task.

Overall, we concluded intrinsic position uncertainty (and the extrinsic position uncertainty in [127]) provide a plausible explanation of the asymmetric masking reported in both studies.

3.2 Measurement of detectability with varying similarities

All experimental procedures in this section were approved by the University of Texas Institutional Review Board (IRB). Informed consent was obtained from all participants. The study included three male participants, aged 20–25. They all had normal or corrected-to-normal acuity. In a trial, the observer's head was stabilized with a chin and head rest.

The stimuli in the experiments were generated with MATLAB 2022a and the Psychophysics Toolbox [101, 102]. The stimuli were displayed with a resolution of 120 pixels per visual degree on a well calibrated Sony GDM-FW900 cathode-ray-tube

(CRT) monitor. The monitor had a display size of 1920 x 1200 pixels, a refresh rate of 85 Hz, and a bit depth of 8. Prior to display on the screen, the stimuli were clipped to the upper 99th percentile gray level, gamma-compressed, and quantized to gray levels in the range of 0-255. The mean luminance of circular background patches on the screen was always 50 cd/m^2, which was equal to the luminance outside the patch.

We measured the psychometric functions with respect to the target amplitude a in a simple yes/no detection task. The target amplitude was defined to be the squared root of the sum of the squared pixel values. For plotting convenience, we divided the actual RMS amplitude by 97.8. Besides the definition in Section 1.3, our detection task here has only a single, deterministic target, and only a single possible location (the center of display) that the target can be present at. The target was a horizontal 4-cpd raised-cosine windowed sine wave target in cosine phase, added to the background. It had a diameter of 96 pixels, or 0.8 visual degrees. The background had a diameter of 516 pixels, or 4.3 visual degrees.

The natural backgrounds in the experiment were from a large database of calibrated, high-resolution (4284 $\times$ 2844), 14-bit per color images, as mentioned in [17]. They were then converted to grayscale and binned along the dimensions of luminance, RMS contrast, and amplitude-spectrum similarity. The resulting joint histograms contained 1000 bins, with 10 along each dimension. It is important to point out that those statistics were computed only from backgrounds cropped into the size of the target in this experiment, though extended regions around the patches were also displayed in a trial.

Mathematically, amplitude-spectrum similarity

$$S_A = \frac{A_T}{||A_T||} \cdot \frac{A_D}{||A_D||} \tag{3.1}$$

where A_T is the amplitude spectrum matrix of the target or the template, A_D is the amplitude spectrum matrix of the mean-subtracted background, and $||\cdot||$ is the Euclidean norm.

To closely emulate the real-world scenario of visual search and separate the effect of phase similarity from other factors, we first chose the two bins with fixed levels of luminance and contrast (the second-highest levels for both) and two different levels of amplitude-spectrum similarity (the second-lowest level and the second-highest level), with the average $S_A = 0.18$ and 0.38, respectively. Then we randomly sampled natural backgrounds from only one of the bins for all trials in a session (i.e., conditions were blocked). Similarity in phase was not blocked, but only analyzed after the experiment was completed. In a daily search, the target most likely has a random phase relationship with the background across many fixations, instead of always being in or out of phase. The luminance and contrast levels were picked as typical conditions in natural environments. The levels of amplitude-spectrum similarity were picked to allow possibly noticeable effect size while still having thousands of image patches to sample from.

Per amplitude-spectrum similarity level and per participant, there were 2000 trials (10 target amplitudes $\times$ 50 trials $\times$ 4 sessions). Target amplitude was also blocked, the same as in the detection task in the last chapter.

On each trial, a central fixation cue was given for 750 ms and then extinguished for 250 ms (Figure 3.1). Then a stimulus was displayed for 250 ms, that is the typical fixation duration during natural overt search [50, 51]. The observer was asked to

focus at the center of the stimulus display and make no saccade. The presentation duration was short enough to allow only one central fixation before response. The target was present for half of the trials, and if present, always at the very center of the display. A natural background was randomly sampled without replacement from a certain level of the amplitude-spectrum similarity. Each human observer was asked to press the left arrow key to respond "target-absent", and the right arrow key to respond "target-present". Auditory feedback was given at the end of each trial on whether the response was correct.

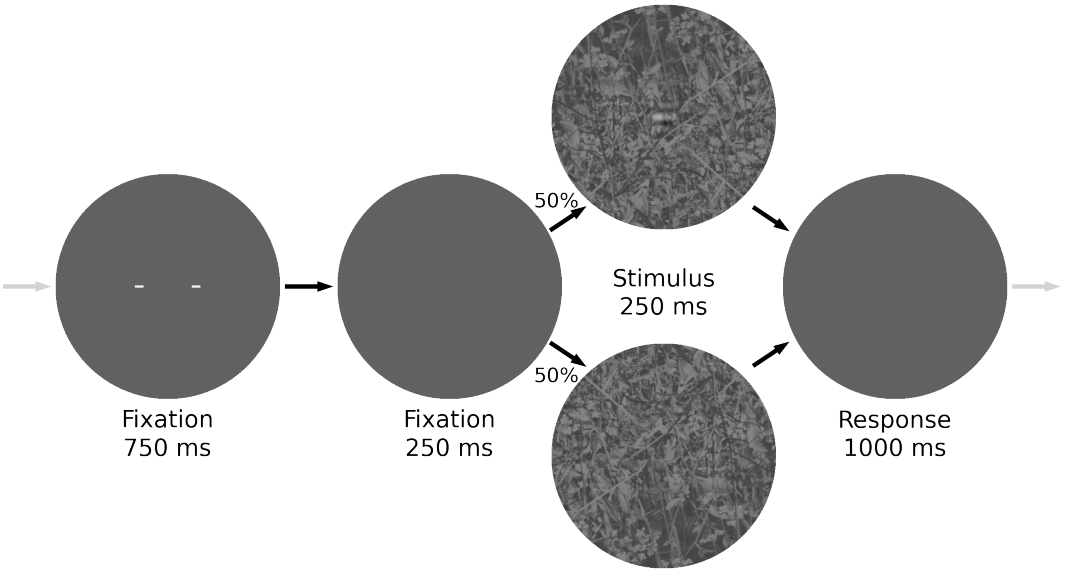


Figure 3.1: Timeline of a trial for our detection task in natural images.

The similarity in phase structure between the target and the background is defined as image similarity conditioned on the amplitude-spectrum similarity. Image similarity is simply the cosine similarity, or normalized dot product in spatial domain:

$$S_I = \frac{T}{||T||} \cdot \frac{D}{||D||} \tag{3.2}$$

In the following discussions, we may use $S_{I|A}$ instead of S_I to represent phase similarity and emphasize the fact that the natural backgrounds were first blocked in amplitude-spectrum similarity, though there is no difference in the computation of its value.

During analysis, for each level of amplitude-spectrum similarity, the backgrounds were sorted into five bins (quintiles) of image similarity. For the high amplitude-spectrum similarity condition, the levels of image similarity had average values of -0.15, -0.06, 0.00, 0.06, and 0.15. For the low amplitude-spectrum similarity condition, the levels of image similarity had average values of -0.05, -0.02, 0.00, 0.02, and 0.05.

Figure 3.2 shows examples of the natural stimuli from the two levels of amplitude-spectrum similarity and the two extreme levels of image similarity (first and fifth quntiles). As a crude visual examination without jumping into a conclusion, the target seems to be harder to see when the amplitude-spectrum similarity is high and when the image similarity is low given the same amplitude-spectrum similarity.

For each amplitude-spectrum similarity level and each target amplitude within an image similarity quntile, the hit and correct rejection rates were converted to detectability d' and criterion γ with Equation 1.6 in the description of Signal Detection Theory. Then d' was used to calculate the generalized maximum proportion correct $PC = \Phi(d'/2)$. Finally, the amplitude threshold was defined as the maximum likelihood estimate $\hat{\alpha}$ in a generalized cumulative Gaussian function:

$$PC(a|\alpha, \beta) = \Phi\left[\frac{1}{2}(\frac{a}{\alpha})^{\beta}\right] \tag{3.3}$$

where α and β are the slope and shape parameters. The amplitude threshold

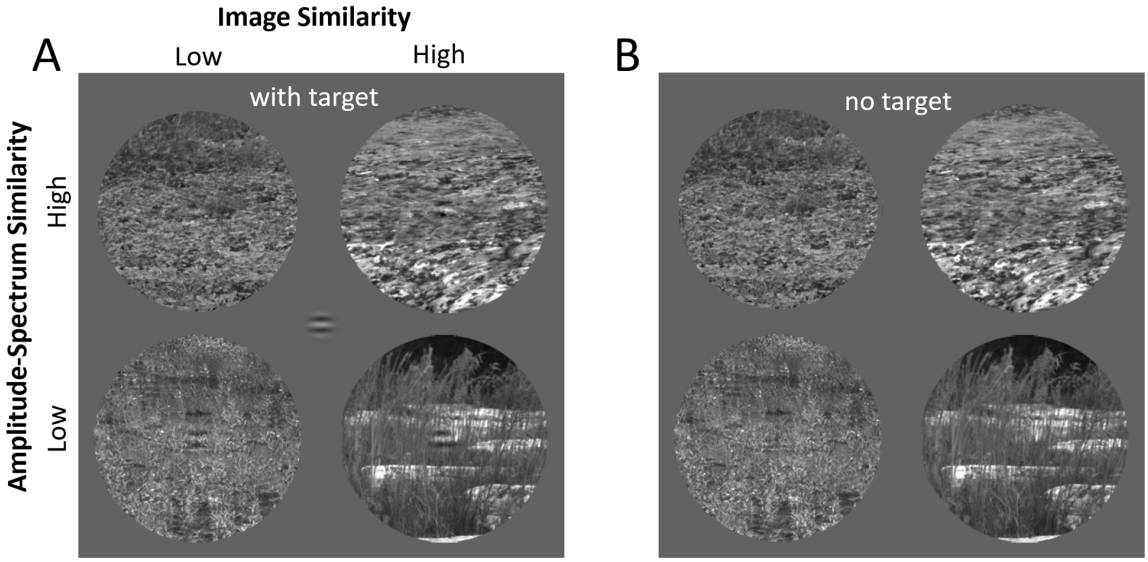


Figure 3.2: Masking by amplitude-spectrum similarity and image similarity in natural images.

corresponds to the target amplitude where $d' = 1$.

3.3 Direct measurement of intrinsic position uncertainty

Intrinsic position uncertainty refers to the phenomenon that, even if a target is always presented at exactly the same physical location on the display screen, and this condition is perfectly understood by a human observer, the human observer will still accept, intentionally or not, target-like features in (a small region of) the surrounding background [19–25], because the observer has and acknowledges the unavoidable internal noise of knowing the exact target location. In other words, slightly displaced "features" in the noise background that correspond to the target are considered as evidence of the target.

We directly measured the intrinsic position uncertainty in two of our observers (aged 20–25) with this experiment. They all had normal or corrected-to-normal acu-

ity. In a trial, the observer's head was stabilized with a chin and head rest. All experimental procedures in this section were approved by the University of Texas Institutional Review Board (IRB). Informed consent was obtained from all participants.

The target, the timeline of a trial, the size of the background, the hardware and software specifics for image generation, the response and feedback methods were identical to those in the detection experiment (Figure 3.3). However, the background content was simply uniform gray with a thin black line marking the border. The target was clearly visible with an RMS contrast of 4%. The mean luminance of circular background patches was always 50 cd/m^2, which was equal to the luminance outside the patch on the screen. Human observers were asked to report whether the target was shifted to the left or right of the center of the circular background region.

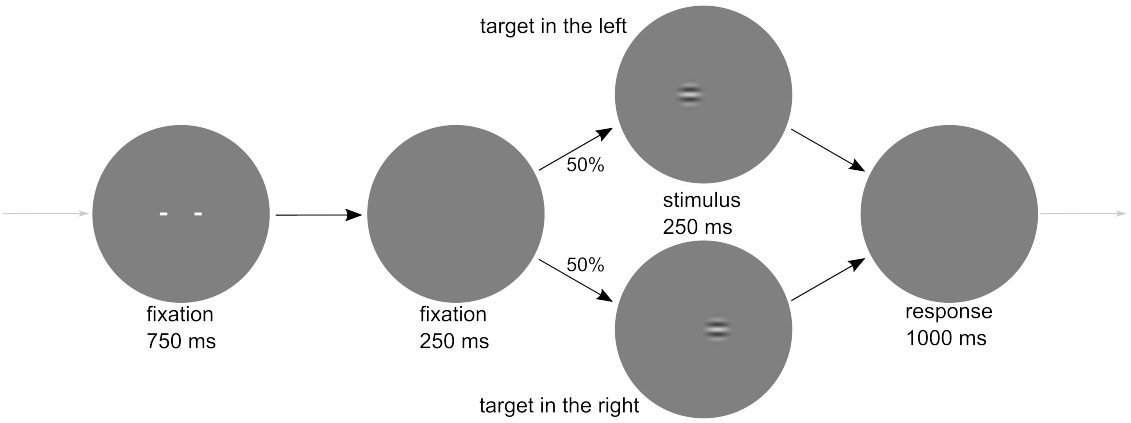


Figure 3.3: Timeline of a trial for our position discrimination task in gray background.

Psychometric functions were measured for 10 levels of location displacement with 120 trials (30 trials $\times$ 4 sessions) per level. We considered two possible distribution shapes for the intrinsic position uncertainty: two-dimensional Gaussian with standard deviation σ, and two-dimensional uniform with radius ρ. We fitted them by maximizing the likelihood to the psychometric functions. For details in the derivation

of probabilities in the confusion matrix based on intrinsic position uncertainty, see Appendix E.

The position uncertainty level in the gray background was used in the template matching model. We assumed that the uncertainty level would serve as the lower bound of the total position uncertainty in natural images.

3.4 Phase similarity effects on human and model observers

The psychometric data and fitting for the human observers in the detection task are shown in Figure 3.4. As can be seen for all observers, the bias-corrected (maximum) proportion correct gradually increases as the target amplitude increases.

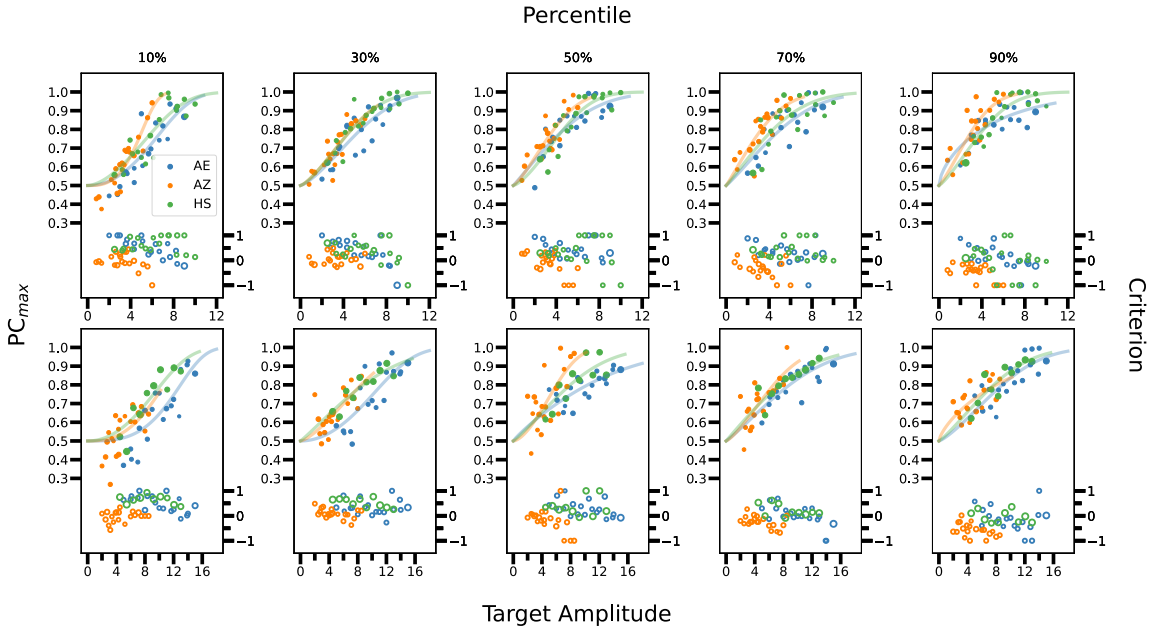


Figure 3.4: Psychometric data and fitting of the detection task. Left axis: estimated bias-corrected maximum proportion correct; right axis: criterion in the unit of the standard deviation in SDT. Top row: low amplitude-spectrum similarity; bottom row: high amplitude-spectrum similarity. Each column represents one of the quintiles.

Detection thresholds in decibels were plotted as a function of image similar-

ity for the two levels of amplitude-spectrum similarity (Figure 3.5a). For all three observers, thresholds declined monotonically with image similarity and were higher when the amplitude-spectrum similarity was higher. In other words, the target was more detectable when the target and the background were less similar in amplitude spectrum, and more similar in phase structure.

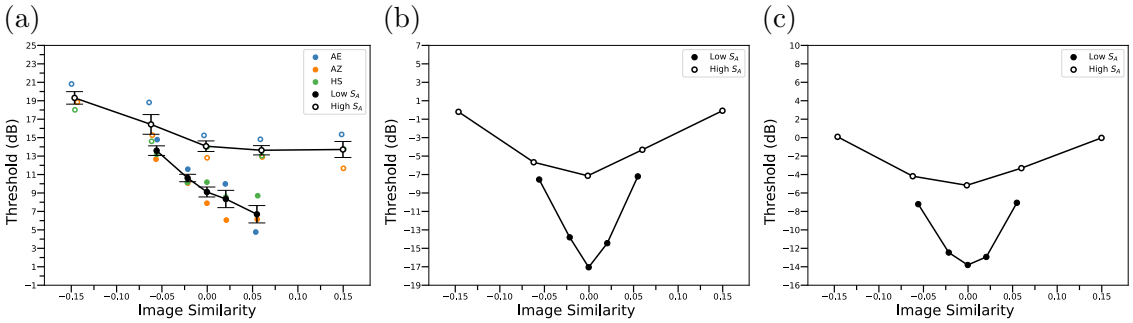


Figure 3.5: Amplitude thresholds of human (a) and model observers (b, c). (a) Colored circles: individual observers. Black circles: the average observer. Error bar: ±1 standard error across the observers. (b) Thresholds of the simple template matching model (Equation 2.2). (c) Thresholds of the eye-filtered template matching model (Equation 2.16).

However, Figure 3.5b shows that the simple template matching (TM) observer (Equation 2.2) has amplitude thresholds that do not decline monotonically with image similarity, but form symmetrical U-shaped functions with the minimum at an image similarity of 0.0, or when the target and the background are approximately orthogonal in phase. Clearly, the detection pattern of the HVS with regard to similarity in phase is very different from that of a simple template matching model, despite its past success in predicting the effects of luminance, RMS contrast, and amplitude-spectrum similarity [17].

The U-shaped thresholds with image similarity are intuitive because the template response to the target stays the same independent of the orthogonality of the

background, while the template response to the background is weaker and hence less variable when the background is more orthogonal to the target. When the target and the background are orthogonal in phase, signal-to-noise ratio and d' (Equation 1.5) are the highest and the threshold is the lowest.

Furthermore, as shown in Figure 3.5c, we still found a U-shaped function after incorporating the eye filter mentioned in Equation 2.15 with the same parameter values. The eye filter was used to model the foveal amplitude transfer function (ATF) of the early visual system, fitted to the contrast sensitivity function in the ModelFest Dataset [106], and normalized to a peak of 1.0. The shallower U-shape indicates eye-filtering causes target detectability to vary as a function of similarity in phase.

3.5 Interaction between phase similarity and position uncertainty

In this section, I will present the results of the interactive effect of phase similarity and intrinsic position uncertainty (IPU). Intrinsic uncertainty refers to the phenomenon that, even if a target is always presented at exactly the same physical location on the display screen, and this condition is understood by a human observer, the human observer will still accept, intentionally or not, target-like features in (a small region of) the surrounding background [19–25], because the observer has and acknowledges the unavoidable internal noise of the exact target location. In other words, slightly displaced "features" in the noise background are considered as evidence of the target.

Position uncertainty often does not change the qualitative pattern of detection thresholds in a model. For instance, the effect of amplitude-spectrum similarity on detection thresholds still had the same pattern, given the intrinsic position uncertainty

[128]. However, we discovered that a small amount of the intrinsic position uncertainty significantly changed the effect of similarity in phase on detection thresholds.

We directly measured the position uncertainty of two of our observers with a position discrimination task. Figure 3.6a shows the bias-corrected proportion correct as a function of the amplitude of displacement from the center of the background. As the displacement amplitude increased, the discrimination accuracy gradually increased, while the criterion remained relatively constant.

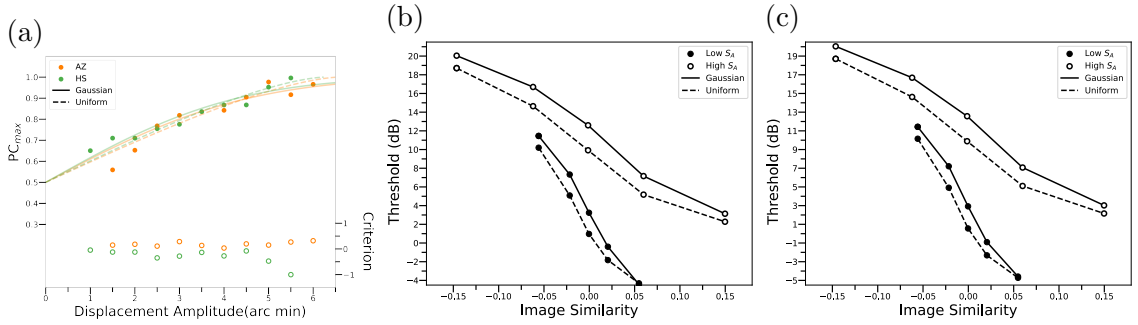


Figure 3.6: Measurement and effect of position uncertainty. (a) Psychometric data and fitting of the position discrimination task. Solid circles: bias-corrected proportion correct; open circles: criterion. (b) Thresholds of the max-UETM observer (Equation 3.4). (c) Thresholds of the sum-UETM observer (Equation 3.8).

The discrimination threshold, defined as the displacement amplitude when $d' = 1.0$, was approximately 2 arc min for both human observers. This is consistent with previous measures under similar conditions (e.g., the "bullseye" thresholds reported in [129]). Under the Gaussian assumption of the intrinsic position uncertainty (Equations E.2, E.3), the standard deviation was 3.4 arc min. Under the uniform assumption (Equations E.7, E.8), the radius was 6.5 arc min. The maximum likelihood fits using the Gaussian and uniform distributions are almost equally good, so our data in this position discrimination task do not distinguish between the two assumptions,

or some intermediate assumptions, on the shape of the uncertainty distribution.

Now we implement this level of intrinsic position uncertainty on top of eye-filtering to template matching, with a method different from the last chapter (Equation 2.22). We computed the response as

$$R_{uer} = \max_{\vec{x}\in\mathbb{U}} \left[\ln p_U(\vec{x}) + \frac{a}{\sigma_e^2} D_e(\vec{x}) \cdot T_e \right] \tag{3.4}$$

$$T_e = \mathbb{F}^{-1}\{E\,\mathbb{F}\{T\}\} \tag{3.5}$$

$$N_e(\vec{x}) = \mathbb{F}^{-1}\{E\,\mathbb{F}\{N(\vec{x})\}\} \tag{3.6}$$

$$D_e(\vec{x}) = N_e(\vec{x}) - \overline{N_e(\vec{x})} \tag{3.7}$$

where $\mathbb{U}$ is the set of all considered uncertain locations (vectors starting from the actual target location), $\vec{x}$ is a specific uncertain location, $p_U(\vec{x})$ is the uncertainty distribution, either Gaussian or uniform (using the estimated standard deviation or radius parameter), a is the target amplitude, E is the eye filter, $\overline{N_e}$ is the average luminance of the eye-filtered background, and σ_e is the standard deviation of the eye-filtered template matching response (Equation 2.16). For the high and low amplitude-spectrum similarity conditions, $\sigma_e = 250.2$ and 91.5, respectively. I abbreviate this model as the max-UETM observer.

Similarly, we have the sum-UETM observer:

$$R_{uer} = \ln \sum_{\vec{x}\in\mathbb{U}} \left\{ p_U(\vec{x}) \exp[\frac{a}{\sigma_e^2} D_e(\vec{x}) \cdot T_e] \right\} \tag{3.8}$$

The reason we express the uncertainty distribution with a log-prior is that detection under position uncertainty is effectively a visual search task without indicating the target "location" (Equation 1.25). We chose the max and sum rules because the

max rule is the maximum a posteriori search strategy when the task is to indicate either the target location or that the target is absent (Equation 1.20; see also [59]), and the sum rule is the maximum a posterior search strategy when the task is to indicate whether the target is present or absent (Equation 1.22).

Figures 3.6b and 3.6c show the performance of the max-UETM and sum-UETM observers, assuming Gaussian and uniform uncertainty distributions with parameters estimated from the position discrimination experiment. Both predictions are similar to human performance in Figure 3.5a. Indeed, the sum-UETM response is dominated by the location giving the max-UETM response. An intrinsic position uncertainty with the discrimination threshold of just 2 arc min is able to produce a highly asymmetric effect of image similarity like in human detection thresholds.

As expected, the UETM observers have thresholds substantially higher than those of the ETM observer, on average by about 20 dB, or an order of magnitude. The difference in their thresholds is least when the target and the background are most in phase (with the highest image similarity). The predicted average difference in thresholds between the low and high amplitude-spectrum similarity conditions remain about the same.

Now I explain this asymmetric effect intuitively with attraction and repulsion with Figure 3.7. When the background is in phase with the absent target, the location producing the max-UETM response or dominating the sum-UETM response tends to be attracted to the actual target location. When the background is out of phase with the absent target, the location producing the max-UETM response or dominating the sum-UETM response tends to be repulsed from the actual target location. Attraction mitigates the threshold increase due to position uncertainty, and repulsion aggravates the threshold increase due to position uncertainty. Those effects override the original

U-shaped function due to the reduced variance of orthogonal backgrounds.

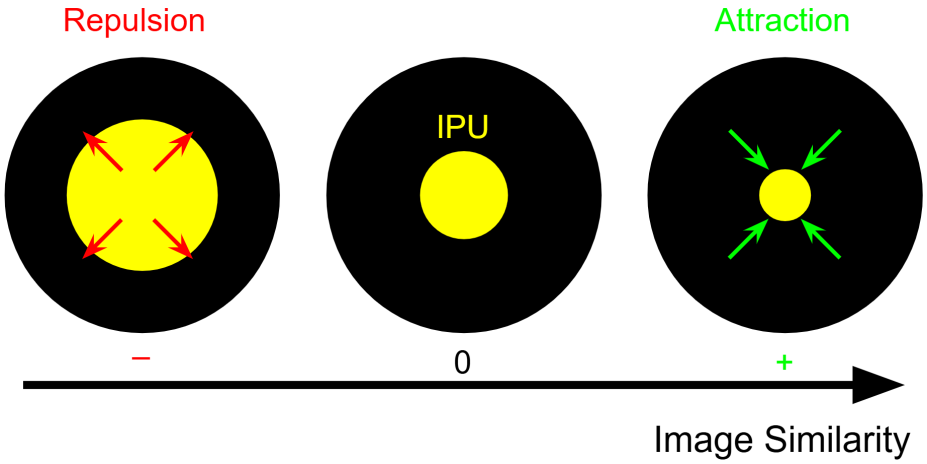


Figure 3.7: Attraction and repulsion of similarity in phase to intrinsic position uncertainty (IPU). Black circles illustrate the background. Green color indicates the target is in phase with the background. Red color indicates the target is out of phase with the background.

Furthermore, we noticed that in Figure 3.4, for the first image similarity bin where target and background were most out of phase, the bias-corrected (maximum) proportion correct was systematically below chance (a negative detectability) when target amplitude was low. That is because the stimulus was still out of phase with the template when the target was added to the background. The uncertain template response, no matter by the max or sum rule, favors locations that turn the original out-of-phase structure to be more in phase with the template. Therefore, the response under position uncertainty becomes higher when the target is absent, and lower when the target is present, opposite to the expectation of the classical signal detection model. In other words, because of the interaction between the phase similarity structure and position uncertainty, the ideal observer would reverse the inequality direction in the decision rule, responding target-present when the template of the target is less matched. With image similarity unblocked, it is highly unlikely

for the human observer to recognize such a situation and reverse the decision rule on a trial-by-trial basis to perform above chance.

As we mentioned in the introduction section of this chapter, the partial masking factor $||T_p||$ is also an impactful factor [79, 80] to the detection thresholds: $a_t \propto L \cdot C \cdot S_A/||T_p||$. However, we have not controlled its level directly. Could the asymmetric detection pattern we observed actually be the result of asymmetry in the partial masking factor? Figure 3.8 shows that is not the case. In all cases, partial masking factor has very similar distributions, and the only potential systematic effect is its slightly higher value in high amplitude-spectrum similarity condition, which would lower the thresholds uniformly in that condition by only 10%, or 0.9 dB.

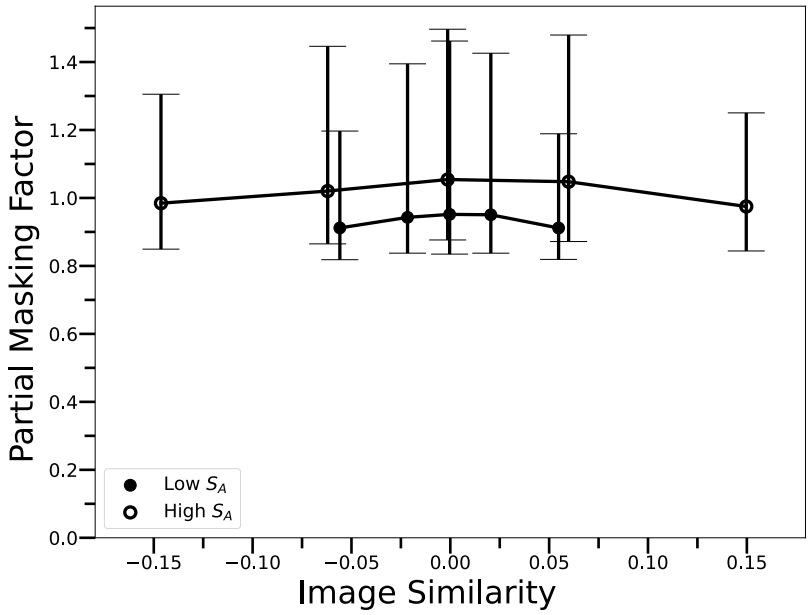


Figure 3.8: Partial masking factor of the natural images by amplitude-spectrum similarity and image similarity levels. Means and 67% confidence intervals were marked. The analyzed natural images here are the exact same images used in the experiment.

The absolute value of the cosine similarity between two real-valued vectors $\vec{u}, \vec{v}$ equals to the magnitude of the cosine similarity between their complex-valued,

Fourier-transformed vectors (termed Fourier cosine similarity), that is

$$\frac{|\vec{u} \cdot \vec{v}|}{||\vec{u}|| \cdot ||\vec{v}||} = \frac{|\mathbb{F}\{\vec{u}\} \cdot \mathbb{F}\{\vec{v}|\}|}{||\mathbb{F}\{\vec{u}\}|| \cdot ||\mathbb{F}\{\vec{v}\}||} \tag{3.9}$$

Furthermore, amplitude-spectrum similarity is no less than the Fourier cosine similarity, because the Fourier cosine similarity is maximized when the phases are totally aligned.

$$\frac{A_{\vec{u}} \cdot A_{\vec{v}}}{||A_{\vec{u}}|| \cdot ||A_{\vec{v}}||} \geq \frac{|\mathbb{F}\{\vec{u}\} \cdot \mathbb{F}\{\vec{v}|\}|}{||\mathbb{F}\{\vec{u}\}|| \cdot ||\mathbb{F}\{\vec{v}\}||} \tag{3.10}$$

As the calculation of similarities for two-dimensional matrices is no different from one-dimensional vectors, we obtain that $S_a \geq |S_i|$. Amplitude-spectrum similarity is the upper bound of the absolute value of image similarity for any target and background.

Therefore, given a fixed level of amplitude-spectrum similarity, we hypothesize a benefit for the HVS to normalize image similarity based on its range of value. We plotted the relationship between amplitude-spectrum and image similarity in varying targets and backgrounds (Figure 3.9). It is quite common, such as for the sine wave target and the $1/f$ noise background, that the quntiles of image similarity expand out as amplitude-spectrum similarity level increases. If image similarity is normalized and expressed as the exact quantiles (gray dots), then the amplitude-spectrum similarity and the normalized image similarity dimensions are orthogonalized.

When the detection thresholds of the human observers (Figure 3.5a) are plotted as a function of the normalized image similarity, as shown in Figure 3.10, threshold curves become fairly parallel, indicating a separability between the amplitude similarity and normalized image similarity for predicting detection performance.

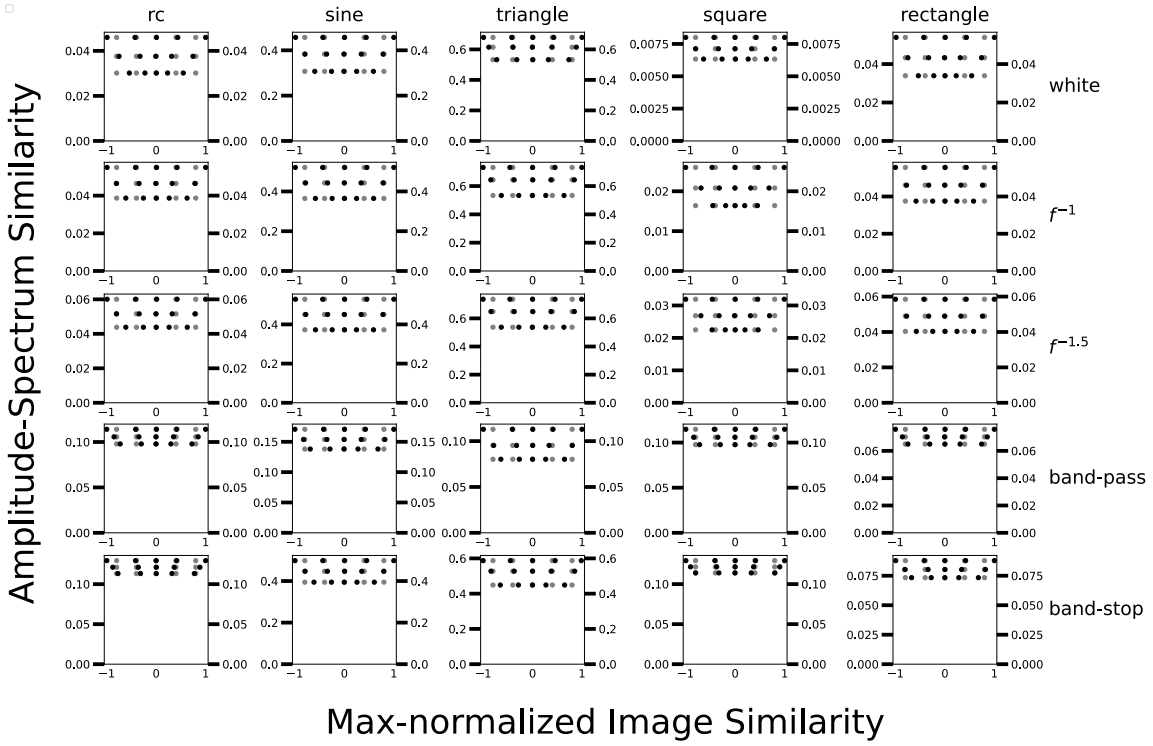


Figure 3.9: Relationship between amplitude-spectrum and image similarity. Target: a raised-cosine blob, and raised-cosine-windowed sine, triangle, square, and rectangle waves. The rectangular wave has a duty cycle of 10% (Figure 2.3). Background: white, $1/f$, $1/f^{1.5}$, band-pass, and band-stop noises (see Figure 2.1 for examples). The total number of image patches per condition is 100,000. Those patches were first binned into three amplitude-spectrum similarity levels, and then binned per S_a level into five image similarity levels. The max-normalized image similarity is the image similarity normalized by the maximum absolute image similarity across all 15 bins. Black dots: median values of each bin; gray dots: quntiles from [-1,1].

In addition to the aforementioned model observers, we also considered versions where the template responses are a mixture of simple and complex template responses. A complex template combines the response of the simple template (T) with the response of the same simple template with its phase spectrum shifted by 90 degrees ($T_\perp$), that is

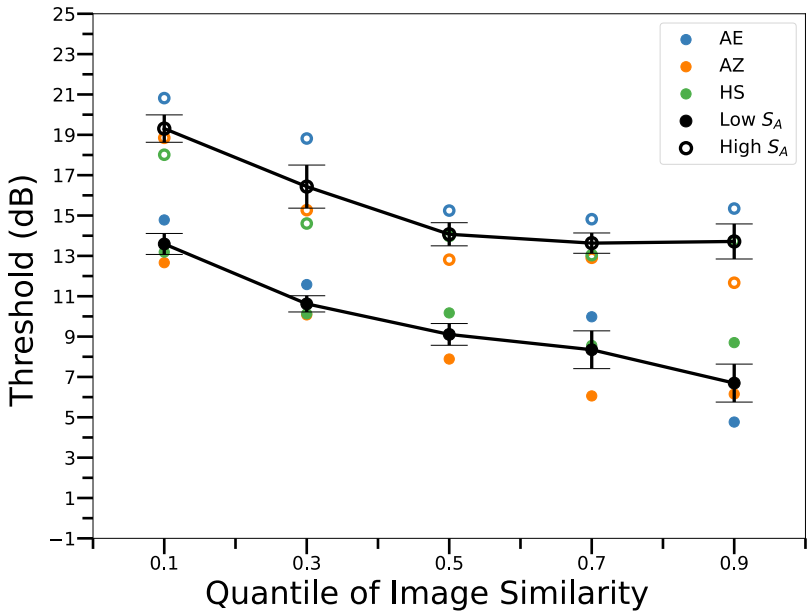


Figure 3.10: Amplitude thresholds of the human observers as a function of the quantiles of the image similarity. Colored circles: individual observers. Black circles: the average observer. Error bar: ±1 standard error across the observers.

$$R_c = \sqrt{(D \cdot T)^2 + (D \cdot T_\perp)^2} \tag{3.11}$$

where D is the mean-subtracted stimulus. As the template in our experiment is a cosine phase wavelet, the phase-shifted template will be in sine phase. Therefore, Equation 3.11 gives the response of a classical complex cell response with orientation and spatial frequency tuning matched to the target [130].

To capture the population pooling from both the simple and complex cells of the HVS for visual detection, mixture models linearly combines simple and complex template responses with a weighting parameter α ranging from 0 to 1. For example, the mixed template matching (MTM) and the mixed, eye-filtered template matching (METM) observers are

$$R_m = \alpha D \cdot T + (1-\alpha)\sqrt{(D \cdot T)^2 + (D \cdot T_\perp)^2} \tag{3.12}$$

$$R_{me} = \alpha D_e \cdot T + (1-\alpha)\sqrt{(D_e \cdot T)^2 + (D_e \cdot T_{\perp e})^2} \tag{3.13}$$

where T_e is the eye-filtered template, and $T_{e\perp}$ is the eye-filtered, phase-shifted template.

The mixed, uncertain, eye-filtered template matching (MUETM) observer is

$$R_{mue} = \alpha R_{uer} + (1-\alpha)\sqrt{R_{uer}^2 + R_{uer}'^2} \tag{3.14}$$

$$R'_{uer} = \max_{\vec{x}\in\mathbb{U}} \left[\ln p_U(\vec{x}) + \frac{a}{\sigma_e^2} D_e(\vec{x}) \cdot T_{e\perp} \right] \text{(max rule)} \tag{3.15}$$

$$R'_{uer} = \ln \sum_{\vec{x}\in\mathbb{U}} \left\{ p_U(\vec{x}) \exp[\frac{a}{\sigma_E^2} D_e(\vec{x}) \cdot T_{e\perp}] \right\} \text{(sum rule)} \tag{3.16}$$

We presented the effect of incorporating complex templates into model observers with Figure 3.11, both with and without position uncertainty. When there are only complex template responses, detection thresholds are highly asymmetrical, even without position uncertainty; simultaneously, including (uniform) position uncertainty almost has no effect on the detection thresholds.

These seemingly puzzling results can be explained by the fact that a complex template effectively sums energy over the template region independent of the spatial phase, similar to applying the sum rule over the template region. Thus, applying the sum or max rule on top of the complex template responses over the uncertainty region may produce little additional effect.

If the stimulus presentations in different image similarity bins were blocked, then in the low percentile trials, it might have been possible for the human observers

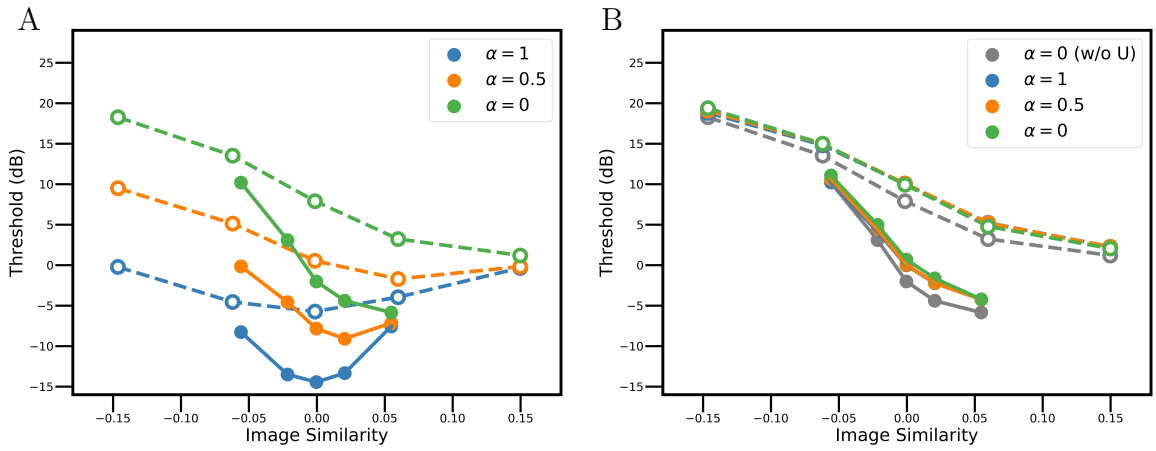


Figure 3.11: Amplitude thresholds of simple-complex mixture template matching models. A. Thresholds of the METM observer (Equation 3.13). Three cases are plotted: only-simple ($\alpha = 1$), only-complex ($\alpha = 0$), and an even mix of simple-complex ($\alpha = 0.5$). The blue curve is a re-plot of Figure 3.5c. B. Thresholds of the MUETM observer (Equation 3.14). The gray curve is a re-plot of the green curve in A.

to learn that the image similarity is negative and detect the target as a reduction in the template response. We simulated performance of the max-UETM and sum-UETM observers that can flip the inequality direction of the decision rule for each amplitude level whenever the distributions of template responses with target absence and presence are reversed in location.

As can be seen in Figure 3.12a, when the image similarity is highly negative (which is more likely given high amplitude-spectrum similarity), as the target amplitude increases from 0, the detectability first increases from 0 to a peak, then decreases back to 0, and afterward increases without a limit. This interesting pattern results from the fact that the target-present response distribution first shifts lower than and away from the target-absent response distribution, and then shifts back to and eventually cross over the target-absent distribution.

Figures 3.12b and 3.12c show the pattern of detection thresholds when the

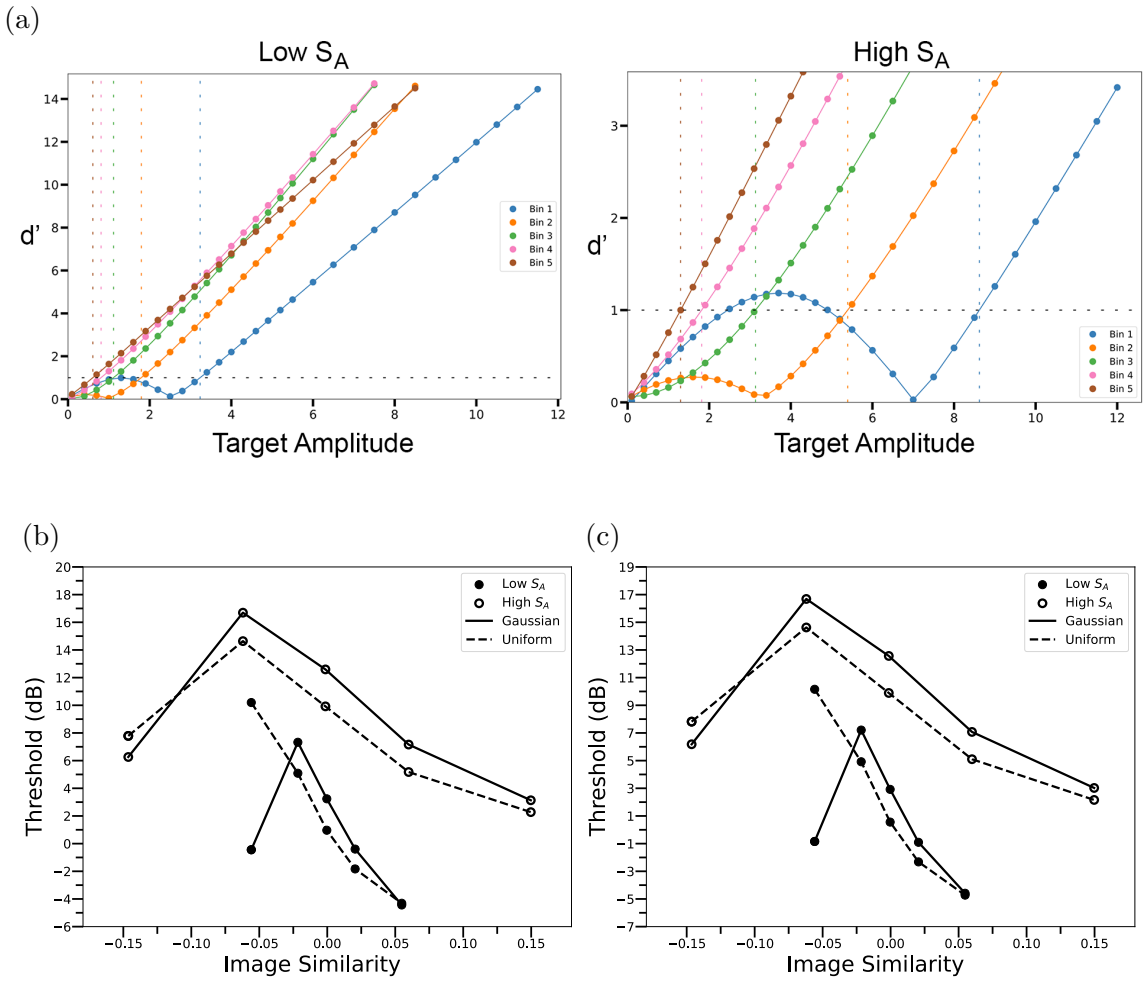


Figure 3.12: Uncertain template observers under conditions with image similarity and target amplitude also blocked (besides amplitude-spectrum similarity), in natural images. (a) Detectability as a function of target amplitude. (b) Thresholds of the max-UETM observer (Equation 3.4). Here, the threshold is defined as the lowest target amplitude that allows d'=1. (c) Thresholds of the sum-UETM observer (Equation 3.8).

image similarity and target amplitude are also blocked. Compared to when they are unblocked (Figures 3.6b and 3.6c), thresholds in this condition have a substantial drop when the image similarity is highly negative, but still maintain the general asymmetric phase effect in most conditions.

3.6 Discussion

Previous studies [17, 79] have shown the major factors that affect human visual detection in natural backgrounds include background luminance, background RMS contrast, amplitude-spectrum similarity, and partial masking factor. Rideaux et al. [127] investigated the effect of similarity in phase between derivative-of-Gaussian targets and natural background in covert visual search, and found the targets were more visible when aligned with the background in phase, but amplitude-spectrum similarity showed little effect on search performance.

Here, we measured this phase effect using the cosine similarity between a wavelet target and natural backgrounds, with fixed background luminance, contrast and two levels of amplitude-spectrum similarity. We found that both amplitude-spectrum similarity and image similarity are significant in predicting target detectability in visual detection. Phase similarity is an important fifth dimension. Specifically, a target is more detectable when it is less similar to the background in spectral amplitude, and more similar to the background in phase.

All unsophisticated template matching models (e.g., TM, ETM) failed to predict this phase-asymmetric pattern of the human observers, despite their previous success for other feature dimensions [17, 128]. Though intrinsic position uncertainty only scaled the thresholds mostly uniformly across varying levels of those dimensions, it has a very different effect for similarity in phase. We directly measured the small level of position uncertainty (2 arc min) through a position discrimination task, and incorporated it into template matching models. Surprisingly, including position uncertainty results in the asymmetry pattern along the (phase-dependent) image similarity.

Similarity in phase modulates the effect of intrinsic position uncertainty on visual detection. When target and background are in phase, the uncertain template response, no matter by sum or max rule, is attracted to the actual target location, effectively reducing the level of position uncertainty; when target and background are out of phase, the uncertain template response is repelled from the actual target location, effectively increasing the level of position uncertainty.

We admit there are still some quantitative differences in detection thresholds between the human observers and the uncertain template matching models. The predicted asymmetric effect is about 16 dB, larger than the observed effect (6-7 dB). We considered the complex templates (Figure 3.11) and incomplete templates (e.g., ignoring half of the pixels in the target), but the size of the asymmetric effect still persists. We incorporated a reasonable level of internal noise and reduced the effect size to about 10 dB. Thus, at this point, we only have a partially satisfactory explanation for why the effects of image similarity in human visual detection are smaller than predicted.

How the image similarity exactly interacts with other detection-relevant dimensions of natural images remains unknown. We have shown normalized image similarity (Figure 3.10) seems to have an effect on detection performance separable to amplitude-spectrum similarity. That was a promising first step. For further research, natural images can be binned into all five dimensions and used as backgrounds for the measurement of human detection thresholds.

Important biological factors were included in our template matching models, such as the human contrast sensitivity function, intrinsic position uncertainty, and normalization. Foveation was only acknowledged implicitly due to the nature of our detection task. However, as is well known, and as we will show in Chapters 4 and 5,

foveation needs to be incorporated in visual search tasks.

The variables that affect the level of intrinsic position uncertainty are diverse and complex. The size and shape of the target compared to those of the background, the direction of uncertainty, and the idiosyncrasy of a specific HVS, all contribute to the level of intrinsic position uncertainty. Furthermore, intrinsic position uncertainty increases rapidly with the retinal eccentricity [23]. An important direction for future research is to develop a theory that predicts the vector fields of intrinsic position uncertainty for arbitrary targets, backgrounds, and observers. In the meantime, it can be directly measured and estimated from well-designed position-discrimination tasks, paired with the main experiment, including but not limited to a visual detection or search task.

In our visual detection task, simple and complex template matching models perform equally well given the intrinsic position uncertainty (Figure 3.11). That implies in cases where energy at irrelevant phases are to be discarded for best performance, because of position uncertainty, it would still be beneficial to pool over simple and complex cell responses for a larger sample size.

We name the second zero point in Figure 3.12a as the break-even point, and the local peak before the break-even point as the early ceiling point. We noticed that if the model did not flip the direction of the decision rule before the break-even point, all d' values before that point would have been mirrored along the amplitude axis and become negative, similar to what we observed in the experiment (Figure 3.4). The amplitude values at both points depend on the similarity in shape (e.g. spatial frequency and phase) between the target and the background, the contrast of the background, and the level of position uncertainty. For future research on the effect of similarity in phase on human visual detection, one could design straightforward

psychophysics experiments to explore the existence (and quantitative relationships) of those two points.

Chapter 4: Covert Search

Abstract

Cued detection and covert search tasks of human observers were measured for a wavelet target in white noise backgrounds. The detectability map measured in the detection task was used to predict the optimal possible performance in the search task, assuming statistical independence of responses from the potential target locations. Surprisingly, we found the average human observer and all individual observers had search accuracy slighly better than the Bayes-optimal searcher, despite humans' substantial loss of sensitivity in the fovea, and the implausibility of neurally replicating the complex Bayes-optimal search rule. We show three factors that can quantitatively explain these seemingly paradoxical results: (1) Many extremely simple and fixed heuristic decision rules are sufficient to obtain near-optimal search performance. (2) Foveal neglect primarily affects only the central target location out of many potential locations. (3) Spatially correlated noise lowers detection performance but has little or no impact to search performance. These findings have broad implications for understanding visual search tasks and other identification tasks in humans and other animals.

4.1 Introduction

We have been focusing on visual detection in the last two chapters. As laid out in Section 1.2, human behavior in the single-location detection is a sensible building

block for the much more complex search behavior. In this chapter, I will present the bridging from detection to covert search (search without saccades). We discovered the surprisingly supraoptimal search performance of human observers. Much content in this chapter is included in this pre-print article [131].

In Section 1.3, I defined visual search as the aim and activity of reducing uncertainty about the location of a physical object or the distribution of particular information based on the sensing and perceiving of light within the field of vision. The search task here is specified in the following ways. First, the human observers performed no saccades during the search. Second, potential target locations were well separated by at least 1.6 visual degrees. Third, the observers were asked to respond whether the target was present or absent, and the location of the target if present.

Carefully controlled studies of covert search typically present the stimuli briefly to prevent saccades and with the potential target locations placed at a fixed distance from the fixation location to keep the target visibility at different locations approximately constant [75, 78, 132–134]. In some studies, the task is simply to indicate whether the target is present or absent. In other studies, the target is always present, and the task is to indicate the location of the target. The target is either a single shape, a limited set of directly defined shapes, or a typically unlimited set of shapes based on a semantic category. The background is made of distracting objects, or stochastic noises.

The human visual system (HVS) has foveated spatial resolution, high in the direction of gaze and rapidly declining into the periphery. Thus, this biological factor needs to be directly incorporated into any representative theory and modeling of human visual detection and search.

When all the potential target locations are at a fixed retinal eccentricity and

the task is to report whether a single target is present or absent, there is a wide range of conditions where human search accuracy is consistent with the optimal decision rule, given statistical independence at those locations [75, 76, 78, 132]. However, the design choice of a fixed eccentricity display is not representative of natural search, where potential target locations are more uniformly distributed across the visual field. Also, a target's visibility slightly varies around a circle at a fixed retinal eccentricity [55–57, 77].

Therefore, we designed well-separated target locations with varying eccentricities in white noise background. Then we measured detectability of a target at each potential target location when the location is known, and apply the detectability map through the Bayesian statistical decision theory in Section 1.9 to predict quantitatively the best possible performance in the search task when the location of the target is unknown, assuming the responses are statistically independent at the potential target locations. We also carefully interleaved the detection and search sessions to minimize differences in practice effects for the two tasks. The predictions from this strictly controlled paradigm provide the normative benchmark for evaluating the effects of various potential stimulus and neural factors on search performance. For examples, hypothesized factors that cause a Bayesian observer's performance to fall below the measured human performance can be confidently rejected.

The results are surprising. First, all four human observers performed the search task slightly better than the prediction of the Bayes-optimal search rule, given the measured detectability when the target locations were cued and the assumption of statistical independence of responses at the different locations. Second, the Bayes-optimal searcher takes into account the prior probability of the target being present at each potential location (with a prior map), as well as the detectability of the target

at each potential target location (with a d' map). It seems implausible that during the course of the experiment, the observers could precisely learn the prior map and their own d' map, and then optimally apply this information to make responses. Third, in the search task, the four observers showed a substantial loss of sensitivity in the fovea, a phenomenon coined "foveal neglect" in a recent study of covert search in continuous noise background [59]. Specifically, it was shown that the reduction in accuracy in the fovea is not due to bias in estimating the prior probability, but to a reduction in detectability in the fovea. The reduction of foveal d' was explained by the hypothesis that there is a limited total attentional gain resource and that this gain is distributed efficiently across neurons in V1.

We show that three factors can explain the seemingly paradoxical results. First, we discovered it is not necessary to know precisely the d' and target-present prior maps. Extremely crude and fixed heuristic decision rules, in combination with local normalization (e.g., luminance and contrast gain control), are surprisingly sufficient to obtain near-optimal search performance. Second, foveal neglect primarily affects only the central target location. Third, spatially correlated noise corresponding to about 45% of the total noise variance is sufficient to predict the supraoptimal search performance, even with the foveal neglect and the heuristic decision rules.

These findings have several important implications. The near-optimality of highly heuristic decision rules is a promising sign to greatly simplify the development of a general theory of visual search, without and with saccades. For example, even though under natural conditions the actual d' map changes with every fixation during visual search, the central decision mechanisms can assume a nearly fixed simple d' map and still approach optimal performance closely. These simple decision mechanisms would not die out throughout natural selection. Furthermore, the near-optimal

search performance allows the possibility of individual differences in decision rules with diverse heuristics. Also, the highly controlled experimental conditions of the current study provide even stronger evidence for the phenomenon of foveal neglect compared to Walshe and Geisler [59]. Finally, our results reveal a new potential effect of correlated neural noise for human visual search behavior.

4.2 Methodology and experiments

All experimental procedures in this section were approved by the University of Texas Institutional Review Board (IRB). Informed consent was obtained from all participants. The study included three male participants, aged 19–26. They all had normal or corrected-to-normal acuity. In a trial, the observer's head was stabilized with a chin and head rest.

The stimuli in the experiments were generated with MATLAB 2023a and the Psychophysics Toolbox [101, 102]. The stimuli were displayed with a resolution of 30 mega-pixels per visual degree (with each mega-pixel occupying a 2 x 2 screen pixel region) on a well calibrated Sony GDM-FW900 cathode-ray-tube (CRT) monitor. The monitor had a display size of 1920 x 1200 pixels, a refresh rate of 85 Hz, and a bit depth of 8. Prior to display on the screen, the stimuli were clipped to the upper 99th percentile gray level, gamma-compressed, and quantized to gray levels in the range of 0-255.

The mean luminance of circular background patches was always 60 cd/m^2, which was equal to the luminance outside the patch on the screen. Circular cues were used to indicate possible target locations. Light cues had a luminance of 66 cd/m^2, and dark cues had a luminance of 51 cd/m^2.

The overall background had a diameter of 1200 pixels, or 20 visual degrees. Backgrounds centered on each potential target location were statistically independent samples of high-contrast Gaussian noise. Each patch had a root-mean-square (RMS) contrast of 20%. For the 19-location configuration, the background had a diameter of 3.5 visual degrees, and the potential locations were at the center of background patches, separated by 4 visual degrees. The target was a vertical 6-cpd raised-cosine windowed sine wave target in cosine phase, adding to the background. It had a diameter of 48 pixels, or 0.8 visual degrees, and a single, fixed amplitude level per human observer, where the bias-corrected detection accuracy at the center of the display is approximately 95%, or a d' of 4.5.

On each trial, location cues were given for 750 ms and then extinguished for 250 ms (Figure 4.1). For the detection task, the only possible target location was cued with a dark cue, while other locations were cued with light cues. Then a stimulus was displayed for 250 ms, that is the typical fixation duration during natural overt search [50, 51]. The observers were asked to focus at the center of the stimulus display and make no saccade. The presentation duration was short enough to allow only one central fixation before response. The target was present for half of the trials, and if present, always at the very center of one of the background patches. For the detection task, the only possible target location was indicated before stimulus presentation, and the observers were asked to right-click to respond "target-absent", and left-click to respond "target-present". For the search task, the observers were asked to right-click anywhere on the screen to respond "target-absent", and to left-click within a background patch to respond "target-present" at that corresponding location. In both cases, the observers had up to 3000 ms to respond, with the cueing display presented. Auditory feedback was given at the end of each trial on whether

the response was correct.

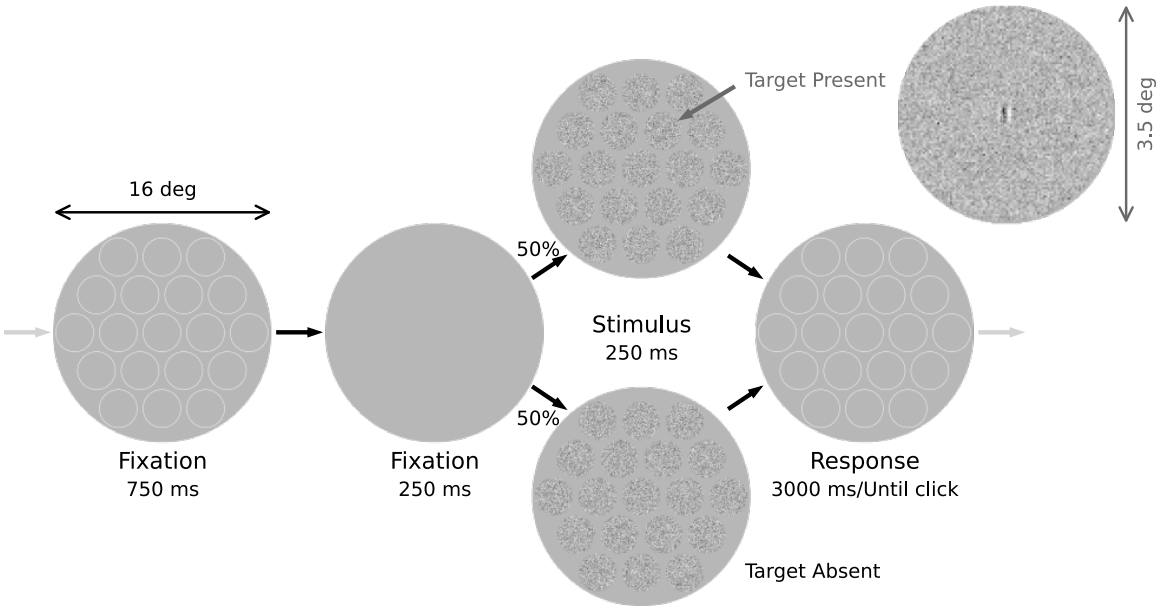


Figure 4.1: Timeline of a cued detection or search trial. For the cued detection trial, one of the light cues was replaced by a dark cue to indicate the only possible target location.

We ran preliminary search trials with highly visible targets and found that the human observers made no errors in clicking on the target locations, indicating that their search performance was not limited by spatial memory and motor control. Furthermore, we used a reverse counterbalancing design, where the observers completed the detection task and all search tasks (19, 7, 61, and 91 locations) in two opposite orders.

In the following section, I will compare human search performance with that of the Bayes-optimal and heuristic searchers. The metrics for comparison, including overall accuracy, correct rejection, hit and miss rates at each location, were calculated based on simulated responses. Nevertheless, here I list the analytic expressions of those metrics given the derivations in Appendix C.

Bayes-optimal searcher

For the Bayes-optimal searcher (Equation 4.19), the hit rate at location y is

$$p(\hat{x} = y|y) = \int_{-(\frac{lp_y}{d'_y} + \frac{d'_y}{2})}^{\infty} \phi(z) \prod_{y' \neq y} \Phi(q_h(y, y', z))dz \tag{4.1}$$

$$q_h(y, y', z) = \frac{1}{d'_{y'}} \left[\ln \frac{p_y}{p_{y'}} + d'_y z + \frac{1}{2}(d'^2_y + d'^2_{y'}) \right] \tag{4.2}$$

The miss rate at location y is

$$p(\hat{x} = 0|y) = \Phi(-\frac{d'_y}{2} - \frac{lp_y}{d'_y}) \prod_{y' \neq y} \Phi(\frac{d'_{y'}}{2} - \frac{lp_{y'}}{d'_{y'}}) \tag{4.3}$$

The false alarm rate to location y is

$$p(\hat{x} = y|0) = \int_{\frac{d'_y}{2} - \frac{lp_y}{d'_y}}^{\infty} \phi(z) \prod_{y' \neq y} \Phi(q_{fa}(y, y', z))dz \tag{4.4}$$

$$q_{fa}(y, y', z) = \frac{1}{d'_{y'}} \left[\ln \frac{p_y}{p_{y'}} + d'_y z + \frac{1}{2}(d'^2_{y'} - d'^2_y) \right] \tag{4.5}$$

The false hit rate from location y to location y' is

$$\forall y' \neq y, p(\hat{x} = y'|y) = \int_{\frac{d'_{y'}}{2} - \frac{lp_{y'}}{d'_{y'}}}^{\infty} \phi(z)\Phi(q_{fh1}(y, y', z)) \prod_{y'' \neq y, y'} \Phi(q_{fh2}(y', y'', z))dz \tag{4.6}$$

$$q_{fh1}(y, y', z) = \frac{1}{d'_y} \left[\ln \frac{p_{y'}}{p_y} + d'_{y'} z - \frac{1}{2}(d'^2_y + d'^2_{y'}) \right] \tag{4.7}$$

$$q_{fh2}(y', y'', z) = \frac{1}{d'_{y''}} \left[\ln \frac{p_{y'}}{p_{y''}} + d'_{y'} z + \frac{1}{2}(d'^2_{y''} - d'^2_{y'}) \right] \tag{4.8}$$

The correct rejection rate is

$$p(\hat{x} = 0|0) = \prod_y \Phi(\frac{d'_y}{2} - \frac{lp_y}{d'_y}) \tag{4.9}$$

A heuristic searcher

For the heuristic searcher given by Equation 4.21, The hit rate at location y is

$$p(\hat{x} = y|y) = \int_{-(\frac{lp_y}{\hat{d}'_y} + \frac{\hat{d}'_y}{2})}^{\infty} \phi(z) \prod_{y' \neq y} \Phi(q_h(y, y', z))dz \tag{4.10}$$

$$q_h(y, y', z) = \frac{1}{\hat{d}'_{y'}} \left[\ln \frac{p_y}{p_{y'}} + \hat{d}'_y z + \hat{d}'_y d'_y + \frac{1}{2}(\hat{d}'^2_{y'} - \hat{d}'^2_y) \right] \tag{4.11}$$

The miss rate at location y is

$$p(\hat{x} = 0|y) = \Phi(\frac{\hat{d}'_y}{2} - d'_y - \frac{lp_y}{\hat{d}'_y}) \prod_{y' \neq y} \Phi(\frac{\hat{d}'_{y'}}{2} - \frac{lp_{y'}}{\hat{d}'_{y'}}) \tag{4.12}$$

The false alarm rate to location y is

$$p(\hat{x} = y|0) = \int_{\frac{\hat{d}'_y}{2} - \frac{lp_y}{\hat{d}'_y}}^{\infty} \phi(z) \prod_{y' \neq y} \Phi(q_{fa}(y, y', z))dz \tag{4.13}$$

$$q_{fa}(y, y', z) = \frac{1}{\hat{d}'_{y'}} \left[\ln \frac{p_y}{p_{y'}} + \hat{d}'_y z + \frac{1}{2}(\hat{d}'^2_{y'} - \hat{d}'^2_y) \right] \tag{4.14}$$

The false hit rate from location y to location y' is

$$\forall y' \neq y, p(\hat{x} = y'|y) = \int_{\frac{\hat{d}'_{y'}}{2} - \frac{lp_{y'}}{\hat{d}'_{y'}}}^{\infty} \phi(z)\Phi(q_{fh1}(y, y', z)) \prod_{y'' \neq y, y'} \Phi(q_{fh2}(y', y'', z))dz \quad (4.15)$$

$$q_{fh1}(y, y', z) = \frac{1}{\hat{d}'_y}\left[\ln\frac{p_{y'}}{p_y} + \hat{d}'_{y'}z - \hat{d}'_y d'_y + \frac{1}{2}(\hat{d}'^2_y - \hat{d}'^2_{y'})\right] \quad (4.16)$$

$$q_{fh2}(y', y'', z) = \frac{1}{\hat{d}'_{y''}}\left[\ln\frac{p_{y'}}{p_{y''}} + \hat{d}'_{y'}z + \frac{1}{2}(\hat{d}'^2_{y''} - \hat{d}'^2_{y'})\right] \quad (4.17)$$

The correct rejection rate is

$$p(\hat{x} = 0|0) = \prod_y \Phi(\frac{\hat{d}'_y}{2} - \frac{lp_y}{\hat{d}'_y}) \quad (4.18)$$

4.3 Comparison of human and model observers in visual search

The target detectability for the average human observer is shown in Figure 4.2a. The average d' across all locations is 2.17, and the overall proportion correct is 84.2%. While there are some individual differences in these maps across the four observers, they show the same qualitative pattern: highest detectability in the fovea, intermediate at the six locations nearest the fovea, and poorest in the remaining 12 locations, with relatively lower detectability in the upper and lower visual fields (Figure 4.3, first column). This qualitative pattern is consistent with previous studies [55–57, 77].

The average detectability in the covert search task is shown in Figure 4.2b. Here, the detectability was computed from the hit rate at each target location and the overall correct rejection rate. The average d' across all locations is 2.17, and the overall proportion correct is 69.8%, which is considerably lower than that in the

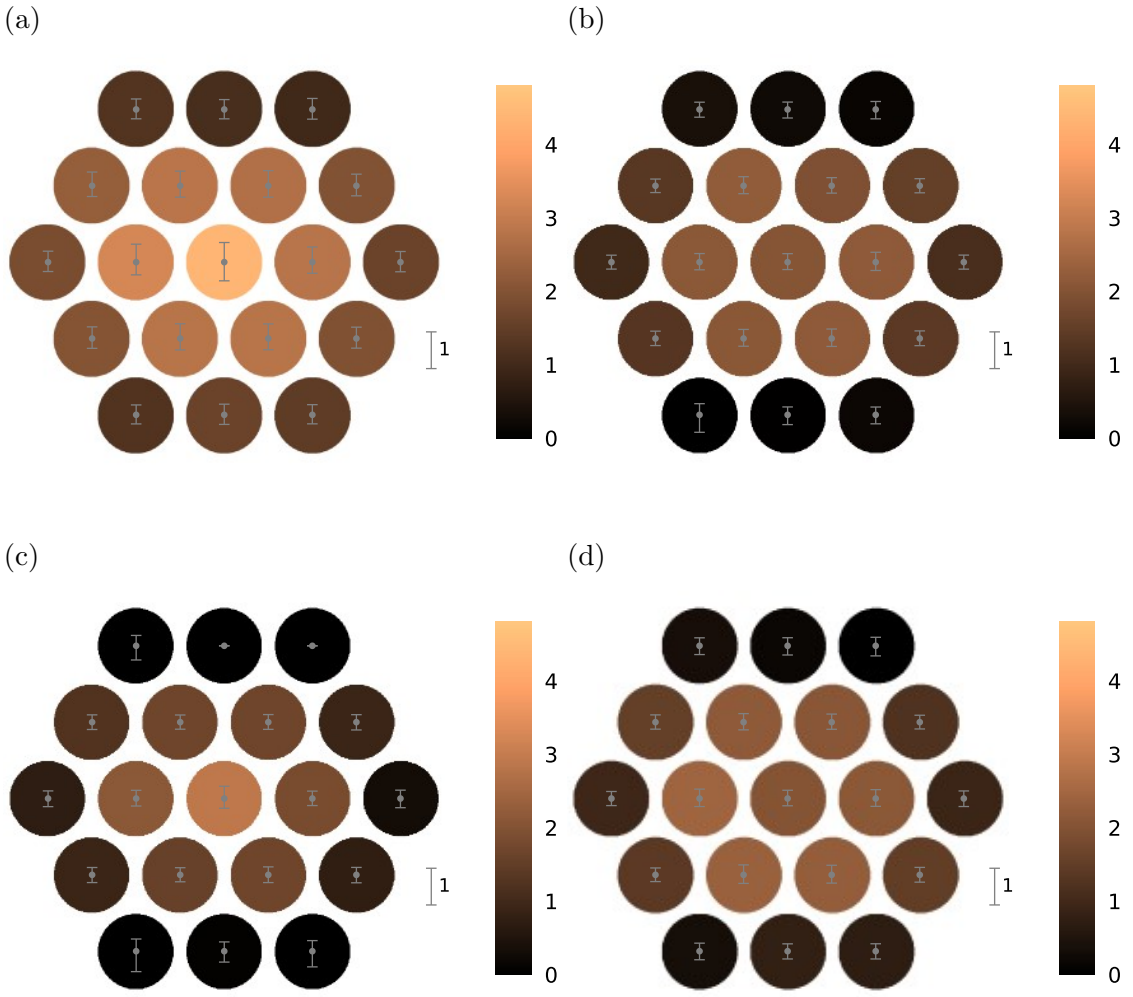


Figure 4.2: Detectability map in the detection and search tasks. (a) Average human d' map in the detection task. (b) Average human d' map in the search task. (c) The d' map of the Bayes-optimal searcher given the average human d' map in the detection task. (d) The d' map of the best-fit heuristic searcher given the average human d' map in the detection task, with correlated noise and foveal neglect. Error bars are bootstrapped 95% confidence intervals.

cued detection task. This pattern is seen in all four observers (Table 4.1). Although there is a falloff in d' with eccentricity, the d' values within the central 7 locations are much more similar than those in the detection task. This pattern remains true for

individual observers (Figure 4.3, second column).

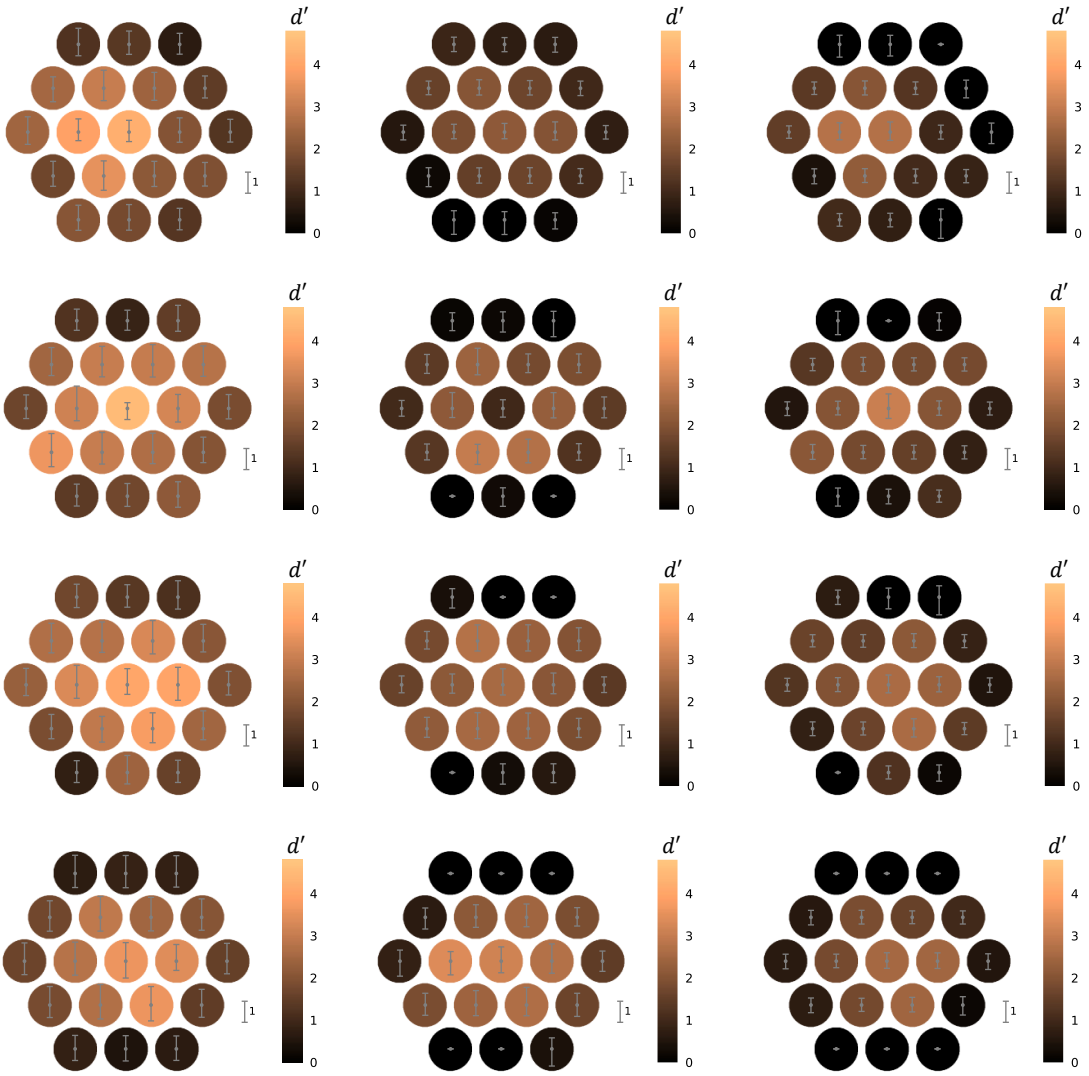


Figure 4.3: Detectability map for four individual observers in the detection and search tasks. Rows 1-4 corresponds to individual observer P1-P4. First column: d' map in the detection task; second column: d' map in the search task; third column: d' map of the Bayes-optimal searcher given the individual d' map in the detection task. Error bars are bootstrapped 95% confidence intervals.

The gray bars in Figure 4.4 show the pattern of correct responses and errors across retinal locations averaged over all four human observers. Note that the false

hit rate is the proportion of trials where the observer reported a location different from the target location that was in the region of interest.

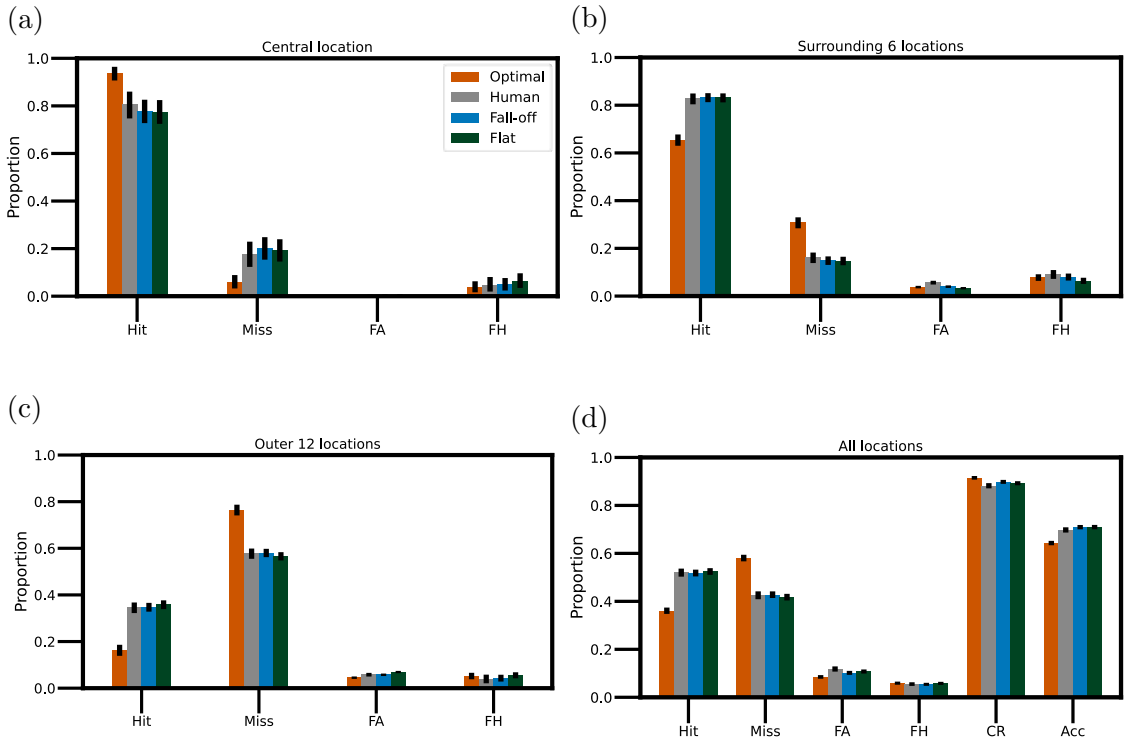


Figure 4.4: Correct responses and errors in the (19-location) search task by retinal eccentricity. (a) Histogram of hits, misses, false alarms (FA) and false hits (FH) in the central location, for the average observer (gray), the Bayes-optimal searcher given the d' map of the average observer in detection (orange), the best-fit heuristic observer given an assumed d' map that falls off, the average human d' map in the detection task, correlated noise and foveal neglect (blue), and the best-fit heuristic observer given a flat assumed d' map, the average human d' map in the detection task, correlated noise and foveal neglect (dark green). (b) Histogram of the surround six locations. (c) Histogram of the outer 12 locations. (d) Histogram for all locations. The correct rejection rate and overall accuracy are also included. (Number of trials N=6800. Error bars are bootstrapped 95% confidence intervals. Fall-off heuristic: log-likelihood = -11758, AIC = 23529, BIC = 23570. The flat heuristic is worse: log-likelihood = -11787, AIC = 23584, BIC = 23618. The Bayes-optimal searcher is the worst: log-likelihood = -12039, AIC = BIC = 24078. The fall-off model is $e^{274.5}$ times as probable as the Bayes-optimal searcher.)

To understand the relationship between the performance in the detection and covert search tasks, we simulated the behavior of Bayes-optimal covert searchers that use the Bayes maximum a posteriori decision rule (follow Section 1.9 for its derivation)

$$\hat{x} = \underset{x \in \mathbb{X}}{arg\,max} \left[\ln p_x + d'_x (R'_x - d'_x/2) \right] \tag{4.19}$$

where $\mathbb{X}$ is the target location set that also includes the target-absent "location", x is a potential target location, $\hat{x}$ is the estimated target location, p_x is the prior probability that the target is at location x, d'_x is the detectability of the target at location x (that we measured in the cued detection task), R'_x is the normalized response on that trial at location x. Recall that we have directly measured d'_x in the cued detection task.

Figure 4.2c and the orange bars in Figure 4.4 show covert search performance using the optimal decision rule, given the measured d' map in Figure 4.2a. The uniform target-present prior probability was used in the experiment, and statistical independence of responses was assumed from potential target locations. This statistical independence is plausible because the targets were small, the potential target locations were well separated, and the random noise backgrounds were statistically independent. While there is some general qualitative agreement between the Bayes-optimal searcher and the average human searcher, several puzzling differences emerge.

First, the overall accuracy of the average human observer is slightly higher than that predicted by the optimal decision rule (Figure 4.4d). This is also true for the individual human observers and for different numbers of potential target locations (Table 4.1).

Second, although the overall accuracy of the human observers is higher than

the prediction of the Bayes-optimal decision rule, their performance is suboptimal at the central location (Figure 4.4a; also compare Figures 4.2b and 4.2c). This result remains true for the four individual observers (Figure 4.3, second and third columns), which confirms the foveal neglect phenomenon in a recent study of covert search in continuous noise background [59]. In principle, foveal neglect is expected to guarantee that human searches worse than the Bayes-optimal searcher.

Finally, it is implausible that human observers implement calculations exactly equivalent to the Bayes optimal decision rule. The optimal decision rule requires weighting the response at each potential target location by the detectability of the target at that location and adding the log prior probability of the target appearing at that location (Equation 4.19). Learning all 19 detectability-prior pairs in the course of the experiment seems unlikely. Worst yet, under natural conditions the d' map is different on every fixation, even for the same target, because the masking properties of the background are different on every fixation. Also, the prior probability map varies depending on the scene context. For the optimal decision rule to be implemented under natural conditions, the HVS would need sophisticated neural mechanisms to estimate in parallel, during each fixation, the d' map over the visual field for any desired target, and the prior map from the current scene context. If the human observers were indeed using heuristic decision rules, how could they exceed the performance predicted by the optimal decision rule?

We argue there are three factors that together could explain the results. First, a wide range of extremely simple heuristic decision rules can achieve near-optimal overall search performance. Second, correlated neural noise causes the measured d' values in the detection task to be an underestimate of the effective d' values in the search task. Third, foveal neglect primarily affects only the central location of out of

the 19 location.

We compared performance of the Bayes-optimal searcher and heuristic searchers for a wide range of possible d' maps. We constructed each d' map as a function of eccentricity e (distance from the fovea in degrees of visual angle), with a peak value parameter d'_{max} and a half-fall eccentricity parameter e_2, so that

$$d'(e) = \frac{d'_{max} e_2}{e + e_2} \tag{4.20}$$

For example, the human d' map in the detection task (Figure 4.2a) is best fit with a d'_{max} of 4.69 and an e_2 of 4.68 visual degrees. In Figure 4.5 we considered 25 conditions of the actual d' maps (or the baseline d' maps for Figure 4.5f), with d'_{max} taking the values of 3.0, 4.5, 6.0, 7.5, 9.0 and e_2 taking the values of 1, 3, 5, 7, and 9. The range of e_2 approximately matches the falloff rates in detectability for wavelet targets, from about 1 to 16 cycles per visual degree [135, 136], covering most values of spatial frequencies under the human contrast sensitivity function. Generally, fine targets have small values of e_2, and coarse targets have large values of e_2.

As shown in Figures 4.5d and 4.5e, the overall search accuracy of the Bayes-optimal searcher increases rapidly as either e_2 or d'_{max} increases, when the target is present in half of the trials or always present. That is expected because the actual d' map is holistically higher when either parameter is higher.

A heuristic searcher is any model that does not use the same decision process as the Bayes-optimal searcher. For now, we consider a family of models with heuristic d' map only, that is

$$\hat{x} = \underset{x \in \mathbb{X}}{arg\,max} \left[\ln p_x + \hat{d}'_x (R'_x - \hat{d}'_x / 2) \right] \tag{4.21}$$

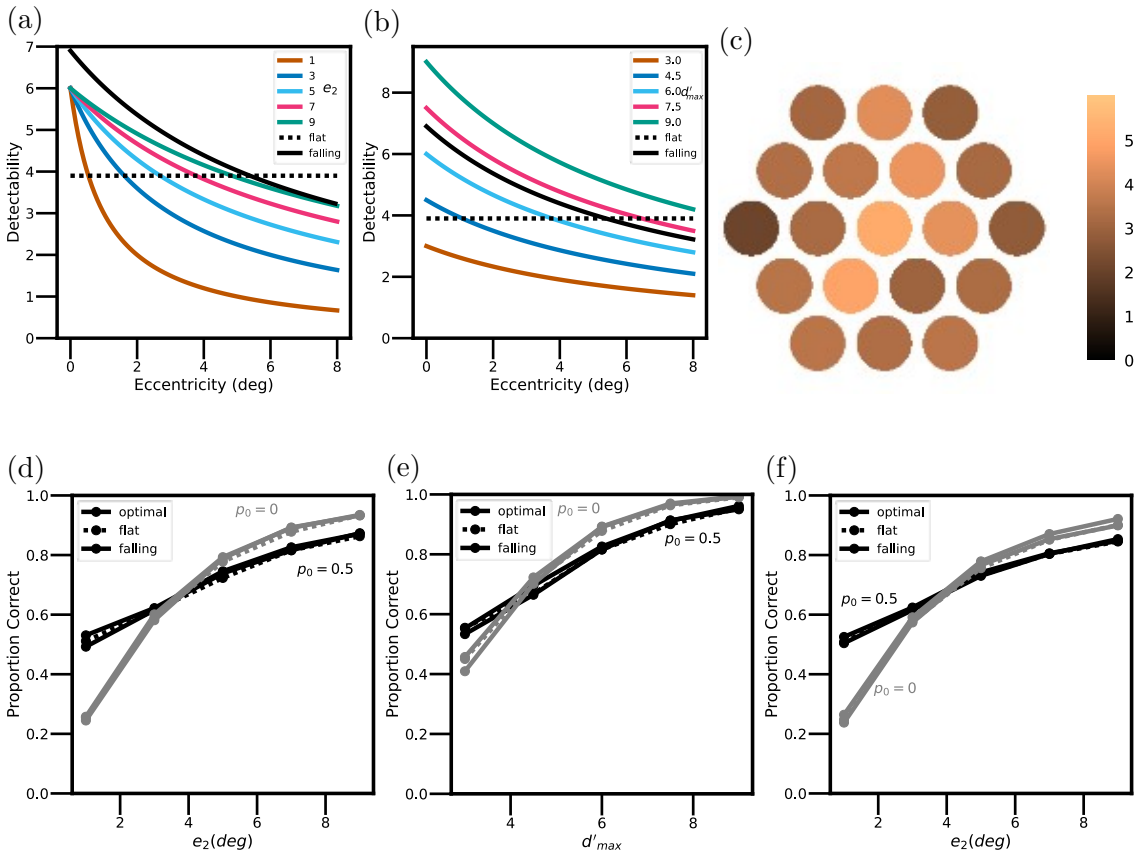


Figure 4.5: Optimal and heuristic searchers. (a) Actual d' maps with a d'_{max} of 6.0 and a range of e_2, in colored curves. The best-fit (across all 25 conditions) heuristic with a flat d' map has a d'_{max} of 3.9 (and $e_2 = \infty$), in the dotted line. The best-fit (across all 25 conditions) heuristic has a d'_{max} of 6.9 and the same fall-off rate as the best fit fall-off rate of the average human observer in search ($e_2 = 7.0$), in the dashed curve. (b) Overall search accuracy for Bayes-optimal and heuristic searchers in (a), with the target absent rate p_0 of 0.5 and 0.0. (c) Actual d' maps with an e_2 of 7.0 and a range of d'_{max}, in colored curves. The two heuristic searchers are the same as those in (a). (d) Overall search accuracy for Bayes-optimal and heuristic searchers in (c), with the target absent rate p_0 of 0.5 and 0.0. (e) An example of the d' map that varies randomly per trial. The baseline d' map has a d'_{max} of 6.0 and an e_2 of 7.0. (f) Overall search accuracy for Bayes-optimal and heuristic searchers for the baseline d' maps with a d'_{max} of 6.0 and e_2 ranging from 1 to 9 visual degrees.

where $\hat{d}'$ is the assumed d' map in the decision process.

One of the simplest heuristic decision rules is to assume a completely flat d' map. The best fit heuristic searcher with a flat d' map has search performance nearly identical to that with optimal decision rules (Figures 4.5d and 4.5e). A slightly more complex implementation is to assume a d' map with a fixed peak and a fixed fall-off rate in the decision process. The best fit heuristic search with a fall-off d' map also has search performance nearly identical to that with optimal decision rules. Both observations hold whether the target is present in half of the trials or always present.

We also found a wide range of the assumed fall-off rates $\hat{e}_2$ give almost equivalent levels of overall search accuracy (Figures 4.6a and 4.6c), as long as the assumed peak $\hat{d}'_{max}$ is adjusted accordingly to maximize search accuracy. That means, the shape of a simple heuristic is not a major factor on the performance lag of that heuristic searcher to the Bayes-optimal searcher. Over the 25 conditions with a wide range of combinations of the actual d'_{max} and e_2, the best-fit heuristic searcher maintains a performance lag less than 4% in most cases (Figures 4.6b and 4.6d).

Under natural conditions, the properties of the background scene vary over space, and hence the d' map is generally different with every new fixation. We simulated this situation by starting with a baseline d' map and varying per trial the actual d' value at each location, according to a normal distribution with a standard deviation of 20% of the base value. Figure 4.5c shows a single example of a random d' map, where the baseline d' map has a d'_{max} of 6.0 and an e_2 of 7.0. Figure 4.5f shows the heuristic decision rules with fixed flat and fall-off d' maps still search nearly optimally even when the actual d' map randomly varies on each trial.

Given the varying shapes of 25 baseline d' maps, a heuristic searcher maintains a performance lag less than 4% in most cases (Figures 4.7a and 4.7b). When the

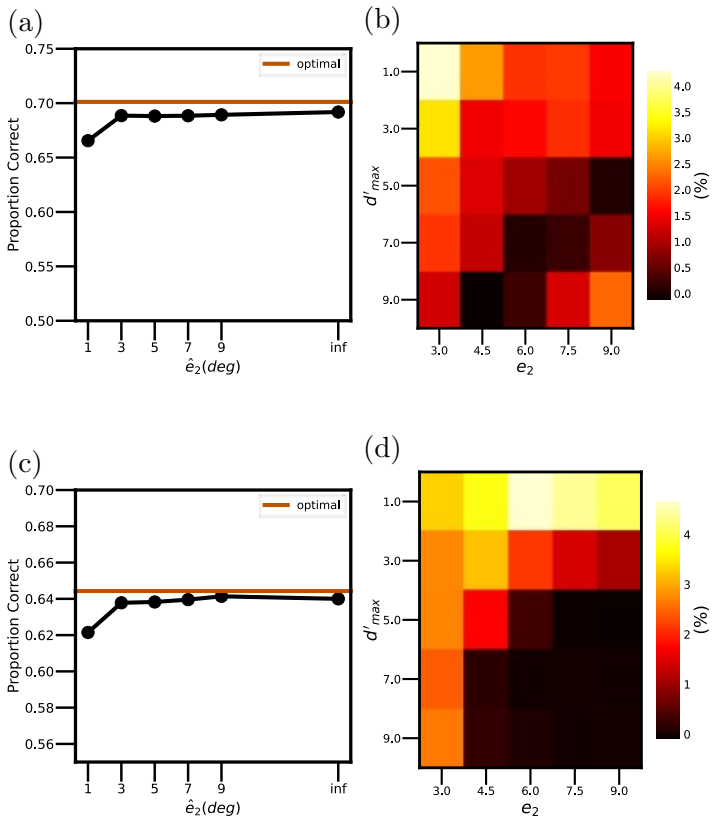


Figure 4.6: Comparison of Bayes-optimal and heuristic searchers. (a) The overall proportion correct over 25 conditions with d'_{max} = 3, 4.5, 6, 7.5, 9 and e_2 = 1, 3, 5, 7, 9, for the Bayes-optimal (orange) and heuristic (black) searchers. The Bayes-optimal searcher uses the optimal decision rule in each condition, while the heuristic searchers use a fixed assumed d' map across all 25 conditions. For each heuristic, the assumed fall-off rate $\hat{e}_2$ is first fixed, and then the assumed peak $\hat{d}'_{max}$ was fitted to maximize overall proportion correct across all 25 conditions. The target-absent prior was 0.5. (b) The heatmap of the performance lag, defined as the difference between the proportion correct of the optimal search and that of a fixed heuristic ($\hat{d}'_{max} = 6.9$, $\hat{e}_2 = 7.0$, fitted to maximize overall proportion correct across all 25 conditions). (c) The overall proportion correct as in (a), but with a target-absent prior of 0.0. (d) The heat map of the performance lag as in (b), but with a target-absent prior of 0.0.

target-absent prior is 0.5, the average performance lag across all conditions is 1.57%. When the target-absent prior is 0.0, the average performance lag across all conditions is 2.08%.

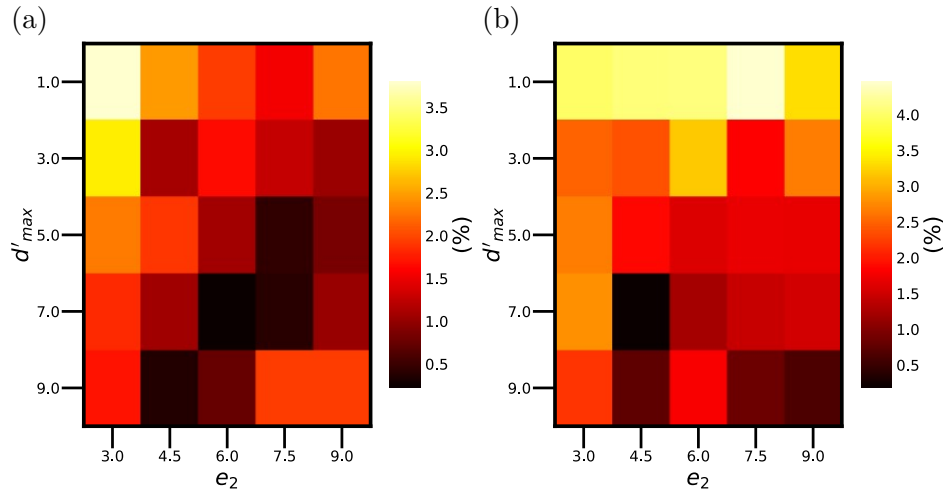


Figure 4.7: Comparison of the Bayes-optimal searchers and a heuristic searcher given random d' maps. (a) The heatmap of the performance lag, defined as the difference between the proportion correct of the Bayes-optimal search and that of a fixed heuristic ($\hat{d}'_{max} = 6.9$, $\hat{e}_2 = 7.0$, fitted to maximize overall proportion correct across all 25 conditions). Baseline d' maps $d'_{max} = 3$, 4.5, 6, 7.5, 9 and $e_2 = 1$, 3, 5, 7, 9. In each trial, the actual d' map is a random sample of the multi-variate independent Gaussian distribution, with the baseline d' map as the mean and 20% of the mean value as the standard deviation. The Bayes-optimal searcher uses the exact sampled d' map on every trial. The target-absent prior was 0.5. (b) The heatmap of the performance lag of the same heuristic searcher to the Bayes-optimal searcher when the target-absent prior was 0.0.

These results strongly suggest that the HVS uses a highly heuristic rule in covert search, given little or no benefit in implementing the optimal rule. Given that many heuristics achieve near-optimal overall search performance, can we estimate the specific heuristic rule used by human observers from behavioral data? The answer appears to be a partial yes. But we need to first incorporate two biological factors into our search model—foveal neglect and correlated neural noise.

The efficiency-limited-foveated (ELF) searcher proposed by Walshe and Geisler [59] allocates attentional sensitivity gain in V1, based on the retinotopic map of the

primary visual cortex [137]. Here, we use the same retinotopic mapping for gain modulation. Figure 4.8 shows how the backgrounds at the 19 locations map between retinal space and cortical space. With the same area size in retinal space, background patches with smaller eccentricity occupy more area in V1. The location relative to the map center is flipped both horizontally and vertically (independent to the vertical flipping from image space to retinal space). For example, a top left location in the retina corresponds to a location at the lower right occipital lobe (and the right temporal lobe downstream).

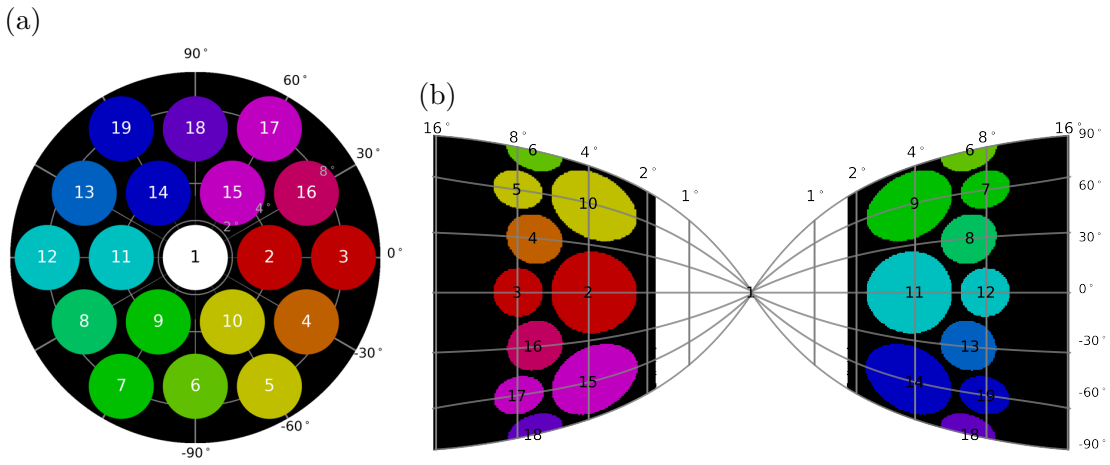


Figure 4.8: Retinotopic mapping of background patches between (a) retinal space and (b) cortical space. Colors indicate the orientation of a location with regard to the display center. Iso-orientation and iso-eccentricity contours are matched in two plots and marked with gray lines.

A fixed total amount of attentional sensitivity gain is distributed over this retinotopic map and based on the measured anatomical density of the ganglion cells in the human retina [41, 42]. The attentional sensitivity gain map in V1 is modeled as a Weibull function

$$g(\vec{i}) = g_f + (g_p - g_f)\left[1 - e^{-(i/a)^b}\right] \tag{4.22}$$

where $\vec{i} = (i, j)$ is a coordinate in V1, i represents a coordinate along the horizontal meridian direction, g_f, g_p are the gains in the fovea and periphery, respectively, and a, b are the steepness and shape parameters of the gain modulation. The attentional sensitivity gain $g(x)$ at each location x is applied as

$$x = x(\vec{i}) \qquad d'_g(x) = g(x)d'(x) \tag{4.23}$$

We assume the existence of noise with common sources that are added to the responses at all potential target locations. These common sources cause the total noise at the different locations to be partially correlated. For simplicity, we further assume that the independent noise and common noise are both Gaussian distributed, with standard deviations of σ and σ_0, respectively. Thus, the total noise variance at each target location is $\sigma^2 + \sigma_0^2$. In the detection task, the correlated noise component necessarily lowers the detectability. The detectability in detection is $d'_d = a/\sqrt{\sigma^2 + \sigma_0^2}$. However, the common noise has little or no effect on the optimal decision rule in the search task. For example, with a heuristic rule using a flat d' map, the effect of the correlated noise on d' is cancelled out by the max rule, due to the same amount of increase or decrease for responses at all potential target locations in each trial. Then the detectability in search is $d'_s = a/\sigma$. Because of the correlated noise, the peripheral gain parameter g_p can exceed 1.0. The estimated value of the periphery gain provides an estimate of the proportion of the total variance due to correlated noise:

$$\frac{\sigma_0}{\sigma + \sigma_0} = 1 - \hat{g}_p^{-2} \tag{4.24}$$

Also, we note that the optimality of the max rule still holds even when the response is correlated across locations (see Section 1.9).

Combining these two factors, we found the heuristic searcher that best fits the average human performance in covert search. This heuristic searcher is fitted with a maximum likelihood method focusing on 13 metrics; they are the overall correct rejection rate plus the hit, miss, false alarm, and false hit rates at the central location, the surrounding six locations, and the outer 12 locations. The best fitting model has a foveal gain of 0.780, a peripheral gain of 1.348, an assumed d'_{max} of 3.5, and an assumed e_2 of 7.0. Figure 4.9 shows the gain modulation and correlated noise. Attentional gain in the fovea is 58% of that in the periphery. The amount of peripheral gain above 1.0 is due to correlated noise, implying that 45% of the total noise variance is correlated. The d' map of this heuristic searcher is shown in Figure 4.2d, and its thresholds were plotted as the blue bars in Figure 4.4.

When we forced the heuristic d' map to be constant over potential target locations, the maximum likelihood fit is a bit worse than the aforementioned heuristic searcher with an assumed d' map that decreases along eccentricity, but still much better than the Bayes-optimal searcher (Figure 4.4, quantitative comparison in the caption). This searcher with a flat heuristic d' map has a foveal gain of 0.761, a peripheral gain of 1.342, and an assumed d' map of 3.1. The estimated effect sizes of foveal neglect and correlated noise are almost the same.

We also measured human search performance in a 7-location search task, where the target could be present in the central seven location locations, while the back-

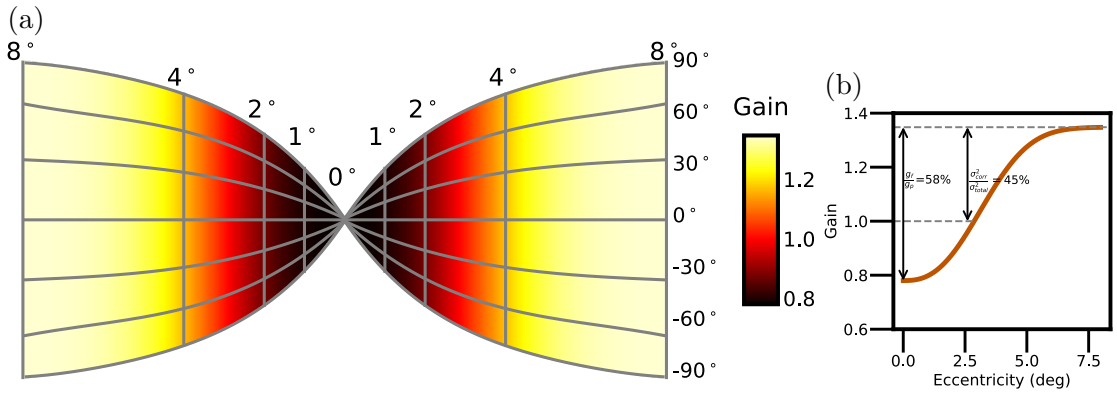


Figure 4.9: Foveal neglect and correlated noise. (a) Retinotopic gain map. The flattened V1 sheet has a constant density of neurons. The grid of contours shows the retinal locations of the cortical neurons' receptive fields. (b) Attentional gain along the horizontal meridian.

ground patches still appeared in all 19 locations. Figure 4.10 shows this task cannot distinguish between the heuristic searchers with a flat or fall-off assumed d' map, though both of them fit human behavior much better than the Bayes-optimal searcher. The fall-off heuristic has a foveal gain of 0.966, a peripheral gain of 1.259, an assumed d'_{max} of 4.6, and an assumed e_2 of 15.8. The flat heuristic has a foveal gain of 0.860, a peripheral gain of 1.325, and an assumed d' map of 3.7. The estimated levels of foveal neglect and correlated noise are slightly lower in this task compared to the 19-location search task.

As can be seen in Figure 4.11, all individual observers in both the 19-location and 7-location tasks had a certain and often similar level of foveal neglect and correlated noise.

In summary, both flat and fall-off heuristic searchers with the combination of very simple heuristic d' map, foveal neglect and correlated noise, provide a plausible explanation of the seemingly paradoxical detection and search results in human

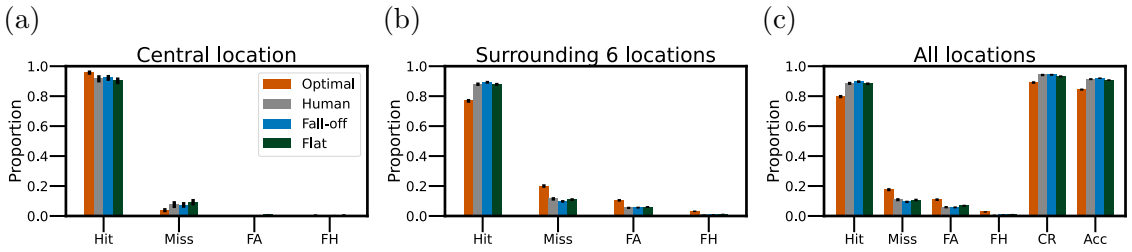


Figure 4.10: Correct responses and errors in the 7-location search task by retinal eccentricity. (a) Histogram of hits, misses, false alarms (FA) and false hits (FH) in the central location, for the average observer (gray), the Bayes-optimal searcher given the d' map of the average observer in detection (orange), the best-fit heuristic observer given an assumed d' map that falls off, the average human d' map in the detection task, correlated noise and foveal neglect (blue), and the best-fit heuristic observer given a flat assumed d' map, the average human d' map in the detection task, correlated noise and foveal neglect (dark green). (b) Histogram of the surround six locations. (c) Histogram for all locations. The correct rejection rate and overall accuracy are also included. (Number of trials N=6800. Error bars are bootstrapped 95% confidence intervals. Fall-off heuristic: log-likelihood = -8228, AIC = 16468, BIC = 16509. The flat heuristic is comparable: log-likelihood = -8230, AIC = 16470, BIC = 16504. The Bayes-optimal searcher is the worst: log-likelihood = -8405, AIC = BIC = 16810. The fall-off model is $e^{1.0}$ times as probable as the flat heuristic model and e^{171} times as probable as the Bayes-optimal searcher.)

performance, that human observers had better overall search accuracy than the Bayes-optimal searcher.

How generalizable are these results with varying numbers of potential target locations? We measured human search performance when the number of target locations is 7, 61, or 91. I have mentioned the configuration of the 7-location search task earlier. For the 61-location configuration, the background had a diameter of 1.9 visual degrees, and the potential locations are still at the center of background patches, separated by 2.2 visual degrees. For the 91-location configuration, the background had a diameter of 1.6 visual degrees, and the potential locations are still at the center of background patches, separated by 1.8 visual degrees. In both configurations, all

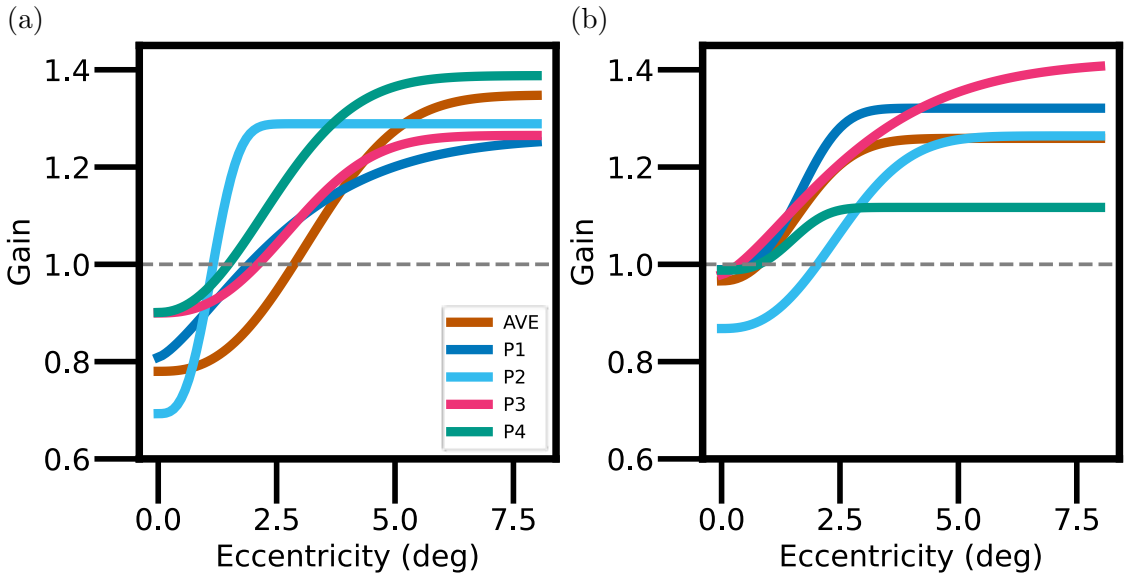


Figure 4.11: Attentional sensitivity gain in individual human observers. (a) Estimated gain in the 19-location search task, for the average human observer and the four individual observers, in the same order as the four rows in Figure 4.3. (b) Estimated gain in the 7-location search task.

potential target locations still cover the central 16 visual degrees of the visual field. We also ran preliminary search trials here with highly visible targets to make sure that the human observers made no errors in clicking on the target location, so that their search performance was not limited by spatial memory and motor control.

Comparison of the human and model d' maps for search in varying numbers of locations is shown in Figure 4.12. Similar to the case when the number of target location is 19 (Figures 4.2b and 4.2c), near-foveal region has much more similar d' values compared to that from the predictions of the Bayes-optimal searcher; that implies some level of the foveal neglect effect.

Table 4.1 summarizes the proportion correct for the average human observer, four individual observers in all detection and search tasks, and the proportion cor-

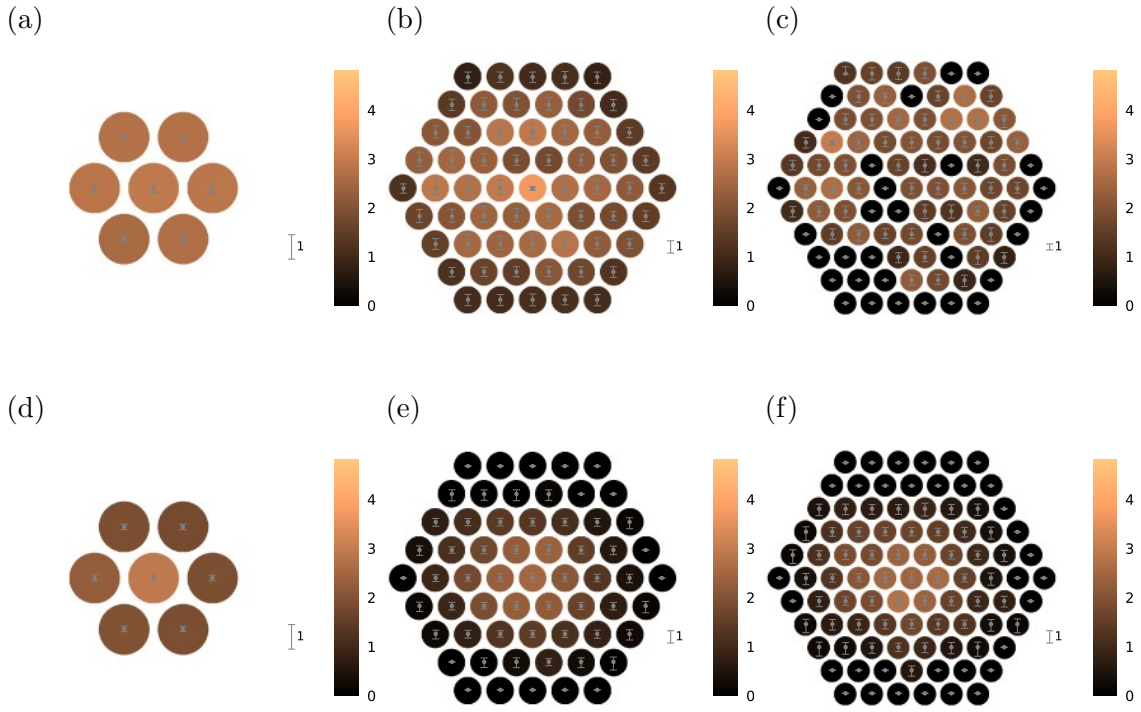


Figure 4.12: Search d' maps for varying numbers of search locations. (a) Search d' map of the average human observer when the target was known to appear only at one of the central 7 locations in half of the trials, while the background patch still appeared at all 19 locations. (b) Search d' map of the average human observer when the target appeared in one of the 61 locations in half of the trials. (c) Search d' map of the average human observer when the target appeared in one of the 91 locations in half of the trials. (d) Search performance of the Bayes-optimal searcher in the 7-location search task, given the central seven d' values from the 19-location d' map in Figure 4.2a. (e) Search performance of the Bayes-optimal searcher in the 61-location search task, with a d' map interpolated and extrapolated from the 19-location d' map in Figure 4.2a. (f) Search performance of the Bayes-optimal searcher in the 91-location search task, with a d' map interpolated and extrapolated from the 19-location d' map in Figure 4.2a. Error bars are bootstrapped 95% confidence intervals.

rect of their corresponding Bayes-optimal searchers. The human observers have a significantly higher overall accuracy than the corresponding Bayes-optimal searcher in every single search configuration. The HVS indeed performs supraoptimal (covert)

visual search.

Task	No. of locations	Human/ Optimal Searcher	O1	O2	O3	O4	Average
Detection	19	Human	0.8167	0.8596	0.8522	0.7862	0.8325
Covert Search			0.6762 ***	0.7069 ***	0.7225 ***	0.6958 ***	0.6978 ***
		Optimal	0.6391	0.6841	0.6869	0.6237	0.6430
	7	Human	0.8979 ***	0.9156 ***	0.9512 ***	0.8917 ***	0.9135 ***
		Optimal	0.8193	0.8561	0.8871	0.8417	0.8438
	61	Human	0.7025 ***	0.8319 ***	0.8388 ***	/	0.791 ***
		Optimal	0.5883	0.6316	0.6426	/	0.5978
	91	Human	0.7331 ***	/	/	/	0.7331 ***
		Optimal	0.5873	/	/	/	0.5873

Table 4.1: Proportion correct in all experiments for the four human observers and the average human observer. The overall accuracy of the Bayes-optimal searcher was computed given the corresponding individual or combined d' map, and interpolated and extrapolated when necessary. Asterisks indicate the p-values of the human accuracy to the accuracy distribution of the corresponding Bayes-optimal searcher, that *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.00001.

Now we consider a family of models with heuristic prior map only (no heuristic deviation of the d' map), that is

$$\hat{x} = \underset{x \in \mathbb{X}}{arg\,max} \left[\ln \hat{p}_x + d'_x(R'_x - d'_x/2)\right] \tag{4.25}$$

The target-absent prior p_0 is an essential element of the overall prior map to consider. We parameterized the rest of the prior map (when the target was present), as a function of the retinal eccentricity

$$p(e) \propto (1 - p_0)\frac{e_p}{e + e_p} \tag{4.26}$$

where e_p is the fall-off parameter.

This is a more arbitrary choice compared to the parameterization of the d' map, as no "foveation" is commonly expected in the prior map based on the scene context. Nevertheless, we used this family of prior maps to demonstrate the effect of heuristic priors on covert search performance.

Consider the prior space of $p_0 = 0.5$ or 0.0, e_p ranging from 0.2 to ∞, as shown in Figure 4.13a. No matter if the target was present in half of the trials or always present, assuming p_0 correctly and e_p incorrectly is sufficient to achieve near-optimal search performance, within a difference of 2% to the Bayes-optimal searcher (Figures 4.13b and 4.13c). As expected, the largest performance lag occurs when the assumed e_p is most different from the actual e_p.

As the actual e_p increases, the overall search accuracy decreases rapidly from 85% to 60%, because the target appears more commonly in the periphery where it is less detectable than in the fovea (Figure 4.13d). Nevertheless, assuming a single, fixed, flat target-present prior map, with the correct target-absent prior, is near-optimal across all search conditions.

Search performance is fairly sensitive to the heuristics of target-absent prior. As shown in Figures 4.13e and 4.13f, when the actual target-absent prior p_0 is 0.5, overall search accuracy is near-optimal when the estimated p_0 is around 0.4 to 0.7. When the actual target-absent prior p_0 is 0.0, overall search accuracy is near-optimal when the estimated p_0 is less than around 0.1.

Nevertheless, ignoring locations where targets sometimes appear is detrimental to overall search accuracy (Figure 4.14). We simulated this situation by assuming zero priors progressively by "ring". We refer to rings here as hexagon-edged circles

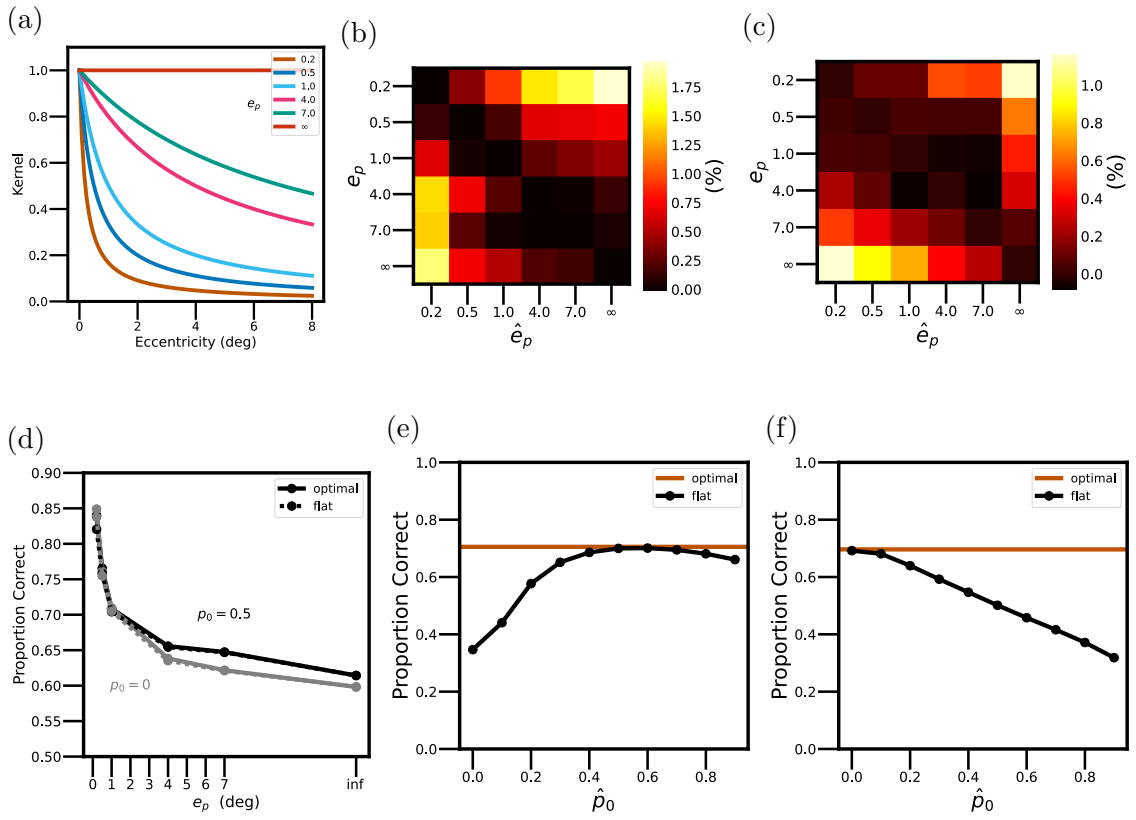


Figure 4.13: Effect of heuristic priors on covert search performance. (a) The actual target-present prior probability kernel with six different values of the fall-off parameter e_p. (b) The performance lag, difference in the overall search accuracy between the Bayes-optimal searcher and the heuristic searchers with a certain assumed e_p and a perfect estimation of p_0, when $p_0 = 0.5$. (c) The performance lag when $p_0 = 0.0$. (d) The overall search accuracy of the Bayes-optimal searcher and a single heuristic searcher with fixed, flat target-present prior map and a perfect estimation of p_0, when p_0 are 0.5 and 0.0. Those differences in performance correspond to the last columns in the heatmaps (b) and (c). (e) The overall accuracy averaged over all six conditions of the Bayes-optimal (orange) and heuristic searchers (black) that assume various p_0 values and a flat prior over all target locations, when $p_0 = 0.5$. (f) The overall accuracy averaged over all six conditions of the Bayes-optimal (orange) and heuristic searchers (black) that assume various p_0 values and a flat prior over all target locations, when $p_0 = 0.0$.

counted outwards from the central location, where the 19-location configuration has three rings, the 61-location configuration has five rings, and the 91-location configuration has six rings. Ignoring near-foveal rings hammers proportion correct more than ignoring peripheral rings, despite considerably more target locations existing in peripheral rings. This results from the fact that the target were more detectable near the fovea than in the periphery.

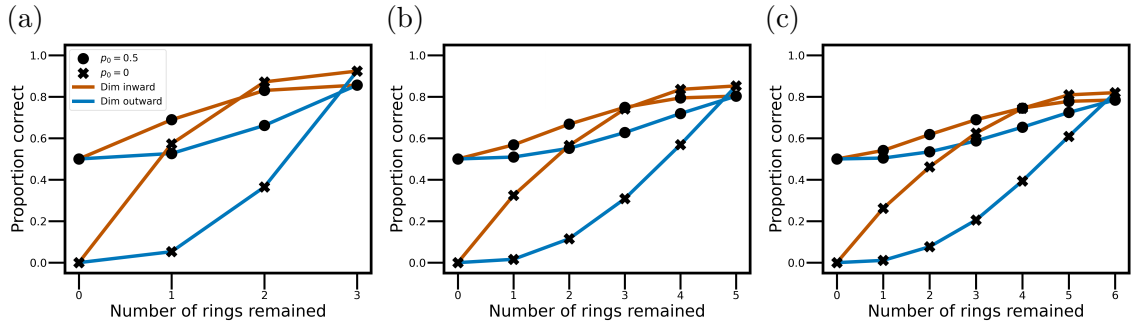


Figure 4.14: Effect of highly heuristic priors on covert search performance. (a) The overall proportion correct in the 19-location search task of the Bayes-optimal searcher and heuristic searchers assuming the target cannot be present at certain target locations (assume local priors of 0), when $p_0 = 0.5$ and 0.0. "Dim inward" means the rings farthest from the center are assumed priors of 0 first. "Dim outward" means the rings closest to the center are assumed priors of 0 first. When no ring is assumed zero prior, the searcher is optimal. (b) The overall proportion correct in the 61-location search task of the Bayes-optimal searcher and heuristic searchers. (c) The overall proportion correct in the 91-location search task of the Bayes-optimal searcher and heuristic searchers. For all searchers, the d' map had a d'_{max} of 7 and an e_2 of 6, and no heuristic d' values were used.

How necessary is an exact normalization of the receptive field response for heuristic searchers to reach near-optimal performance? All the optimal and heuristic decision rules (Equations 4.19, 4.21, 4.25) mentioned so far use the response R' exactly normalized by the standard deviation of the response, which is the standard deviation of the white noise for a simple template matching observer. Assume the mean response

is 1.0 when the target is present, and 0.0 when the target is absent (e.g., a target with an energy of 1 for a simple template matching model). The unnormalized response $R = \sigma R' = R'/d'$, where $d' = 1/\hat{\sigma}$, or $R' = d'R$. Therefore, we consider this family of models with not only heuristic d' map but also heuristic normalization, that is

$$\hat{x} = \underset{x \in \mathbb{X}}{arg\,max} \left[\ln p_x + \hat{d}'^2_x (R_x - 1/2) \right] \tag{4.27}$$

where R_x is the unnormalized receptive field response.

As shown in Figures 4.15a and 4.15b, we considered the same 25 actual d' map conditions in Figures 4.5a and 4.5b, but with two different heuristic searchers that maximize overall search accuracy across all conditions when normalization is applied heuristically. When the target is present in half of the trials, heuristic deviation from the actual d'_{max} and e_2 brings overall accuracy noticeably below the Bayes-optimal searcher (Figures 4.15d and 4.15e). Accuracy falls significantly when those two parameters are overestimated. Those effects are visible, but less in size when the target is always present. In the case where d' map varies randomly per trial, heuristic normalization hammers down search accuracy in most cases (Figures 4.15c and 4.15f). To summarize, exact normalization is necessary for heuristic searchers to feasibly achieve near-optimal performance.

4.4 Discussion

Cued detection and covert search performance were measured for a wavelet target in Gaussian white noise under carefully controlled conditions. The detectability map measured in the cued detection task was used to predict covert search performance of the Bayes-optimal searcher, assuming statistically independent sensory

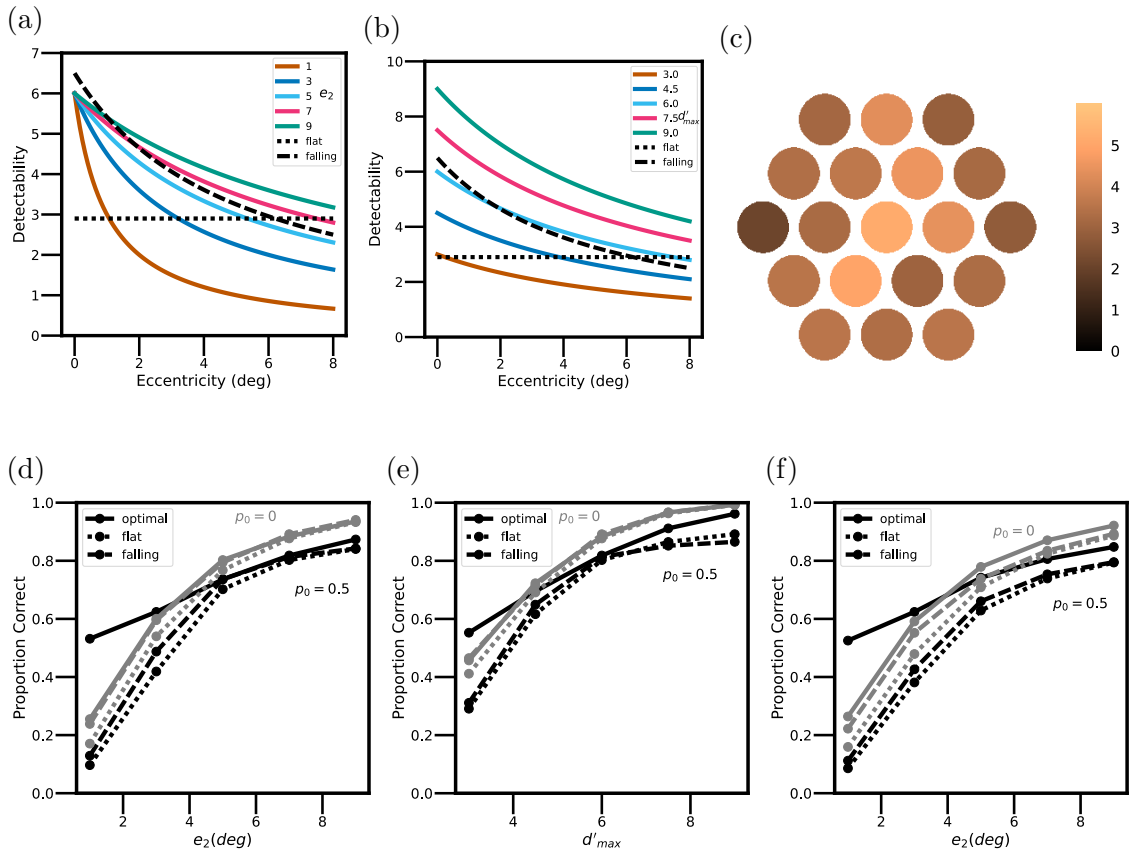


Figure 4.15: Effect of heuristic normalization on the covert search performance of searchers with heuristic d' maps. (a) Actual d' maps with a d'_{max} of 6.0 and a range of e_2, in colored curves. The best-fit (across all 25 conditions) heuristic with a flat d' map has a d'_{max} of 2.9 (and $e_2 = \infty$), in the dotted line. The best-fit (across all 25 conditions) heuristic has a d'_{max} of 6.5 and an e_2 of 7.0., in the dashed curve. Both heuristics normalize response based on Equation 4.27. (b) Overall search accuracy for Bayes-optimal and heuristic searchers in (a), with the target absent rate p_0 of 0.5 and 0.0. (c) Actual d' maps with an e_2 of 7.0 and a range of d'_{max}, in colored curves. The two heuristic searchers are the same as those in (a). (d) Overall search accuracy for Bayes-optimal and heuristic searchers in (c), with the target absent rate p_0 of 0.5 and 0.0. (e) An example of the d' map that varies randomly per trial. The baseline d' map has a d'_{max} of 6.0 and an e_2 of 7.0. (f) Overall search accuracy for Bayes-optimal and heuristic searchers for the baseline d' maps with a d'_{max} of 6.0 and e_2 ranging from 1 to 9 visual degrees.

responses from the potential target locations. We found that human performance slightly exceeded the predictions of the Bayes optimal decision rule, despite the computational complexity of the optimal rule, and despite the fact that humans showed a loss of sensitivity in the fovea. We found these seemingly paradoxical results can be explained quantitatively by these three factors: (1) Extremely simple heuristic decision rules, together with the local normalization, can achieve near-optimal performance. (2) Correlated neural noise causes the d' values measured in the detection task to underestimate the effective d' values in the search task. (3) Foveal neglect only reduces noticeably the d' value at the central target location.

We chose a 6-cpd wavelet target because its detectability varies substantially over the search region with a diameter of 16 visual degrees. It would be informative to repeat the measurements for other targets, Given the fact that simple heuristic decision rules are near-optimal for a wide range of d' maps, it is highly likely that our findings will generalize well across a wide range of spatially-localized targets.

We chose white noise background because they have a dense texture similar to natural images yet are statistically simple, and widely used in studies of visual behavior. Natural backgrounds, on the other hand, are statistically complex and non-stationary, so that their masking typically vary across potential target locations. Thus, the d' map varies on each trial with both foveation and variation in the masking properties of the background. We have shown that simple fixed heuristic rules are effectively optimal even when the d' map changes randomly on each trial (Figures 4.5f and 4.7), which suggests that our findings may remain valid over a wide range of stationary and non-stationary backgrounds. A difficulty in directly testing this hypothesis is to measure the d' map for the stimulus at each location for each trial for the calculation of the optimal performance. It is tractable to estimate the d' map for

white noise background with spatially varying luminance and contrast [17, 98, 138]. For natural backgrounds, this estimation is much harder, but some progress has been made [17, 25, 79, 80, 128]. We are optimistic that a model with a simple heuristic decision rule, correlated noise, and foveal neglect will be able to predict human covert search performance in natural backgrounds.

We acknowledge that our theoretical analysis and computational modeling adopt the standard SDT assumption–the observer's decision variable is normally distributed. This assumption is reasonably justified in white noise background. Psychophysical literature has shown linear receptive field responses to natural backgrounds are approximately normally distributed if the responses are normalized by background properties (e.g., luminance and contrast) [16, 17].

Under natural search conditions, the number of potential target locations often varies from one situation to the next, and hence varying the number of potential locations is a key experimental manipulation. When there is just a single potential target location, the search task is reduced to a very simple identification, discrimination, or detection task. In general, as the number of potential target locations increases, search accuracy and speed decrease. The major scientific questions are what stimulus and neural factors are responsible for the decreases, and whether models that incorporate the relevant neural factors can quantitatively predict search performance.

A caveat to results in the 61- and 91-location search tasks is though those results are quite consistent with the 7- and 19-location search tasks, they were measured from fewer trials, and the d' map for the Bayes-optimal searcher was partially extrapolated. Further experiments and analyses are desired for sounder generalization to our conclusions in the varying numbers of target locations.

Normalization is a fundamental property of cortical processing [12–14]. The

present result suggests it may play a more important role in perceptual decision-making than previously appreciated. For detection tasks in natural backgrounds, it has been shown that normalization by local luminance, contrast, and similarity allows near-optimal decision-making with a single fixed decision criterion [17]. The present results expand this conclusion by showing that such normalization also allows extremely simple heuristic decision rules to achieve near-optimal performance for a wide range of natural identification tasks. Without normalization, the heuristics described in this chapter will not perform nearly as well on natural and other non-stationary backgrounds (Figure 4.15f).

The simplicity of near-optimal search rules allows a feasible neural implementation and a sizable space to tolerate individual differences in search strategy.

Where humans tend to fall below the predictions of the optimal decision rule is when the task is to identify the locations of multiple targets [139] or multiple categories of targets, or to identify targets with demands on memory or high-level cognitive computation (e.g., which location contains a number divisible by 13). Theories of covert search in such conditions need to take the memory capacity and/or cognitive load into consideration. Nonetheless, in many real-world situations, observers are covertly searching for targets that require low cognitive effort.

Our results do not prove that correlated noise is the source of the supraoptimal accuracy of the observers in the experiments, but it is a plausible hypothesis, consistent with the evidence for slow modulations in membrane potential and blood oxygen level-dependent (BOLD) signal [140–144]. We showed correlated noise can create a mismatch between the d' values estimated in the cued detection task and the effective d' values in the covert search task. Such mismatches might occur in other identification tasks, which could become powerful tools to investigate the characteristics of

correlated noise.

An interesting possibility is that the nervous system injects correlated variations into the pathways transmitting information to the brain areas that perform identification tasks. Because these correlated variations do not hurt identification performance, they could provide an independent channel for communicating other kinds of information, including reward signals, arousal signals, and global context information. The benefits of this low bit rate communication channel may outweigh the cost of reduced sensitivity in simple yes/no tasks.

Overt search can be roughly characterized as fixation periods separated by saccades. During each fixation, the stimulus information is gathered and processed to reduce uncertainty in the actual target location, and the next fixation location is selected after computation. The optimal decision rule for picking the next fixation location also takes into account the d' and prior maps [77]. An important next step is to perform a Bayesian heuristic decision analysis for fixation selection.

Our findings with regard to heuristic performance are likely to generalize to many other identification tasks, as most identification tasks can be described as making choices between mutually exclusive events. Bayesian heuristic decision analysis described in this chapter may provide useful insights and testable predictions beyond detection and search tasks.

Chapter 5: Heuristic Analysis

Abstract

We conducted a systematic analysis of Bayes heuristic covert searchers and presented the distribution of their performance lag to the Bayes-optimal searchers. Increasing the number of parameters that are used heuristically in the decision process generally increases the performance lag, but near-optimal search performance can still be achieved even when all parameters are highly heuristic, which indicates interaction among heuristic components can cancel out their individual effects on search performance. Heuristic normalization, the absence of log-likelihood centering, and heuristic max rule most likely decrease overall search performance considerably. Though enormous heuristics can reach the same level of overall accuracy in covert search, I demonstrate the possibility of distinguishing different heuristic compositions by comparing the pattern in the location-dependent statistics.

5.1 Introduction

Ideal observer has made valuable contributions to numerous areas of visual perception, such as pattern detection, discrimination, estimation, visual attention, perceptual grouping, shape, depth, and motion perception [29]. As the ideal observer continues to serve as a powerful tool for explaining and predicting visual behavior, one might ask if its computation in most real-world tasks is overly laborious for the human visual system (HVS) to implement. Furthermore, even if the HVS is able to

implement the optimal computation for best performance, is the effort worthwhile that the benefits overweight the costs? What if the task is too volatile, uncertain, complex, and ambiguous to derive the optimal computation?

For those reasons, heuristic decision-making emerges from a multitude of research fields and competes for attention with normative decision-making. Simon [145] coined the term bounded rationality to describe the fact that human rationality is limited and individuals make decisions heuristically instead of acting as mathematically perfect agents. Gigerenzer and Gaissmaier [146] describe heuristics as "strategies that ignore information to make decisions faster, more frugally, and/or more accurately than more complex methods". In other words, heuristic decision-making is the efficient cognitive process that uses information suboptimally and sometimes still achieve near-optimal or even supraoptimal performance in tasks.

I describe a model as "normative" when it is where heuristics derive from. A normative model is typically optimal or expert (high-performance), with decision parameters matching the sensory parameters. In light of human visual detection and search, we ask questions below on heuristic decision-making:

- When do heuristic rules cause performance lag to normative searchers? When do heuristics become detrimental to performance?
- How normatively or heuristically does the human visual system search in a certain search condition? How could one measure what heuristics, if any, are being used? How could different heuristics be distinguished from each other?

In Chapter 4, we discovered the supraoptimal search performance of human observers and employed the near-optimality of extremely simple heuristic rules to

explain the results. We noticed the heuristics with fall-off and flat d' maps fit almost equally well to experimental data, which means though the decision processes in our detection and search tasks were well bridged, those tasks were not useful to distinguish different possible heuristics the human observers were using. Therefore, in this chapter, we systematically simulated the effects of various heuristic rules on visual search behavior, so those predictions can be used reversely to generate meaningful hypotheses and further experimental testing.

A highly abstract representation of a decision-making process is (w, x, s, f, y), where w is the state of the world or environment, x is the relevant input information, s is the parameters and hyperparameters of the system, f is the computational processing from information to decision when the system is at a certain state, and $y = f(x; s)$ is the decision made. Heuristics can occur in parameters, hyperparameters, and computation of the system, with their effectiveness depending on the (often statistical) state of the world, the availability and uncertainty of relevant input information, and the utility landscape of the decision space.

The maximum a posteriori covert search rule in Equation 4.19 has the receptive field response as the input, the actual d' map as the observation state, the max rule, the log, normalization and summation operations as the computation, the response of target location as the final decision, and the configuration of locations, the target, backgrounds as the state of the world.

We have shown in Figures 4.5, 4.6 and 4.7 that simple fixed heuristic d' maps allows near-optimal performance in the 19-location covert search task, even if the actual d' map varies randomly every trial. We also showed in Figures 4.13 and 4.14 that simplistic heuristic target-present prior maps are sufficient for near-optimal search performance, as long as no target location is ignored, while the target-absent prior

needs to be assumed more accurately. Then Figure 4.15 shows, exact normalization is a critical computation for heuristic searchers to stay close to best performance.

In this chapter, we conducted a systematic analysis of Bayes heuristic covert searchers and presented the distribution of their performance lag to the Bayes-optimal searchers. Increasing the number of parameters that are used heuristically in the decision process generally increases the performance lag, but near-optimal search performance can still be achieved even when all parameters are highly heuristic, which indicates interaction among heuristic components can cancel out their individual effects on search performance. Heuristic normalization, the absence of log-likelihood centering, and heuristic max rule most likely decrease overall search performance considerably. Though enormous heuristics can reach the same level of overall accuracy in covert search, I demonstrate the possibility of distinguishing different heuristic compositions by comparing the pattern in the location-dependent statistics.

5.2 Preliminary exploration of covert search heuristics

To conduct a systematic investigation of the high-dimensional perception space in our covert search task, it is valuable to first itemize the relevant variables. We keep the same hexagonal structure of target locations (e.g., Figure 4.12) and the Gaussian-distributed receptive field response. The radius of the display (from the central location to the horizontal edge location) is kept as 8 visual degrees, though this value can be perturbed in further analysis. Rings can be counted outwards from the central location, with the number of locations at the $k-th$ ring as $n_k = 1, k = 1$ and $n_k = 6(k-1), k > 1$. The number of total locations with k rings is $N_k = 3k^2 - 3k + 1$. For the actual prior map, we have the target-absent prior p_0 and assume the target-present prior map follows

$$p(e) \propto \frac{1 - p_0}{k_p e + 1} \tag{5.1}$$

where k_p is the slope parameter, equivalent to the inverse of the fall-off parameter e_p in Equation 4.26. An k_p of 0 indicates the prior map is flat. The prior map is always normalized to have a sum of $1 - p_0$, so the total prior is 1. For the actual d' map, we have a peak parameter d'_0 (as d'_{max} in the last chapter) and a slope parameter k_d, with

$$d'(e) = \frac{d'_0}{k_d e + 1} \tag{5.2}$$

k_d is equivalent to the inverse of the fall-off parameter e_2 in Equation 4.20. An k_d of 0 indicates the d' map is flat. Lastly, we have the heuristic parameters of the prior and d': $\hat{p}_0$, $\hat{k}_p$, $\hat{d}'_0$, and $\hat{k}_d$.

We prioritize decision-making processes and heuristics that are practical and common in the real world, as it is impossible to explore all heuristics for any problem. Table 5.1 provides an overview of our simulation of more than 20,000 combinations of variables. Rings of 3, 5, and 7 corresponds to 19, 61, and 127 target locations, and low, medium, and high location densities within the visual field. Target-absent priors of 0, 0.5, 0.9 corresponds to the search cases where the target is always present, sometimes present, and rarely present. k_p with values from 0 to 1 covers the cases from when the target is equally likely to show up at every target location, to when the target is much more likely to show up near the fovea. The values of foveal d' from 4 to 10 indicate the target is at least fairly detectable to the observer when being directly focused on; otherwise, the observer may move eyes, head and/or body to have a reasonable chance of locating the target.

Because the data set is large, we employed a machine learning approach for preliminary analysis. We calculated the overall accuracy, hit rate, correct rejection rate, miss rate, false alarm rate, false hit rate, and the performance lag. Then we classified the performance lag with thresholds of 1% and 10%. For each configuration of target locations, our feature variables include p_0, k_p, d'_0, k_d and the differences between them and their heuristics parameters (for better interpretation), and the lag level as the target variable.

Variable Name	Expression	Data Type	Value	Description
n_ring	k	int	3,5,7	number of rings in the potential target locations
p0	p_0	float	0, 0.5, 0.9	target-absent prior (proportion in the trials)
kp	k_p	float	0, 0.1, 0.2, 1	target-present slope parameter
d0	d'_0	float	4,7,10	d' map peak parameter
kd	k_d	float	0, 0.1, 0.2, 1	d' map slope parameter
e_p0	$\hat{p}_0$	float	0 – 0.9	heuristic target-absent prior
e_kp	$\hat{k}_p$	float	0 – 1	heuristic target-present slope parameter
e_d0	$\hat{d}'_0$	float	1 – 13	heuristic d' map peak parameter
e_kd	$\hat{k}_d$	float	0 – 1	heuristic d' map slope parameter
acc		float	0 – 1	overall search accuracy
hit		float	0 – 1	overall hit rate
cr		float	0.7 – 1	overall correct rejection rate

miss		float	0 – 1	overall miss rate
fa		float	0 – 0.3	overall false alarm rate
fh		float	0 – 1	overall false hit rate
d_p0	$\hat{p}_0 - p_0$	float	-0.4 – 0.4	difference in the actual and heuristic target-absent prior
d_kp	$\hat{k}_p - k_p$	float	-1 – 1	difference in the actual and heuristic target-present slope parameter
d_d0	$\hat{d}'_0 - d'_0$	float	-3 – 3	difference in the actual and heuristic d' map peak parameter
d_kd	$\hat{k}_d - k_d$	float	-1 – 1	difference in the actual and heuristic d' map slope parameter
lag		float	0 – 1	difference in the actual and heuristic overall search accuracy
lag_level		category	0,1,2	levels of performance lag. 0 means lag is less than 1%; 1 means lag is between 1–10%; 2 means lag is greater than 10%.

Table 5.1: Scanning space of covert search heuristics.

In the 19-location search task, a simple decision tree model that maximizes Gini gain in the training set is able to achieve a classification accuracy of 93% in the testing set. The two sets were split with a ratio of 75%/25%. Trials with levels 0–2 of performance lag were fairly balanced, with a rough ratio of 2:2:1. For each level of performance lag, the precision, recall, and f1-score range from 91% to 97%. The high accuracy of this decision tree in classification supports its value to predict when heuristics function well and when they fail. For example, Figure 5.1a shows the shapes of the actual and assumed prior maps are not important for performance lag. Also, the difference between the actual and heuristic values of parameters in prior and d' maps is slightly more useful to predict performance lag than the actual values.

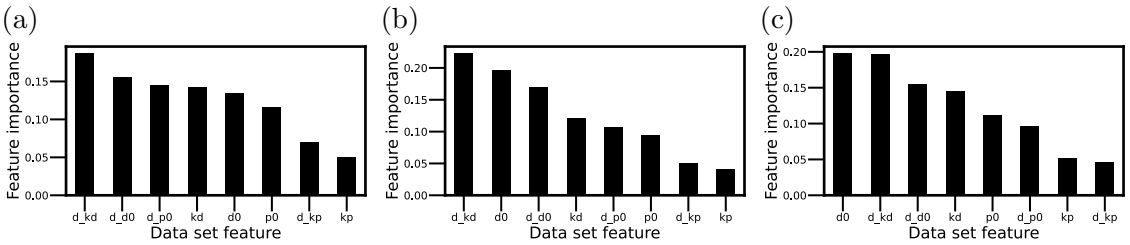


Figure 5.1: Feature importance score on performance lag in (a) 19-location (b) 61-location (c) 127-location covert search. The score was calculated based on the amount of Gini gain at all branches of the decision that use the feature.

The shapes of the actual and assumed prior maps stay unimportant for performance lag as the number of target location increases (Figures 5.1b and 5.1c). At the same time, the peak parameter d'_0 becomes more relevant for predicting performance lag.

The top layers of the best decision tree in the 19-location search are shown in 5.2. The top layers happen to be the same after fitting data in the 61- and 127-location search, so they are not repeatedly plotted. Based on the data set and the decision tree, inaccurate (especially overestimated) heuristics of the target-absent prior could be detrimental to search performance, which is consistent with the results in Figures 4.13e and 4.13f. Furthermore, if the actual d' map has high values throughout the visual field (high d'_0 and low k_d), then the heuristics are more likely to be near-optimal, as the objective, excellent detectability compensates the subjective, simplistic heuristics.

The data set shows some worst cases of performance lag (around 30%) if only a single parameter is heuristic: (1) $\hat{k}_d$ assumes a rapid fall-off of the d' map when the actual d' map is low-valued and flat; (2) $\hat{d}'_0$ underestimates the overall d' map when the actual d' map is low-valued and flat; (3) $\hat{p}_0$ estimates a fair proportion of

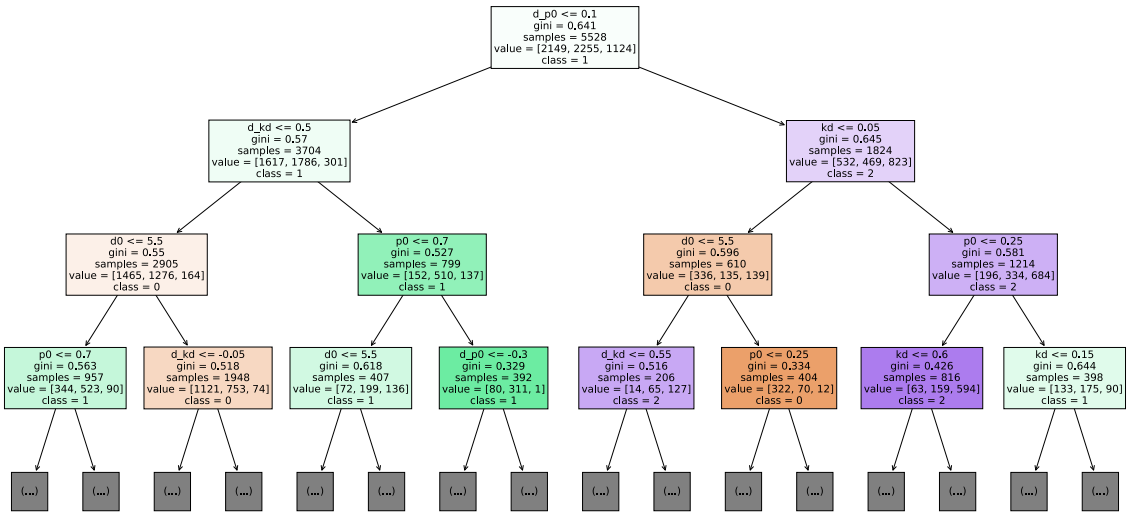


Figure 5.2: The first four layers of a decision tree that well predicts performance lag in the 19-location search task.

target-absence when the target is always present. The co-occurrence of these factors often results in a performance lag above 70%.

On the other hand, zero performance lag can be achieved even if all four parameters are heuristic. Specifically, target-absent prior, d' peak and slope values are severely underestimated, while the target-present prior slope value is arbitrarily heuristic. In fact, though an increase in the number of parameters with values assumed heuristically (shorted as "number of heuristics" for later discussion) increases the median and variance of performance lag, it does not eliminate the cases with zero and near-zero performance lag (Figure 5.3). This interesting result confirms the understanding that it is highly unlikely to find a covert search configuration that allows heuristics to be distinguished by overall search performance alone. Instead, more detailed patterns, such as location-based statistics, are needed for telling apart heuristic components.

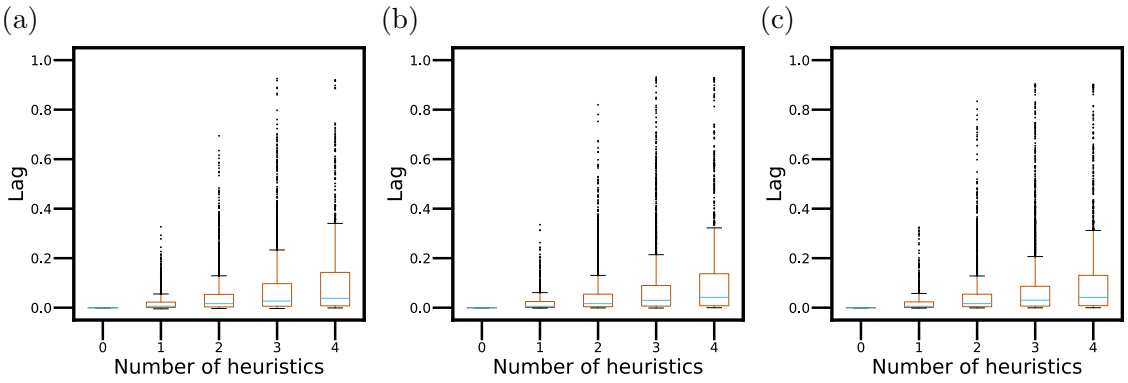


Figure 5.3: Performance lag distribution as a function of the number of heuristics in the (a) 19- (b) 61- (c) 127-location covert search task. The box plot is standard, with the first, second, and third quartiles. Outliers that are beyond 1.5 times of interquartile range from the first and third quartiles are plotted individually.

Consider the effects of overestimation and underestimation of a single parameter. If the target-absent prior is overestimated, then the overall correct rejection and miss rates increase, and the overall hit, false hit, and false alarm rates decrease. If the target-absent prior is underestimated, the effect is reversed for each statistics above. If a target-present prior is overestimated, then the hit, false hit, and false alarm rates at that specific location increases, and the correct rejection and miss rates at that location decrease. If that prior is underestimated instead, the effect is also reversed for statistics at that location.

The effect of bias in a single d' follows a quadratic pattern. When a d' is overestimated and other d' values are estimated ideally, if the target is absent at that location, the mean difference in log-likelihood between the Bayes-optimal and the heuristic searchers is $(-\hat{d}'^2 + d'^2)/2 < 0$, so the heuristic searcher may have a higher correct rejection rate and lower false alarm and false-hit-to rate at that location; if the target is present at that location, the mean difference in log-likelihood between

the Bayes-optimal and the heuristic searchers is $-(\hat{d}' - d')^2/2 < 0$, so the heuristic searcher may have higher miss and false-hit-from rates and a lower hit rate at that location. When a single d' is underestimated, the heuristic searcher may have higher false alarm, miss, false-hit-from and false-hit-to rates, and lower correct rejection and hit rates.

For a covert search task where the target is always present, a single overestimated d' increases the false-hit-from rate and decreases the hit and false-hit-to rate, while a single underestimated d' increases the false-hit-from and false-hit-to rate, and decreases the hit rate.

We predict the effect of a collective bias of d' in the same way. When the actual d' map is flat and all d' values are overestimated by the same amount, the heuristic searcher may have a higher overall correct rejection and miss rates, and lower overall false alarm and hit rates. When the actual d' map is flat and all d' values are underestimated by the same amount, the heuristic searcher may have a higher overall false alarm and miss rates, and lower correct rejection and hit rates. This effect no longer applies when the target is always present.

With the same parametric heuristics, heuristic normalization (Equation 4.27), compared to perfect normalization, typically increases performance lag to the Bayes-optimal searchers, as shown in Figure 5.4, especially when more than two parameters are heuristic. This result is consistent with our observation in Figure 4.15f. Nevertheless, heuristic normalization improves search performance in a minority of cases.

We also consider the heuristic computation where the last term for centering the log-likelihood ratio is dropped, that is

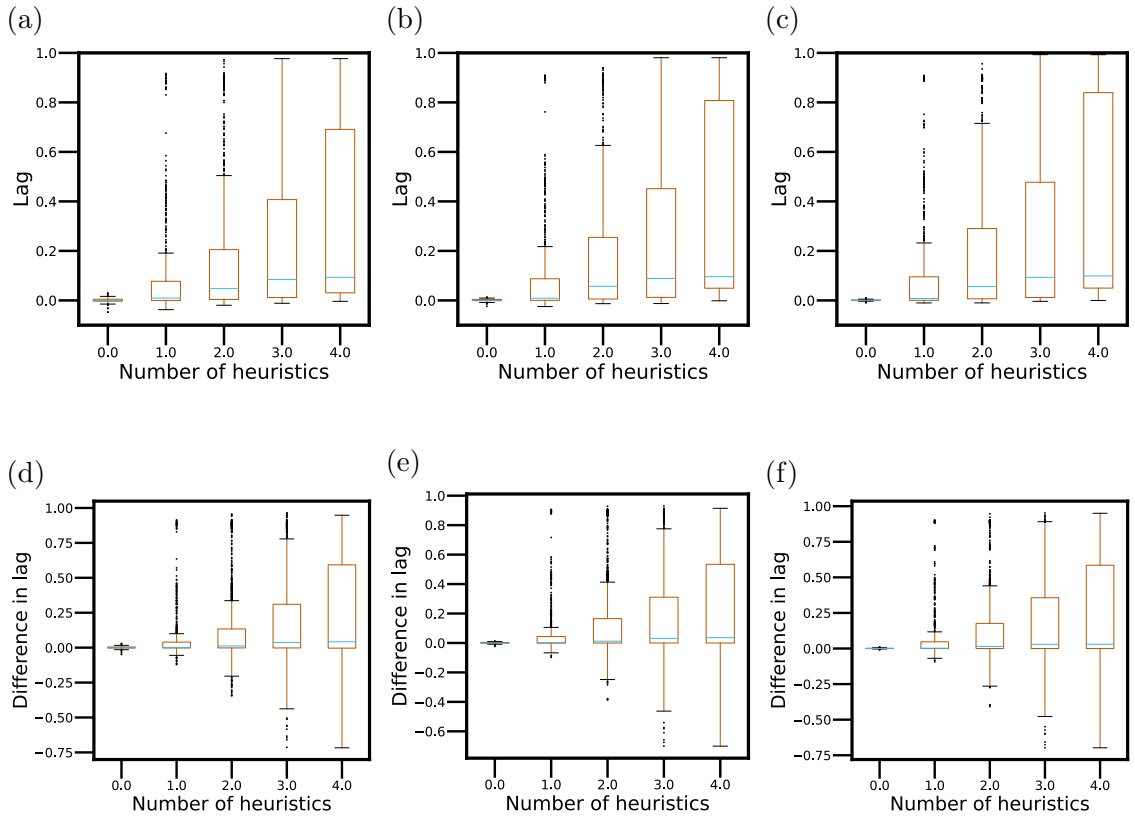


Figure 5.4: Distribution of performance lag and difference in performance lag with heuristic normalization, as a function of the number of heuristics. (a-c) Performance lag is the accuracy difference between the Bayes-optimal and heuristic searchers that normalize heuristically. (d-f) Difference in performance lag is the difference in the performance lags between heuristic searchers with perfect and heuristic normalization. First column: 19-location; second column: 61-location; third column: 127-location.

$$\hat{x} = \underset{x \in \mathbb{X}}{arg\,max} \left(\ln p_x + \hat{d}'_x R'_x \right) \tag{5.3}$$

As shown in Figure 5.5, the centering term $\hat{d}'^2_x/2$ is essential for covert search. Search accuracy typically decreases by 10-50% without centering. Nevertheless, ignoring the centering can improve search performance in a minority of cases.

Next, we simulated the heuristic computation that human covert searchers stop

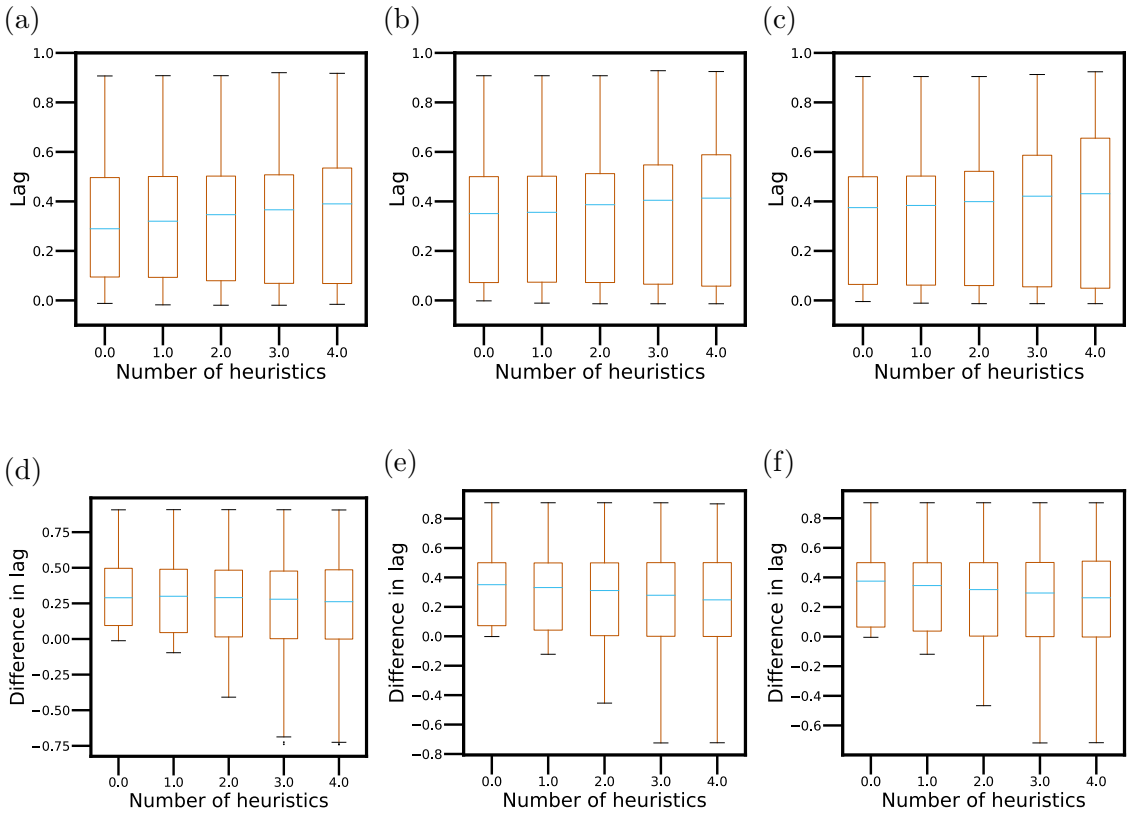


Figure 5.5: Distribution of performance lag and difference in performance lag without centering, as a function of the number of heuristics. (a-c) Performance lag is the accuracy difference between the Bayes-optimal and heuristic searchers that do not center the log-likelihood term. (d-f) Difference in performance lag is the difference in the performance lags between heuristic searchers that center and do not center the log-likelihood ratio. First column: 19-location; second column: 61-location; third column: 127-location.

integrating information in target locations if a large local posterior randomly captures their attention. The heuristic observer first selects the top n largest local responses (including the target-absent response) and then randomly chooses a response among those n cases. Mathematically,

$$\hat{x} = \underset{x \in \mathbb{X}}{arg\,max}(n) \left[\ln p_x + \hat{d}'_x (R'_x - \hat{d}'_x / 2) \right] \tag{5.4}$$

Randomization of the max rule consistently damages covert search performance by 30-50% (Figure 5.6). The random-2-max rule here randomly responds location between the two largest posteriors.

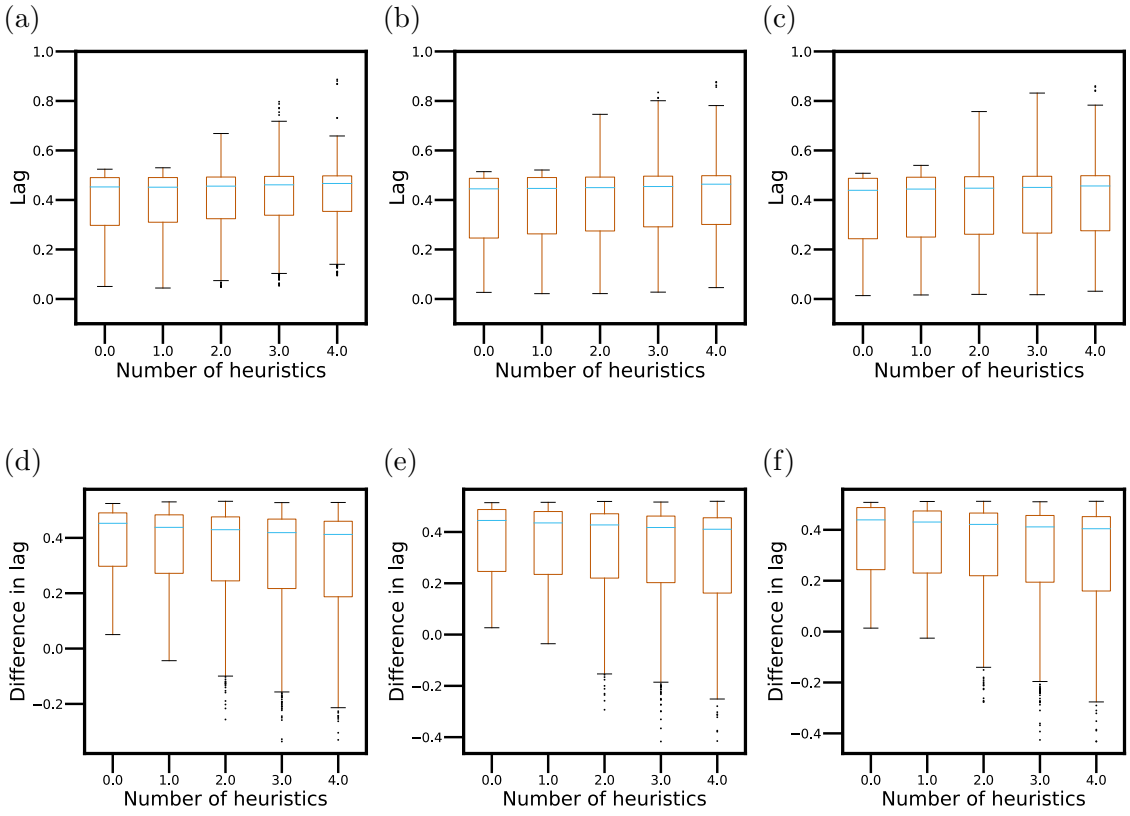


Figure 5.6: Distribution of performance lag and difference in performance lag with the random-2-max rule, as a function of the number of heuristics. (a-c) Performance lag is the accuracy difference between the Bayes-optimal and heuristic searchers that responds randomly among the two largest posteriors. (d-f) Difference in performance lag is the difference in the performance lags between heuristic searchers without and with the random-2-max rule. First column: 19-location; second column: 61-location; third column: 127-location.

Lastly, we varied the d' maps per trial around a baseline d' map. Note that this perturbation is not a heuristic computation, but an extension of the state of the world and relevant input information. The Bayes-optimal searcher used to calculate the performance lag in the random d' condition is able to know the exact d' map on each trial. We found the difference in performance lag typically does not change no matter whether the real-world d' map varies or not (Figure 5.7). Surprisingly, Figure 5.8 shows even if we increased the standard deviation of the random variation to 40% of the base vale (with negative d' set to 0), the difference in performance lag in most cases is still near zero. This means if a heuristic is near-optimal given a stable d' map, it is most likely still near-optimal when the actual d' map varies around with the original d' map as baseline values. For instance, the single fall-off or flat heuristic for the d' map is sufficient to reach near-optimal search accuracy in the randomly varying d' maps (Figure 4.5f).

5.3 Case studies of covert search heuristics

In this section, we focus on comparing the search patterns among heuristic (and Bayes-optimal) searchers that have the same overall accuracy. If a specific configuration of the stimuli results in significant differences in behavioral details, then that configuration can be used as a test to distinguish different heuristic rules.

Figure 5.9 shows such a case. We chose 127 as the number of target locations for more resolution in the pattern of location-dependent statistics. In such an experiment, the method of human response needs to be confirmed not limited by memory and motor precision. As we let $p_0 = 0.5, d'_0 = 4.0$, and both prior and d' maps to be flat, four example searchers with different decision process are able to reach the same level of overall accuracy. Specifically, the high-$\hat{p}_0$ searcher has lower hit, false hit, and

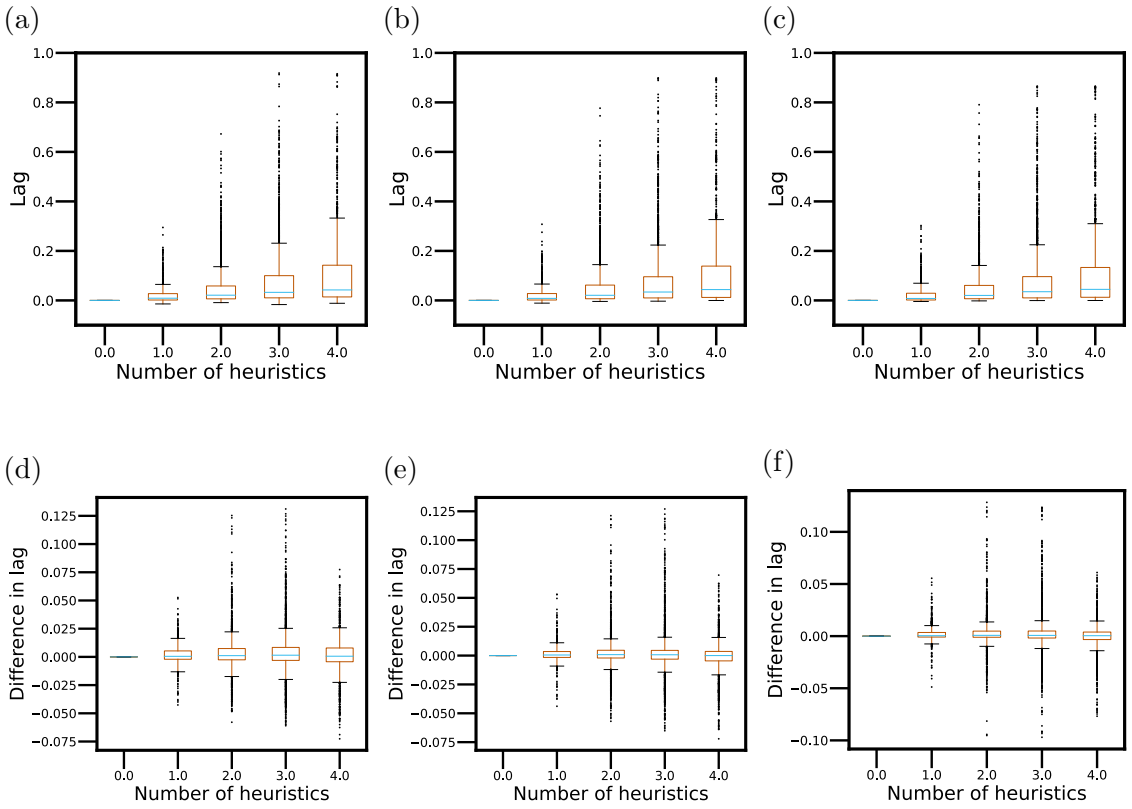


Figure 5.7: Distribution of performance lag and difference in performance lag with randomly varying d' map with a standard deviation of 20% of the base value, as a function of the number of heuristics. (a-c) Performance lag is the accuracy difference between the Bayes-optimal and heuristic searchers. (d-f) Difference in performance lag is the difference in the performance lags between heuristic searchers with and without random variation in d' map. First column: 19-location; second column: 61-location; third column: 127-location.

false alarm rates, and higher correct rejection and miss rates than other searchers. The fall-off and heuristic-normalization searchers have the false hit, false alarm rates higher than those of the Bayes-optimal searcher, and the miss rate higher than that of the Bayes-optimal searcher.

Furthermore, location-dependent statistics as a function of eccentricity are different among the four searchers, as shown in Figure 5.10. The false-hit-from rate

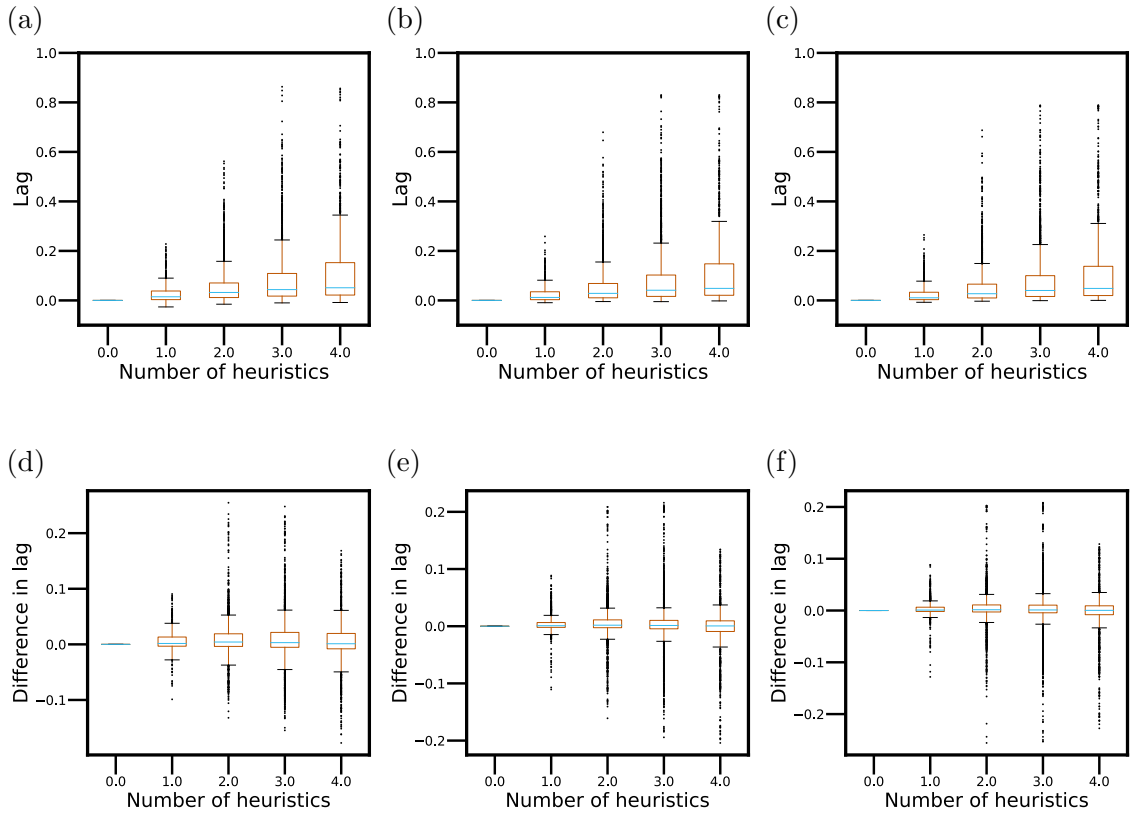


Figure 5.8: Distribution of performance lag and difference in performance lag with randomly varying d' map with a standard deviation of 40% of the base value, as a function of the number of heuristics. (a-c) Performance lag is the accuracy difference between the Bayes-optimal and heuristic searchers. (d-f) Difference in performance lag is the difference in the performance lags between heuristic searchers with and without random variation in d' map. First column: 19-location; second column: 61-location; third column: 127-location.

is the number of false hit trials from locations with a specific eccentricity divided by the number of trials with the target present among those locations. The false-hit-to rate is the number of false hit trials to locations with a specific eccentricity divided by the number of trials with responses among those locations. The fall-off searcher has the hit and false alarm rates increasing rapidly as eccentricity increases, and the false-hit-to and miss rates decreasing rapidly as eccentricity increases, unlike the other

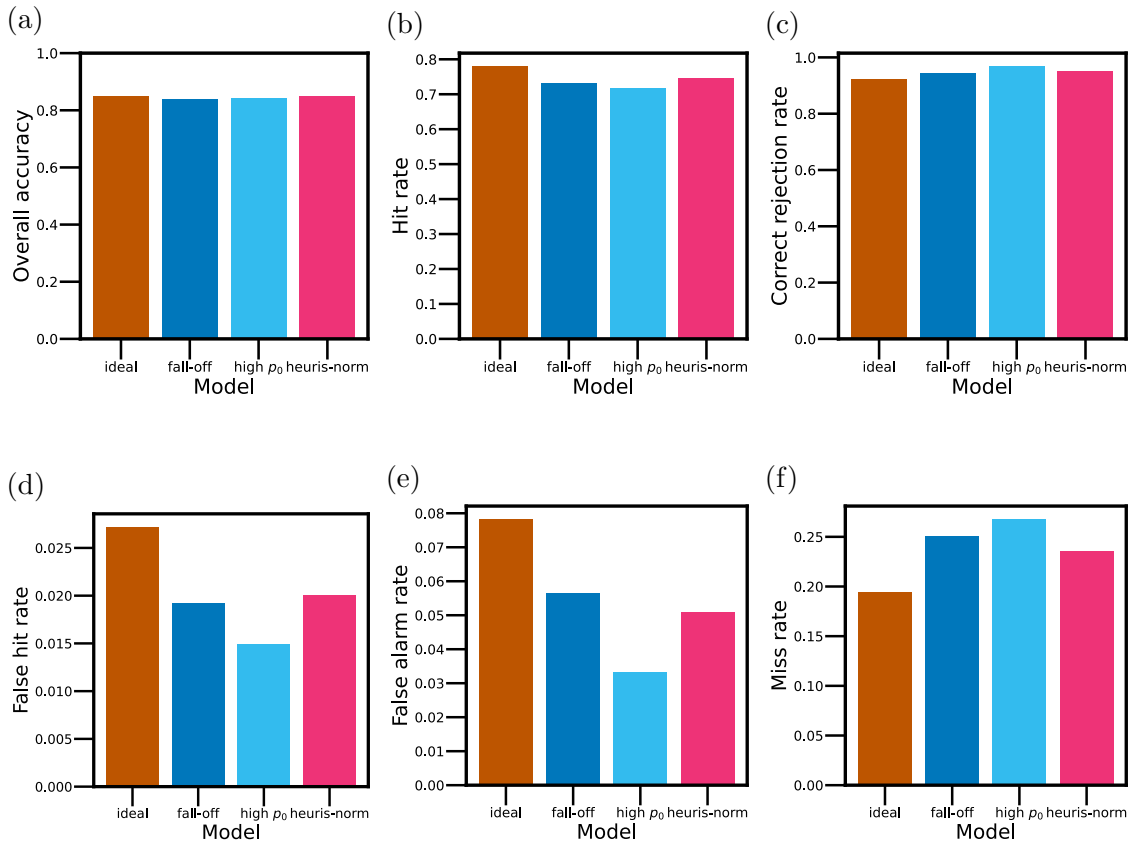


Figure 5.9: Comparison of global statistics among searchers with equal overall accuracy in the 127-location search task. The actual prior and d' maps are flat with $p_0 = 0.5$ and $d'_0 = 4.0$. The fall-off searcher is only heuristic in the d' map, with $\hat{d}'_0 = 7.0$ and $\hat{k}_d = 0.1$, while the high-$\hat{p}_0$ searcher is only heuristic in the prior map with $\hat{p}_0 = 0.7$. The heuristic-normalization searcher uses Equation 4.27 and has a single heuristic parameter $\hat{p}_0 = 0.6$. Statistics include (a) overall accuracy (b) hit rate (c) correct rejection rate (d) false hit rate (e) false alarm rate and (f) miss rate.

three searchers.

Figure 5.11 highlights the trade-off in the search pattern through a heuristic-normalization searcher and a no-centering searcher. These two searchers are in the same configuration of the actual prior and d' maps as above, but they share a search accuracy (60.22%) far from the Bayes-optimal accuracy (85.02%). The heuristic-

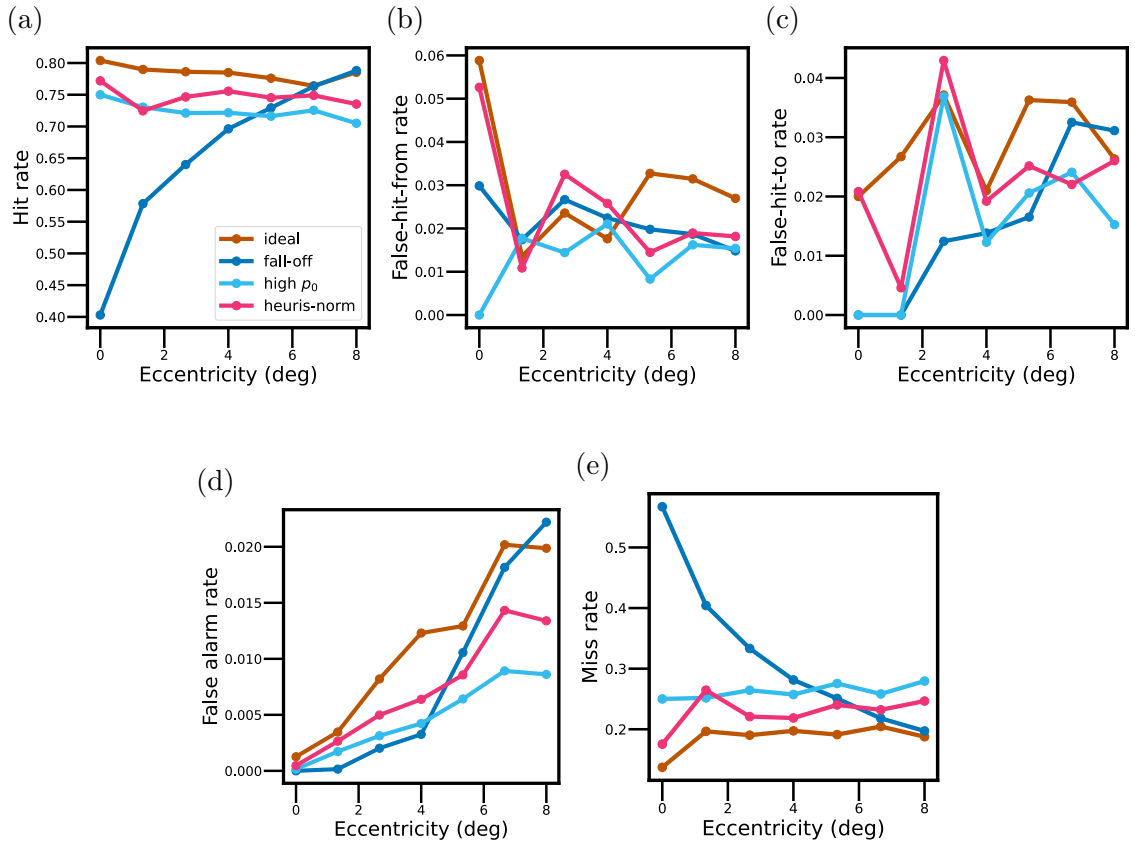


Figure 5.10: Comparison of local statistics among searchers with equal overall accuracy in the 127-location search task. The actual prior and d' maps are flat with $p_0 = 0.5$ and $d'_0 = 4.0$. The fall-off searcher is only heuristic in the d' map, with $\hat{d}'_0 = 7.0$ and $\hat{k}_d = 0.1$, while the high-$\hat{p}_0$ searcher is only heuristic in the prior map with $\hat{p}_0 = 0.7$. The heuristic-normalization searcher uses Equation 4.27 and has a single heuristic parameter $\hat{p}_0 = 0.6$. The following statistics are plotted as a function of eccentricity: (a) hit rate (b) false-hit-from rate (c) false-hit-to rate (d) false alarm rate and (f) miss rate.

normalization searcher has much higher hit and false alarm rates and much lower correct rejection and miss rates than those of the no-centering searcher.

As expected, because both the actual and heuristic maps are flat in both no-centering and heuristic-normalization searchers, local statistics are independent

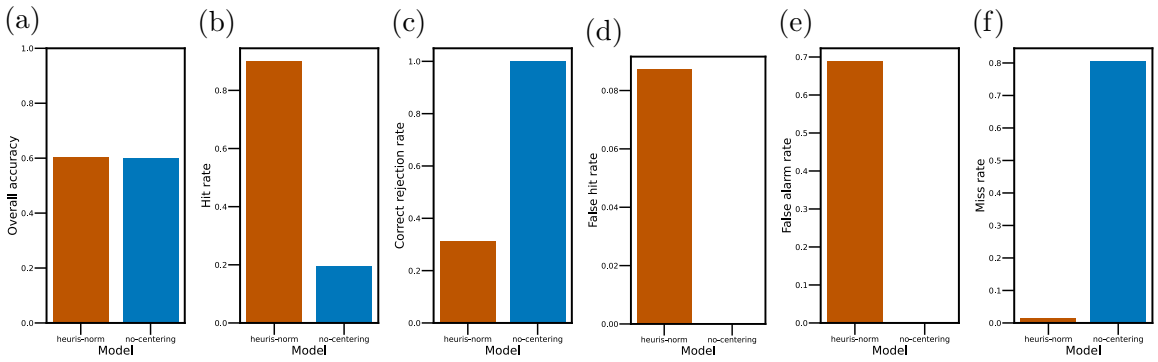


Figure 5.11: Comparison of global statistics among searchers with equal overall accuracy in the 127-location search task. The actual prior and d' maps are flat with $p_0 = 0.5$ and $d'_0 = 4.0$. The heuristic-normalization searcher (Equation 4.27) has a $\hat{p}_0$ of 0.4 and a $\hat{d}'_0$ of 7.0, while the no-centering searcher (Equation 5.3) has a $\hat{d}'_0$ of 1.0. Statistics include (a) overall accuracy (b) hit rate (c) correct rejection rate (d) false hit rate (e) false alarm rate and (f) miss rate.

of eccentricity (Figure 5.12). Specifically, the hit, false-hit-to, and miss rates are constant across eccentricity; the false alarm rate increases as eccentricity increases, merely due to increasing number of locations.

We considered the foveation of the HVS and explored the configuration where the actual d' map decreases at a moderate rate along eccentricity. Four searchers with the same level of search accuracy have varying heuristics in the decision process. As shown in Figure 5.13, the low-$\hat{d}'$ searcher has higher false hit and false alarm rates than other searchers. The high-$\hat{d}'$ and heuristic-normalization searchers have the correct rejection and miss rates higher than those of the Bayes-optimal searcher, and the hit, false hit and false alarm rates lower than that of the Bayes-optimal searcher.

Figure 5.14 shows the differences in their location-dependent statistics. The hit, false-hit-to, false alarm, and miss rates have different speed of change between the high $\hat{d}'$ searcher than the heuristic-normalization searcher, though their global

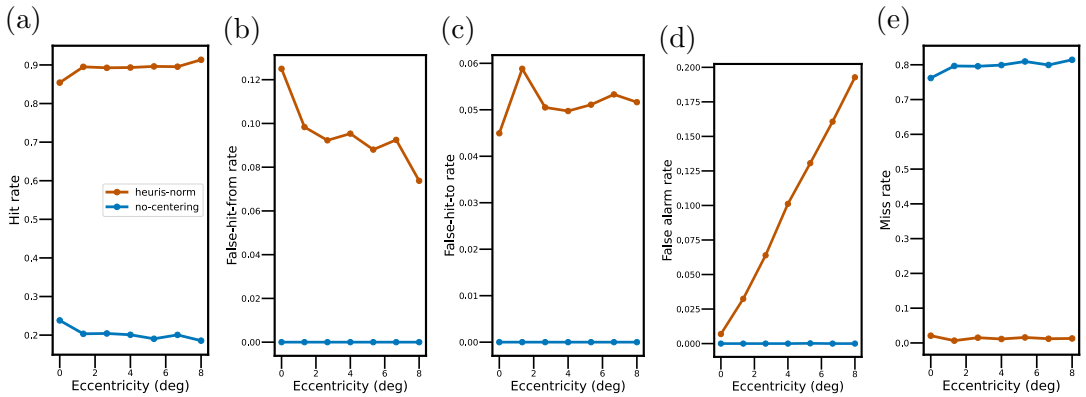


Figure 5.12: Comparison of local statistics among searchers with equal overall accuracy in the 127-location search task. The actual prior and d' maps are flat with $p_0 = 0.5$ and $d'_0 = 4.0$. The heuristic-normalization searcher (Equation 4.27) has a $\hat{p}_0$ of 0.4 and a $\hat{d}'_0$ of 7.0, while the no-centering searcher (Equation 5.3) has a $\hat{d}'_0$ of 1.0. The following statistics are plotted as a function of eccentricity: (a) hit rate (b) false-hit-from rate (c) false-hit-to rate (d) false alarm rate and (f) miss rate.

statistics are similar. Across eccentricity, the low-$\hat{d}'$ searcher has a much steeper false-hit-from rate, a much flatter false-hit-to rate, and an early peak in the false alarm rate. These differences in the detailed search pattern can be served as signs to tell apart different heuristics.

5.4 Discussion

We conducted a systematic analysis of Bayes heuristic covert searchers and presented the distribution of their performance lag to the Bayes-optimal searchers. Though enormous heuristics can reach the same level of overall accuracy in covert search, I demonstrate the possibility of distinguishing different heuristic compositions by comparing the pattern in the location-dependent statistics.

Through the heuristic analysis of Bayesian search, we officially arrive beyond the typical view of Bayesian decision-making, that human perfectly or almost perfectly

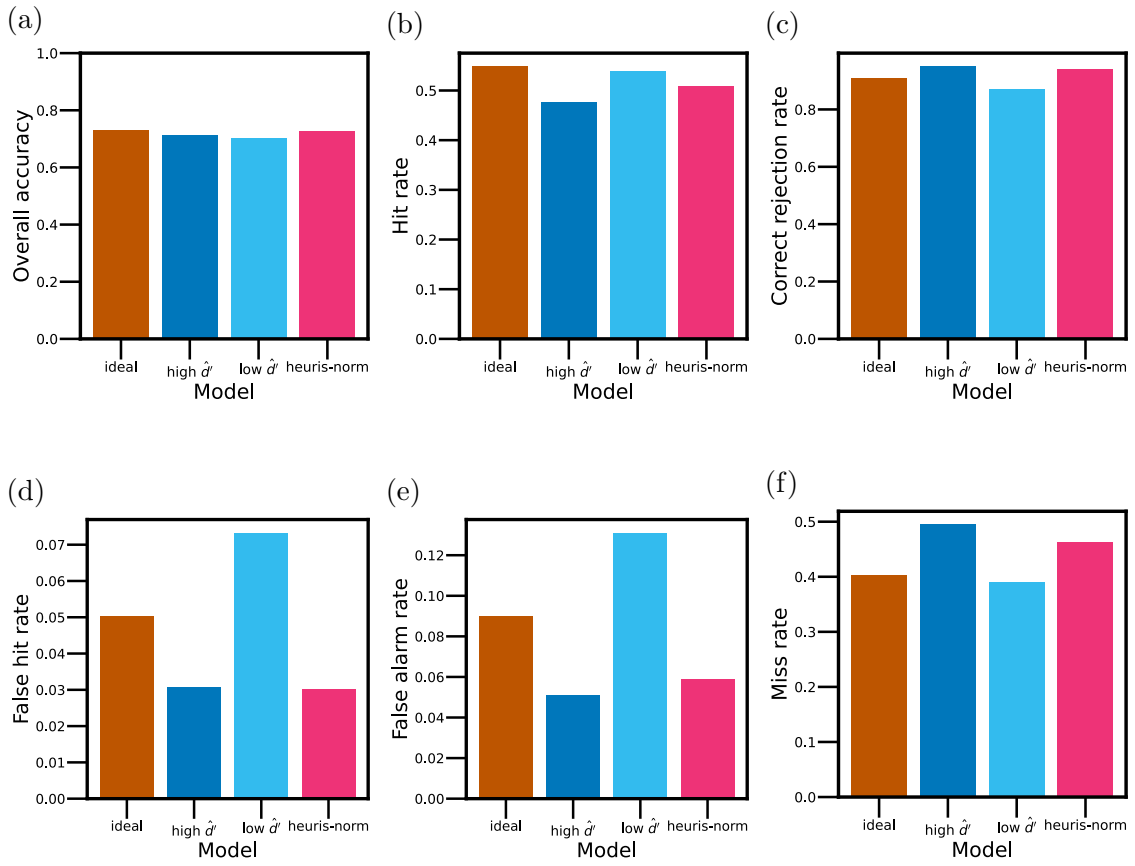


Figure 5.13: Comparison of global statistics among searchers with equal overall accuracy in the 127-location search task. The actual prior map is flat with $p_0 = 0.5$, and the d' map has a d'_0 of 7.0 and a k_d of 0.2. The high-$\hat{d}'$ searcher is only heuristic with $\hat{d}'_0 = 10.0$, while the low-$\hat{d}'$ searcher has a $\hat{p}_0$ of 0.3 and a $\hat{d}'_0$ of 4.0. The heuristic-normalization searcher uses Equation 4.27 and has a single heuristic parameter $\hat{p}_0 = 0.6$. Statistics include (a) overall accuracy (b) hit rate (c) correct rejection rate (d) false hit rate (e) false alarm rate and (f) miss rate.

integrate statistical information, such as incorporating local priors and reliabilities, to make decisions that optimize performance. Our analysis show that for covert search tasks and any other identification tasks that can be expressed with the same mathematical formulation, a broad spectrum of heuristic decision processes not only reduce computation efforts, but also achieve near-optimal performance.

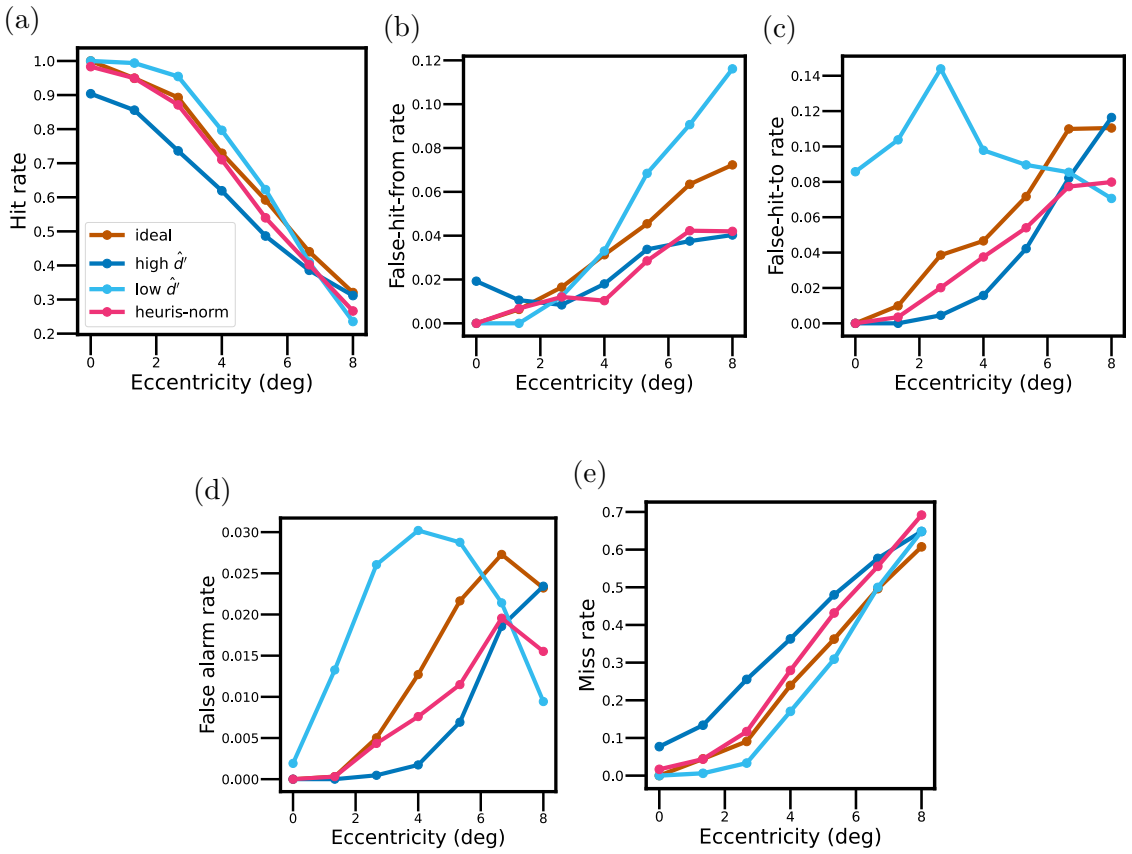


Figure 5.14: Comparison of local statistics among searchers with equal overall accuracy in the 127-location search task. The actual prior map is flat with $p_0 = 0.5$, and the d' map has a d'_0 of 7.0 and a k_d of 0.2. The high-$\hat{d}'$ searcher is only heuristic with $\hat{d}'_0 = 10.0$, while the low-$\hat{d}'$ searcher has a $\hat{p}_0$ of 0.3 and a $\hat{d}'_0$ of 4.0. The heuristic-normalization searcher uses Equation 4.27 and has a single heuristic parameter $\hat{p}_0 = 0.6$. The following statistics are plotted as a function of eccentricity: (a) hit rate (b) false-hit-from rate (c) false-hit-to rate (d) false alarm rate and (f) miss rate.

We acknowledge that though the sensory and heuristic spaces we chose are reasonable, they are not exhaustive. Similar to the measurement of natural scene statistics, one could adventure to measure this natural task statistics for conclusions on heuristics more generalizable than those in this chapter.

Savage [147], who made significant contribution to the modern Bayesian decision theory, distinguished a task environment by its complexity, where a small world refers to a situation where all options along with their probabilities, outcomes and utilities are known, and a large world refers to a situation where some relevant information is unknown or inaccessible, so optimality is ill-defined. In our context of visual detection and search, power-law noise backgrounds and other stationary backgrounds exemplify small worlds, while natural and medical image backgrounds and other non-stationary backgrounds exemplify large worlds. Typically, the Bayes-optimal observer is feasible and optimal in small worlds, and the Bayesian heuristic observers are practical and satisfactory in large worlds. Nevertheless, the heuristic analysis in this chapter shows that near-optimal heuristics can also be common and diverse even in small worlds.

Chapter 6: Conclusion

The overarching research question that I attempt to answer in this dissertation is what and how the human visual system (HVS) computes during detection and search in natural background. We employed a hybrid approach based on Bayesian Decision Theory (BDT) and Signal Detection Theory (SDT), to model, explain, and predict human visual detection and search in natural backgrounds. We compared the behavioral pattern of the HVS in detection and search tasks with that of the Bayes ideal observers, the Bayes observers with biological components, and the Bayes heuristic observers.

Here is a brief summary of our findings. In Chapter 2, we found the HVS fully whitens contrast in space and partially whitens spatial frequencies using the contrast sensitivity function. Including intrinsic position uncertainty in the model increases its explanation power to human detection behavior without introducing any extra parameter. In Chapter 3, we found the target is more detectable to the HVS when the background has a phase structure more similar to that of the target, as a high similarity in phase effectively reduces the effect of intrinsic position uncertainty through the attraction and repulsion mechanisms. We also confirmed the effect of phase-independent similarity, that the target is more detectable when the amplitude spectrum of the background is less similar to that of the target. In Chapter 4, we observed the puzzling phenomenon that the HVS searches better than the prediction of Bayes-optimal decision rule given the human detectability map, despite humans' substantial loss of sensitivity in the fovea, and the implausibility of neurally replicating the complex Bayes-optimal searcher. Correlated internal noise is

a plausible explanation, as it would lower detection performance, but not search performance. We observed the foveal neglect effect in our discrete-location display, that is the loss of foveal detectability due to rational distribution of attentional resources, which confirms a previous report on the search experiments in a continuous display [59]. Furthermore, we discovered extremely simple heuristic decision rules for covert search to achieve near-optimal overall performance, a surprising phenomenon that is essential for explaining the supraoptimal search performance of human observers. Therefore, in Chapter 5, we followed up with a systematic analysis of the Bayesian heuristics in covert search. We found that increasing the number of parameters that are used heuristically in the decision process generally increases the performance lag, but near-optimal and optimal search performance can still be achieved even when all parameters are heuristic, which indicates interaction among heuristic components can cancel out their individual effects on search performance. Heuristic normalization, the absence of log-likelihood centering, and heuristic max rule most likely decrease overall search performance considerably. We also demonstrated the possibility of distinguishing different heuristic compositions that achieve the same level of overall accuracy by comparing the pattern in the location-dependent statistics.

Our discoveries in human detection and search in natural backgrounds can be applied to tools that interact with human vision and tools that emulate human vision. First, knowing how visible a target is to a typical (or professional) human observer allows algorithms and devices to suggest target candidates that are not easily noticeable to human observers but are highly visible to model observers. For example, for threat detection, one could highlight background regions in medical and remote sensing images that give a high template response, yet stay out of phase with the target, because those targets are hard for humans to detect given intrinsic position

uncertainty. The design and placement of visual contents in product packaging, advertisements and transportation can be consulted by how visible they will be on top of backgrounds with varying luminance, contrast, spectrum, and phase-independent and phase-dependent similarities, to achieve the intended goals, such as making the expiration date on a package easy to find, and posting the traffic sign with maximized visibility. A target may also be more detectable by pre-filtering in a way that compliments the contrast sensitivity function for full whitening in spatial frequency. Virtual Reality (VR) and Augmented Reality (AR) can be enhanced by placing virtual contents at proper locations in the visual stimuli and removing naturalistic information that would be filtered away by the HVS to save data transfer bandwidth. Second, knowing when and when not a heuristic search style affects pre-defined metrics of success allows clearer focus and better prioritization to mimic only behaviors that are essential for novice observers to learn from and be trained into expert observers. For instance, the abnormality (e.g., tumors in X-rays, MRIs, and CT scans) detection and search patterns of medical professionals have parts that are crucial to eliminate costly false negative diagnoses, and parts that are flexible heuristics and personal preferences. One could train emerging medical perception workers (e.g., radiologists and pathologists) first with the critical evaluation and computation components of professional searchers. Similarly, when designing a robotic perception-action system, such as autonomous vehicles, the Bayes-optimal detectors and searchers, combined with deep learning algorithms, are capable to grasp the complexity of the real world and navigate it with first principles. A lesson from heuristic analysis is that the heuristic, human-like search strategies can be better than the optimal, computationally expensive strategies, for effort reduction and performance improvement. With heuristic analysis of visual search, one could better understand the diverse, heuristic

behaviors of drivers and how they contribute to traffic accidents, including their inability to perceive traffic rules and risks of collisions, in difficult traffic and weather conditions. The psychophysical measurement of visual detection and search patterns can be used to evaluate physiological states, such as macular degeneration, glaucoma, and cataracts, and cognitive states, such as drunkenness, brain injury, early Alzheimer's disease, and attention deficit hyperactivity disorder (ADHD).

The most restrictive limitation of our Bayesian detection and search framework is statistical clarity. Without knowing the statistical distribution of relevant visual information, the ideal observer is impossible to obtain, and the heuristic searchers are untestable. Therefore, our research requires either the measurement of natural statistics or the artificial construction of visual stimuli. Given statistical clarity, the increasing complexity of the distribution does not invalidate the ideal observer approach, but only demands equally complex mathematical formulation and/or computational operations. In the case that the visual environment is highly volatile, uncertain, complex, and ambiguous, Bayesian ideal observers may be intractable, and even if not, start to fail in predicting human behavior. To stay with Bayesian computation, one often needs to concede to heuristic detection and search models. A promising direction of Bayesian modeling that tolerates statistical obscurity is to combine Bayesian and machine learning algorithms, including the convolution neural network (CNN) and the Transformer neural network. For examples, for non-stationary images, such as natural and medical images, a deep neural network can be used to predict the detectability map (and the prior map) of a target given the background on each trial and integrate those values with Bayes optimal and heuristic computation to predict human detection and search behavior.

Other limitations are much feasible to be overcome. For instance, the number

of human observers and the type of targets can be increased given more resources and time. Current analyses show different individuals detect and search with largely similar patterns, with an overall scaling in thresholds. The same experimental procedures can measure performance also in other natural image databases to verify the generalizability of specific conclusions with regard to natural images. We have not included color and semantics in our visual stimuli and modeling, which are important in real-world search tasks. They can become significant factors that influence the prior and d' maps. Despite the application of the central limit theorem, receptive field responses from a neural population may not be strictly Gaussian, but more Poisson-like [148]. Nevertheless, after including the log-likelihood ratio of two Poisson distributions (into Equation 1.14), the current Bayesian detection and search framework can be generalized. Lastly, the intrinsic position uncertainty is assumed to have a uniform or Gaussian distribution, but in reality may be more anisotropic.

Modifying the Bayes ideal observer with biological components and heuristic decision processes allows our framework not to be limited to Bayes-optimal visual behavior. That implies our experimental design and computational modeling do not necessarily need to be overly simple. It might be true that the HVS is incapable to perform optimally like the ideal observer in many visual tasks, but it is false that the HVS cannot be meaningfully modeled with Bayesian computation in those more complex visual tasks.

My long-term research plan follows the grand roadmap in Section 1.2. First, I will continue to investigate and understand human detection and covert search behavior and computation. The partial whitening in spatial frequency of the HVS can be directly confirmed in natural backgrounds instead of indirectly through power law noises and medical images. One could explore how the asymmetric effect of

phase similarity changes when the stimuli are blocked by phase similarity, testing the existence (and quantitative relationships) of the early d' peak point and the break-even ($d' = 0$) point, so we may know if human observers can learn to flip the direction of the decision rule. Also, I am interested in the similarities and differences of detecting an occluding target, instead of our current additive targets, are as the local contrast modulation, surrounding spatial frequency and phase similarity change. Second, with the knowledge of the HVS in simpler search tasks, one could build Bayesian search models progressively through overt search, embodied search, and social search. A most natural follow-up of our discoveries in covert search is to measure and analyze overt search performance of human observers in similarly discrete display, and then simulate other families of heuristic computation to compare with the ideal overt searcher [77]. At least two new computation components emerge: (1) probability update across fixations; (2) fixation selection. The former can have heuristic memory that update crudely. The latter can have heuristic saccades that do not follow the ideal fixation rule, but rules that require much less spatial integration of information. Last but not least, the aforementioned fundamental research will be translated and applied to real-world search problems. I am particularly interested in the problems in medical image perception. Detection and search misses and delays in diagnosis are much more likely to be perceptual (60-80%) than cognitive [149, 150], and they often cause missed opportunities for early treatment. I am interested in assisting medical professionals to detect and search targets with higher accuracy and speed in medical images, by applying and interpreting the ideal searcher, near-optimal heuristic searchers, and human-like searchers within a specific diagnosis context, such as the digital breast tomosynthesis. One of a few examples mentioned in the early part of this chapter is pre-filtering medical images in a way that compliments the contrast

sensitivity function of the HVS, so full whitening in spatial frequency is achieved. Furthermore, medical images are also much less accessible than natural images for data privacy. That poses a challenge for deep neural networks to have sufficient data for training. Besides augmenting data through generative models, Bayesian models can be combined with neural networks to extract relevant information more efficiently. Overall, our Bayesian detection and search framework is accurate, fast, flexible, and interpretable for applications in medical image perception.

Appendix A: Ideal detection and search rules with utility

As mentioned in Chapter 1.9, the optimal decision rules for visual detection and search are encapsulated in Equations 1.19, 1.20, and 1.22. Here, we discuss the case when the utility function to be maximized is neither likelihood nor posterior probability, but a linear combination of rates in different trial types.

Suppose the utility for responding $\hat{x}$ when target present at x is $u_{x\hat{x}}$. The optimal search rule that maximizes the expected posterior utility is

$$\begin{aligned}
\hat{x} &= \underset{x\in\mathbb{X}}{arg\,max} \left\{ \sum_{x\in\mathbb{X}} \left[u_{x\hat{x}} p_x \prod_{y\in\mathbb{Y}} p(D_y|x=x) \right] \right\} \\
&= \underset{\hat{x}\in\mathbb{X}}{arg\,max} \left\{ \sum_{x\in\mathbb{X}} \left[u_{x\hat{x}} p_x \prod_{y\in\mathbb{Y}} p(D_y|x=x) \right] / \prod_{y\in\mathbb{Y}} p(D_y|x=0) \right\} \\
&= \underset{\hat{x}\in\mathbb{X}}{arg\,max} \left\{ \sum_{x\in\mathbb{X}} \left[u_{x\hat{x}} p_x \prod_{y\in\mathbb{Y}} \frac{p(D_y|x=x)}{p(D_y|x=0)} \right] \right\}
\end{aligned} \tag{A.1}$$

Following the same steps in Equations 1.18 and 1.19, we obtain

$$\hat{x} = \underset{\hat{x}\in\mathbb{X}}{arg\,max} \left[\ln \sum_{x\in\mathbb{X}} u_{x\hat{x}} p_x l_x \right] \tag{A.2}$$

When the decision variable follows a Gaussian distribution with equal variance at each location, we can represent the likelihood ratio as Equation 1.25, so the optimal search rule becomes

$$\hat{x} = \underset{\hat{x}\in\mathbb{X}}{arg\,max} \left\{ \ln \sum_{x\in\mathbb{X}} u_{x\hat{x}} p_x \exp\left[d'_x (R'_x - d'_x/2) \right] \right\} \tag{A.3}$$

Now I derive the optimal detection rule. Because the target can at most be present at one location, the posterior of target presence is the sum of posterior probabilities of target at each present location (mutually exclusive events). The present-absent utility ratio is

$$UR = \frac{\sum_{y\in\mathbb{Y}}\sum_{x\in\mathbb{X}}\left[u_{xy}p_x\prod_{y'\in\mathbb{Y}}p(D_{y'}|x=x)\right]}{\sum_{x\in\mathbb{X}}u_{x0}p_x\prod_{y'\in\mathbb{Y}}p(D_{y'}|x=x)} \tag{A.4}$$

Dividing both the numerator and the denominator by $u_{00}\prod_{y'\in\mathbb{Y}}p(D_{y'}|x=0)$, we have

$$UR = \frac{\sum_{y\in\mathbb{Y}}\sum_{x\in\mathbb{X}}g_{xy}p_x l_x}{\sum_{x\in\mathbb{X}}g_{x0}p_x l_x} \tag{A.5}$$

where $g_{x\hat{x}} = \frac{u_{x\hat{x}}}{u_{00}}$ is the utility ratio between responding $\hat{x}$ when the target is at x and responding target-absent correctly.

When the decision variable follows a Gaussian distribution with equal variance at each location, the present-absent utility ratio becomes

$$UR = \frac{\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}}g_{xy}p_x\exp[d'_x(R'_x - d'_x/2))]}{\sum_{x\in\mathbb{X}}g_{x0}p_x\exp[d'_x(R'_x - d'_x/2)]} \tag{A.6}$$

Therefore, the optimal detection rule that maximizes the expected utility

$$\hat{S} = \begin{cases} a & UR < 1 \\ b & UR > 1 \end{cases} \tag{A.7}$$

If the utility matrix $u_{x\hat{x}} = \mathbb{I}_{\hat{x}=x}$, where $\mathbb{I}$ is the indicator function, then the optimal search rule in Equation A.2 becomes $\hat{x} = \underset{\hat{x}\in\mathbb{X}}{arg\,max}\ln[p(\hat{x})l_{\hat{x}}]$, just as the MAP

search rule (Equation 1.20); the present-absent utility ratio in Equation A.5 becomes $UR = \sum_{y \in \mathbb{Y}} r(y) l_y$, where $r(y)$ is the prior ratio, consistent with Equation 1.21.

Appendix B: Ideal detection and search rules with multiple targets

What are the optimal detection and search rules when multiple deterministic targets can be present and at different locations?

Let each of the n possible target locations independently has a prior distribution with one or none of the m targets present (additive to backgrounds). A template T_z is an element in the template set $\mathbb{T} = \{T_0, T_1, \cdots, T_m\}$, where T_0 is the "zero target" made of a zero matrix, indicating target-absent. Define $\mathbb{Z} = \{0, 1, \cdots, m\}$ and k is the target-presence vector of length n, with each element in $\mathbb{Z}$. For example, if no target is present at any location, then $\vec{k} = (0, \cdots, 0)$.

The prior probability of target $a_z T_z$ present at the location y is p_{yz}. By definition, $\sum_{z=0}^{m} p_{yz} = 1$. The log likelihood ratio at the location y between the presence of target $a_z T_z$ and target absence is defined as ll_{yz}. Notice that the denominator of the ratio can have different values for different locations.

The maximum likelihood response for the detection task in this setting is undefined. The MAP response for the detection task in this setting is

$$\hat{S} = \begin{cases} a & \forall y \in \mathbb{Y}, p_{y0} ll_{y0} > \sum_{z \in \mathbb{Z}, z \neq 0} p_{yz} ll_{yz} \\ b & \exists y \in \mathbb{Y}, p_{y0} ll_{y0} < \sum_{z \in \mathbb{Z}, z \neq 0} p_{yz} ll_{yz} \end{cases} \tag{B.1}$$

where "a" means no target is present at any location, and "b" means at least one target is present at a location.

In the case where background at each location is uniform white noise, then in the same way as Equation 1.14, the log likelihood ratio at the location y for target

$a_z T_z$ is

$$ll_{yz} = \frac{a_z}{\sigma_y^2}(D_y \cdot T_z - \frac{a_z}{2}) \tag{B.2}$$

As long as a decision variable follows Gaussian distributions with equal variance, just as Equation 1.25, we have

$$ll_{yz} = d'_{yz}(R'_{yz} - d'_{yz}/2) \tag{B.3}$$

where d'_{yz} is the detectability for target $a_z T_z$ at the location y, and

$$R'_{yz} \sim \begin{cases} N(0,1) & z \neq k_y \\ N(d'_{yz},1) & z = k_y \end{cases} \tag{B.4}$$

The definitions of the search task can vary in several ways. If the search task asks to report the target-presence vector $\vec{k}$, then the maximum likelihood response

$$\forall y \in \mathbb{Y}, \hat{k}_y = \underset{z \in \mathbb{Z}}{arg\,max}\, ll_{yz} \tag{B.5}$$

The MAP response

$$\forall y \in \mathbb{Y}, \hat{k}_y = \underset{z \in \mathbb{Z}}{arg\,max}\, p_{yz} ll_{yz} \tag{B.6}$$

If the search task asks to report the number of present targets q, then the maximum likelihood response and the MAP response are both

$$\hat{q} = \sum_{y \in \mathbb{Y}} \mathbb{I}_{\hat{k}_y \neq 0} \tag{B.7}$$

while $\hat{\vec{k}}$ is obtained from Equations B.5 or B.6.

Appendix C: Ideal detection and search rules with spatial-temporal correlation

Here I derive the ideal searcher where multiple template responses are measured across spatial location and time are correlated. If the template responses at a location across time or at a time across locations are not correlated, the result below can still be applied, as the covariance matrix becomes closer to an identity matrix.

The detection or search task can be divided temporally into c observation cycles, and the cycle set $\mathbb{C} = \{0, 1, \cdots, c_{max}\}$ (not the complex number set), where "0" is the cycle prior to any observation, c_{max} is the maximum number of cycles across locations. In the same notation as those in the earlier appendices, there are n potential target-present locations, and the target absent "location" is "0".

At the end of the c-th cycle, the probability of target presenting at location x is $p_x(c)$, and the log-likelihood and log-posterior ratios of target presenting at location x versus target absence are $ll_x(c)$ and $lp_x(c)$, respectively. For example, the prior of target absence at the beginning of the task is $p_x(c) = p_0(0)$.

In an overt search, each cycle has three components: (1) information integration; (2) termination evaluation; (3) fixation selection. I will use this search scenario to explain the ideal searcher with temporal correlation.

Information integration

How responses are integrated optimally? If they follow multivariate Gaussian distributions with equal variance, such as $N(\boldsymbol{\mu}_p, \Sigma^2)$ and $N(\boldsymbol{\mu}_a, \Sigma^2)$, then the overall cumulative log-likelihood ratio is

$$LL = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_a)^T \Sigma^{-1} \boldsymbol{R} + \frac{1}{2}(\boldsymbol{\mu}_a^T \Sigma^{-1} \boldsymbol{\mu}_a - \boldsymbol{\mu}_p^T \Sigma^{-1} \boldsymbol{\mu}_p) \tag{C.1}$$

Note that the multiplication in the exponent term of the multivariate normal distribution can be distributed, because Σ^{-1} is symmetric.

Then we re-center the template response so that $\boldsymbol{\mu}_a = 0$, and

$$LL = \boldsymbol{d'}^T \boldsymbol{R'} - \frac{1}{2}\boldsymbol{d'}^T \boldsymbol{d'} \tag{C.2}$$

where $\boldsymbol{d'} = \Sigma^{-1/2}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_a)$ is the normalized detectability, and $\boldsymbol{R'} = \Sigma^{-1/2}\boldsymbol{R}$ is the normalized template response. When the target is present at location y, $\boldsymbol{R'}_y \sim N(\boldsymbol{d'}_y, I)$, and $\forall y' \neq y, \boldsymbol{R'}_{y'} \sim N(\boldsymbol{0}, I)$, where I is the identity matrix.

The number of dimensions of Σ is $(n \cdot c_{max}) \times (n \cdot c_{max})$. In the case where responses correlate either only spatially or only temporally, corresponding parts of the covariance matrix turns into an identity matrix.

After decorrelation, the log-likelihood ratio at location y can be defined as

$$LL_y = \boldsymbol{d'}_y^T \boldsymbol{R'}_y - \frac{1}{2}\boldsymbol{d'}_y^T \boldsymbol{d'}_y \tag{C.3}$$

where $\boldsymbol{d'}_y$ and $\boldsymbol{R'}_y$ have c_{max} dimensions, and Σ_y has $c_{max} \times c_{max}$ dimensions. The log-posterior ratio is $LP_y = \ln p_y + LL_y$ or $LP_y = \ln \frac{p_y}{p_0} + LL_y$, if $p_0 \neq 0$, which is proportional to the exponent of the posterior probability after information integration.

Termination evaluation

How does an observer evaluate and decide if a specific cycle can be the final observation. Strictly speaking, this decision rule needs to collaborate with the decision

rule for fixation selection to ensure optimality. For example, Najemnik and Geisler [77] chose the fixation selection rule that maximizes the expected accuracy after the fixation, assuming a maximum a posteriori (MAP) decision made right after the fixation. Then the optimal termination rule is to end the search right after the fixation, despite that means only one fixation is made for search.

In practice, an observer typically has resources more than what one fixation requires, which means one can give more time and effort to observe and integrate information. Nevertheless, in the next part (fixation selection), I will not derive the optimal observer that maximizes expected accuracy more than one fixation ahead (e.g., choosing the next fixation that maximizes search accuracy after 10 fixations), due to analytical limitations. Instead, I will use the same expected accuracy with one-fixation-ahead as in Najemnik and Geisler [77], and extend to the cases where the template responses can be correlated across fixations, and the target can sometimes be absent.

All that to say, termination evaluation is commonly decoupled from information integration and fixation selection. We can set a single decision criterion γ, and decide termination in the following way:

$$\hat{x}(c) = \begin{cases} \underset{x \in \mathbb{X}}{arg\,max}\, p_x(c) & \underset{x \in \mathbb{X}}{max}\, p_x(c) > \gamma \\ \text{decide later} & \text{else} \end{cases} \tag{C.4}$$

We can also set two separate criteria γ_p and γ_a (assuming both are no less than 0.5) for responding target-present and target-absent:

$$\hat{x}(c) = \begin{cases} \underset{y \in \mathbb{Y}}{arg\,max}\, p_y(c) & \underset{y \in \mathbb{Y}}{max}\, p_y(c) > \gamma_p \\ 0 & p_0(c) > \gamma_a \\ \text{decide later} & \text{else} \end{cases} \tag{C.5}$$

Another timing to terminate the search is when the number of observation cycles reaches a certain threshold value γ_c, which corresponds to the depletion of search resources, such as

$$\hat{x}(c) = \begin{cases} \underset{x \in \mathbb{X}}{arg\,max}\, p_x(c) & \underset{x \in \mathbb{X}}{max}\, p_x(c) > \gamma \text{ or } c > \gamma_c \\ \text{decide later} & \text{else} \end{cases} \tag{C.6}$$

Fixation selection

Define $\mathbb{F}$ as the fixation space that includes all locations the observer can fixate at, such as every pixel on the display. We denote its cardinality $|\mathbb{F}| = F$. Fixation is the most typical action to modify the d' map, given a specific target and background. In 3D images, scrolling can be generalized as "fixation" and join this space.

As pointed out by Najemnik and Geisler [77], if the target is always present, the fixation selection rule that maximizes the expected accuracy with one fixation ahead is

$$\hat{f}(c+1) = \underset{f \in \mathbb{F}}{arg\,max} \left[\sum_{y \in \mathbb{Y}} p_y(c) p\left(\hat{x}(c+1) = y | y, f\right) \right] \tag{C.7}$$

where $p(\hat{x}(c+1) = y|y, f)$ is the probability that the MAP response right after the fixation is location y, given the target is present at y, after fixating at f. In other words, $\hat{x}(c+1) = \underset{y \in \mathbb{Y}}{arg\,max}\, p_y(c+1)$.

Let us consider a more general case where c fixations have been made and expected accuracy with c' fixations ahead. All possible next c' fixations reside in the space $\mathbb{F}_c(c')$. For each location y, the predicted MAP response is correct if the posterior, or the log-posterior ratio is the largest at y:

$$p(\hat{x}(c+c') = y|y, f) = p(\forall y', lp_y(c+c') \geq lp_{y'}(c+c')|y, \boldsymbol{f})$$
$$= p(\forall y' \neq y, lp_y(c) + LL_y(c') \geq lp_{y'}(c) + LL_{y'}(c')|y, \boldsymbol{f}) \tag{C.8}$$

where $LL_y(c')$ is the cumulative likelihood ratio between target presenting at location y and target absence after c' cycles of observation, in the expression of Equation C.2.

Let $\forall y, y', c', Z_y(c'), Z_{y'}(c') \overset{i.i.d.}{\sim} N(0,1)$, $\forall c', R'_y(c') = Z_y(c') + d'_y(f(c'))$ and $R'_{y'}(c') = Z_{y'}(c')$. Combined with Equation C.2, the inequality in Equation C.8 becomes

$$\ln \frac{p_y(c)}{p_{y'}(c)} + \boldsymbol{d'}_y^T \boldsymbol{Z}_y + \frac{\boldsymbol{d'}_y^T \boldsymbol{d'}_y}{2} \geq \boldsymbol{d'}_{y'}^T \boldsymbol{Z}_{y'} - \frac{\boldsymbol{d'}_{y'}^T \boldsymbol{d'}_{y'}}{2} \tag{C.9}$$

Note that $\boldsymbol{f}$, $\boldsymbol{d'}$, $\boldsymbol{Z}_y$, and $\boldsymbol{Z}_{y'}$ have c' dimensions.

Based on the summation of independent Gaussian distributions (Equation 1.15), the inequality can be simplified further into

$$\ln \frac{p_y(c)}{p_{y'}(c)} + ||\boldsymbol{d'}_y|| Z_y + \frac{||\boldsymbol{d'}_y||^2}{2} \geq ||\boldsymbol{d'}_{y'}|| Z_{y'} - \frac{||\boldsymbol{d'}_{y'}||^2}{2} \tag{C.10}$$

or

$$Z_{y'} \leq \frac{\ln \frac{p_y(c)}{p_{y'}(c)} + ||\boldsymbol{d'}_y|| Z_y + \frac{1}{2}(||\boldsymbol{d'}_y||^2 + ||\boldsymbol{d'}_{y'}||^2)}{||\boldsymbol{d'}_{y'}||} \tag{C.11}$$

where $Z_y, Z_{y'} \overset{i.i.d.}{\sim} N(0,1)$, $||\boldsymbol{d'}|| = \sqrt{\boldsymbol{d'}_{y'}^T \boldsymbol{d'}_{y'}}$ is the vector norm of $\boldsymbol{d'}$.

As this inequality needs to be true for all locations that are not y, we can condition on Z_y and obtain

$$p(\hat{x}(c+c') = y|y, \boldsymbol{f}) = \int_{-\infty}^{\infty} p(z_y) \prod_{y' \neq y} p(z_{y'} \leq q(y, y', \boldsymbol{f}, z_y)|y, f) dz_y$$
$$= \int_{-\infty}^{\infty} \phi(z) \prod_{y' \neq y} \Phi(q(y, y', \boldsymbol{f}, z)) dz \tag{C.12}$$

where ϕ and Φ are the PDF and CDF of the standard normal distribution, and

$$q(y, y', \boldsymbol{f}, z) = \frac{\ln \frac{p_y(c)}{p_{y'}(c)} + ||\boldsymbol{d'}_y||z + \frac{1}{2}(||\boldsymbol{d'}_y||^2 + ||\boldsymbol{d'}_{y'}||^2)}{||\boldsymbol{d'}_{y'}||} \tag{C.13}$$

Then a series of fixations can be determined by:

$$\hat{\boldsymbol{f}} = \underset{\boldsymbol{f} \in \mathbb{F}_c(c')}{arg\,max} \left[\sum_{y \in \mathbb{Y}} p_y(c) p\left(\hat{x}(c+c') = y|y, \boldsymbol{f}\right) \right] \tag{C.14}$$

To this point, I derived the maximum expected (c'-fixation-ahead) accuracy observer when template responses correlate across fixations and the target is always present.

If the target has an absent prior of $p_0(0)$, then the fixation selection rule includes the target absent "location", that is

$$\hat{f}(c+c') = \underset{f \in \mathbb{F}_c(c')}{arg\,max} \left[\sum_{x \in \mathbb{X}} p_x(c) p\left(\hat{x}(c+c') = x|x, \boldsymbol{f}\right) \right] \tag{C.15}$$

Following the derivation, I notice $p(\hat{x}(c+1) = y|y, \boldsymbol{f})$ ($\forall y \in \mathbb{Y}$, see Equation C.8) now includes the inequality $lp_y(c) + LL_y(c') \geq 0$. That changes the lower bound of the expected accuracy integral (Equation C.12), so that

$$p(\hat{x}(c+c')=y|y,\boldsymbol{f}) = \int_{-(\frac{lp_y(c)}{||\boldsymbol{d'}_y||}+\frac{||\boldsymbol{d'}_y||}{2})}^{\infty} \phi(z)\prod_{y'\neq y}\Phi(q(y,y',\boldsymbol{f},z))dz \tag{C.16}$$

The computation of $q(y,y',\boldsymbol{f},z)$ still follows Equation C.13. The last term to calculate is

$$\begin{aligned} p(\hat{x}(c+c')=0|0,\boldsymbol{f}) &= p(\forall y, 0 \geq lp_y(c) + ||\boldsymbol{d'}_y||Z_y - \frac{||\boldsymbol{d'}_y||^2}{2}|0,\boldsymbol{f}) \\ &= \prod_{y\in\mathbb{Y}}\Phi\left(\frac{||\boldsymbol{d'}_y||}{2} - \frac{lp_y(c)}{||\boldsymbol{d'}_y||}\right) \end{aligned} \tag{C.17}$$

In summary, when template responses are correlated temporally, the ideal searcher normalizes those responses and corresponding d' values by the covariance matrix, considers the variation in responses after the next fixation, and selects a fixation location that maximizes the expected accuracy.

Appendix D: Multidimensional power-law noise

We name a noise in n dimensions following a radially symmetric power-law in spectrum as a power-law noise. Its power spectral density (PSD) $P(\vec{f}) = P(f)$, where $\vec{f}$ is the spatial frequency, $f = ||\vec{f}||$ is the amplitude of the spatial frequency, and $P(f)$ is the one-dimensional, cross-sectional radial power spectral density, different from the one-dimensional, total radial power spectral density $P_t(f)$; the total noise power

$$P = \int P_t(f)df = \int S_{n-1}(f)P(f)df \tag{D.1}$$

where $S_{n-1}(f) = \frac{2\pi^{n/2}}{\Gamma(n/2)}f^{n-1}$ is the area of an (n-1)-sphere.

If $P(f) \propto f^{-\beta}$, then $P_t(f) \propto f^{n-\beta-1}$, the one-dimensional, cross-sectional radial amplitude spectral density $A(f) = \sqrt{P(f)} \propto f^{-\beta/2}$, the one-dimensional, total radial amplitude spectral density $A_t(f) \propto \sqrt{P_t(f)} \propto f^{(n-\beta-1)/2}$. If $n \neq \beta$, the total noise amplitude $A \propto \sqrt{P} \propto f^{(n-\beta)/2}$.

In my dissertation, $1/f$ noise specifically refers to the two-dimensional power-law noise where $\beta = 2$, with $P(f) \propto f^{-2}$, $A(f) \propto f^{-1}$, $P_t(f) \propto f^{-1}$, $A_t(f) \propto f^{-1/2}$. The total power

$$\int_{f_{min}}^{f_{max}} P_t(f)df \propto \ln(\frac{f_{max}}{f_{min}}) \tag{D.2}$$

We allow the $P_t(0)$ component of a power-law image to be either 0 or another arbitrary value. The minimum non-zero frequency (that is still radially complete) is

$f_{min} = 1/s$ cycles per pixel, where s is the number of pixels of the shorter side of the image. The maximum frequency is $f_{max} = 1/2$ cycles per pixel (Nyquist frequency). Therefore, the total power scales with the log of image radius, that is $\ln(s/2)$.

Appendix E: Confusion in position discrimination by uncertainty

In a binary position discrimination task where the target is either shifted a displacement amplitude of a to either the left or the right of the center, we define a simple decision rule for the response as

$$\hat{S} = \begin{cases} l, & x < \gamma \\ r, & x > \gamma \end{cases} \tag{E.1}$$

where S is the position state of the target, $\hat{S}$ is an estimate of that state, l indicates the target is shifted to the left of the center, r indicates the target is shifted to the right of the center, x is the is perceived horizontal location of the target, γ is a location criterion parameter.

The two-state confusion matrix categorizes trials into the following types: (1) true left rate: $P(\hat{S} = l|S = l)$; (2) true right rate: $P(\hat{S} = r|S = r)$; (3) false left rate: $P(\hat{S} = l|S = r)$; (4) false right rate: $P(\hat{S} = r|S = l)$.

Figure E.1a illustrates a two-dimensional, isotropic Gaussian model of intrinsic position uncertainty. When the target is on either the left at $(-a, 0)$ or the right at $(a, 0)$, the perceived location follows these two probability density functions:

$$p(x, y|S = l) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}(\frac{x+a}{\sigma})^2 - \frac{1}{2}(\frac{y}{\sigma})^2} \tag{E.2}$$

$$p(x, y|S = r) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}(\frac{x-a}{\sigma})^2 - \frac{1}{2}(\frac{y}{\sigma})^2} \tag{E.3}$$

Over infinite trials, the true left rate

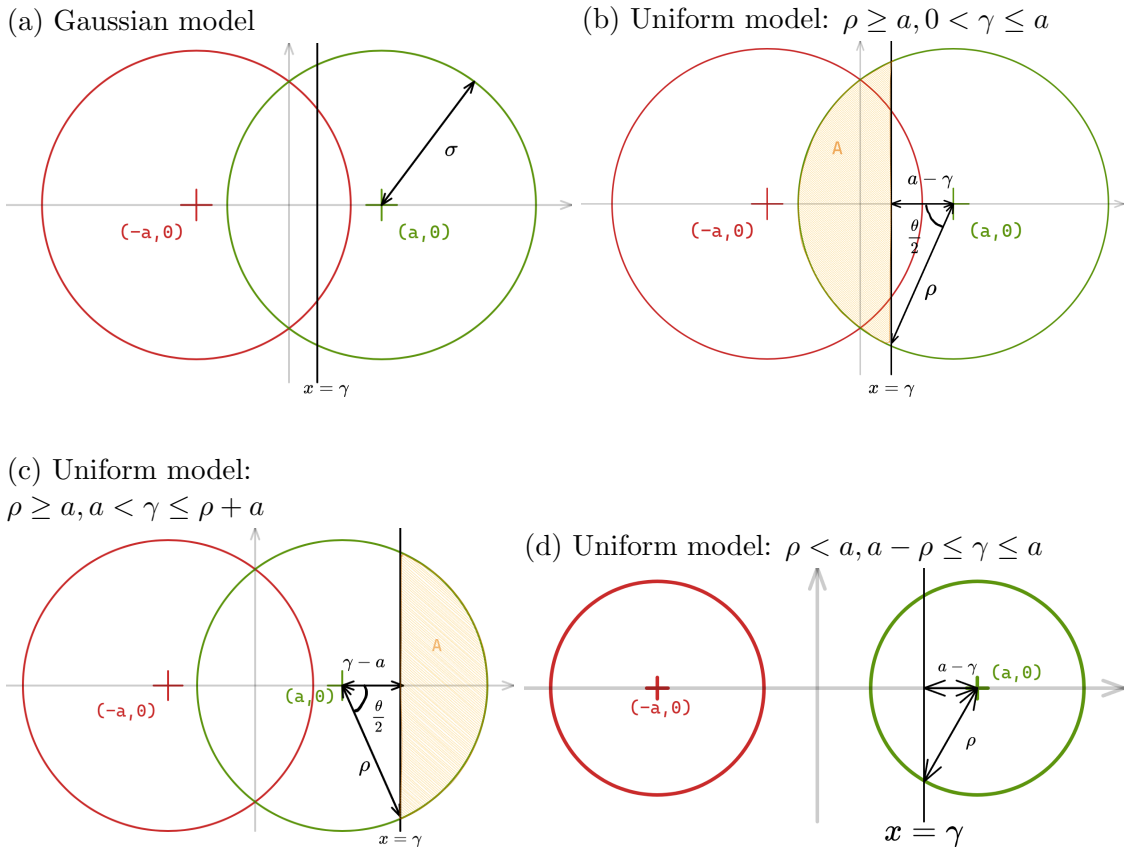


Figure E.1: Gaussian and uniform models for intrinsic position uncertainty. The standard deviation of the Gaussian model is σ. The radius of the uniform model is ρ. The displacement amplitude is a.

$$P(\hat{S} = l|S = l) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\gamma} p(x, y|S = l)dx \right] dy \tag{E.4}$$

$$\overset{z=\frac{x+a}{\sigma}}{=} \int_{-\infty}^{\frac{\gamma+a}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \Phi(\frac{\gamma + a}{\sigma}) \tag{E.5}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

The false right rate $P(\hat{S} = r|S = l) = 1 - P(\hat{S} = l|S = l) = \Phi(-(a + \gamma)/\sigma)$. The false left rate

$$P(\hat{S}=l|S=r)=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\gamma}p(x,y|S=r)dx\right]dy$$
$$\overset{z=\frac{x-a}{\sigma}}{=}\int_{-\infty}^{\frac{\gamma-a}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=\Phi(\frac{\gamma-a}{\sigma}) \tag{E.6}$$

The true right rate $P(\hat{S}=r|S=r)=1-P(\hat{S}=l|S=r)=\Phi((a-\gamma)/\sigma)$.

Alternatively, we consider a two-dimensional, isotropic uniform model of intrinsic position uncertainty. When the target is on either the left at $(-a,0)$ or the right at $(a,0)$, the perceived location follows these two probability density functions:

$$p(x,y|S=l)=\frac{1}{\pi\rho^2}\mathbb{I}_{(0,\rho]}(\sqrt{x^2+y^2}+a) \tag{E.7}$$
$$p(x,y|S=r)=\frac{1}{\pi\rho^2}\mathbb{I}_{(0,\rho]}(\sqrt{x^2+y^2}-a) \tag{E.8}$$

where $\mathbb{I}$ is the indicator function.

The calculation of the confusion matrix in uniform distribution is more complicated. I will first derive the cell values in the confusion matrix for major example cases, and then write down the complete, case-by-case expression of the rate for each trial type.

When $\rho\geq a, 0<\gamma\leq a$ (Figure E.1b), the segment highlighted by orange lines has an area of $A=\rho^2(\theta-\sin\theta)/2$, where the angle $\theta=2\arccos((a-\gamma)/\rho)$, so that $\sin\theta=2(a-\gamma)\sqrt{\rho^2-(a-\gamma)^2}/\rho^2$, and

$$A=\rho^2\arccos(\frac{a-\gamma}{\rho})-(a-\gamma)\sqrt{\rho^2-(a-\gamma)^2} \tag{E.9}$$

The rates in the confusion matrix can be calculated as

$$P(\hat{S}=l|S=r)=\frac{A}{\pi\rho^2}=\frac{1}{\pi}\left[\arccos(\frac{a-\gamma}{\rho})-\frac{a-\gamma}{\rho^2}\sqrt{\rho^2-(a-\gamma)^2}\right] \tag{E.10}$$

$$P(\hat{S}=r|S=r)=1-P(\hat{S}=l|S=r)=\frac{1}{\pi}\left[\arccos(\frac{\gamma-a}{\rho})+\frac{a-\gamma}{\rho^2}\sqrt{\rho^2-(a-\gamma)^2}\right] \tag{E.11}$$

$$P(\hat{S}=r|S=l)=P(\hat{S}=l|S=r)|_{\gamma=-\gamma}=\frac{1}{\pi}\left[\arccos(\frac{a+\gamma}{\rho})-\frac{a+\gamma}{\rho^2}\sqrt{\rho^2-(a+\gamma)^2}\right] \tag{E.12}$$

$$P(\hat{S}=l|S=l)=1-P(\hat{S}=l|S=r)=\frac{1}{\pi}\left[\arccos(\frac{-\gamma-a}{\rho})+\frac{a+\gamma}{\rho^2}\sqrt{\rho^2-(a+\gamma)^2}\right] \tag{E.13}$$

When $\rho \geq a, a < \gamma \leq \rho + a$ (Figure E.1c), the segment highlighted by orange lines still has an area of $A = \rho^2(\theta - \sin\theta)/2$, but the angle $\theta = 2\arccos((\gamma - a)/\rho)$, so that $\sin\theta = 2(\gamma - a)\sqrt{\rho^2 - (\gamma - a)^2}/\rho^2$, and

$$A=\rho^2\arccos(\frac{\gamma-a}{\rho})-(\gamma-a)\sqrt{\rho^2-(\gamma-a)^2} \tag{E.14}$$

Therefore, the rates in the confusion matrix become

$$P(\hat{S}=r|S=r)=\frac{A}{\pi\rho^2}=\frac{1}{\pi}\left[\arccos(\frac{\gamma-a}{\rho})-\frac{\gamma-a}{\rho^2}\sqrt{\rho^2-(\gamma-a)^2}\right] \tag{E.15}$$

$$P(\hat{S}=l|S=r)=\frac{1}{\pi}\left[\arccos(\frac{a-\gamma}{\rho})+\frac{\gamma-a}{\rho^2}\sqrt{\rho^2-(\gamma-a)^2}\right] \tag{E.16}$$

$$P(\hat{S}=l|S=l)=\frac{1}{\pi}\left[\arccos(\frac{-\gamma-a}{\rho})+\frac{a+\gamma}{\rho^2}\sqrt{\rho^2-(a+\gamma)^2}\right] \tag{E.17}$$

$$P(\hat{S}=r|S=l)=\frac{1}{\pi}\left[\arccos(\frac{\gamma+a}{\rho})-\frac{a+\gamma}{\rho^2}\sqrt{\rho^2-(a+\gamma)^2}\right] \tag{E.18}$$

The current two cases share the same mathematical expression of the confusion matrix!

Furthermore, if $\rho < a, a - \rho \leq \gamma \leq a$ (Figure E.1d), we can obtain the same expression for a half of the confusion matrix, with the other half having the value of either 0 or 1.

In summary,

$$P(\hat{S} = r|S = r) = \begin{cases} 0 & \gamma > a + \rho \\ 1 & \gamma < a - \rho \\ \frac{1}{\pi}\left[\arccos(\frac{\gamma - a}{\rho}) + \frac{a-\gamma}{\rho^2}\sqrt{\rho^2 - (a-\gamma)^2}\right] & \text{else} \end{cases} \tag{E.19}$$

$$P(\hat{S} = l|S = r) = \begin{cases} 1 & \gamma > a + \rho \\ 0 & \gamma < a - \rho \\ \frac{1}{\pi}\left[\arccos(\frac{a - \gamma}{\rho}) + \frac{\gamma-a}{\rho^2}\sqrt{\rho^2 - (a-\gamma)^2}\right] & \text{else} \end{cases} \tag{E.20}$$

$$P(\hat{S} = l|S = l) = \begin{cases} 0 & \gamma < -a - \rho \\ 1 & \gamma > \rho - a \\ \frac{1}{\pi}\left[\arccos(\frac{-\gamma - a}{\rho}) + \frac{\gamma+a}{\rho^2}\sqrt{\rho^2 - (\gamma+a)^2}\right] & \text{else} \end{cases} \tag{E.21}$$

$$P(\hat{S} = r|S = l) = \begin{cases} 1 & \gamma < -a - \rho \\ 0 & \gamma > \rho - a \\ \frac{1}{\pi}\left[\arccos(\frac{\gamma + a}{\rho}) - \frac{\gamma+a}{\rho^2}\sqrt{\rho^2 - (\gamma+a)^2}\right] & \text{else} \end{cases} \tag{E.22}$$

Glossary

Amplitude-spectrum Similarity (S_A) The cosine similarity of the target and background amplitude spectra. It is a phase-independent similarity measure.

Bayesian Decision Theory (BDT) A statistical framework of decision-making that combines prior knowledge and new observation based on Bayes' theorem.

Contrast Sensitivity Function (CSF) A function that describes the sensitivity of the human visual system to gratings with different levels of spatial frequencies.

Eye-filtered Template Matching (ETM) A template matching model that accounts for the eye filtering of the target and the background.

Eye-filtered, Reliability-weighted Template Matching (ERTM) A template matching model that accounts for the eye and contrast-normalization filtering of the target and the background.

Human Visual System (HVS) The eye and the parts of the central nervous system that give humans the sense of vision.

Image Similarity (S_I) The cosine similarity of the target and background in the spatial domain. It is a phase-dependent similarity measure.

Intrinsic Position Uncertainty (IPU) The uncertainty on the target position due to internal neural noise, no matter if the target location is always fixed (without extrinsic uncertainty).

Lateral Ganglion Nucleus (LGN) A structure in the thalamus that relays visual information from the retina to the visual cortex.

Linearly Filtered Gaussian (LFG) Noise A Gaussian noise with linear filters applied to the spatial and spatial frequency domains.

Log-likelihood Ratio (LLR) The log of a ratio of likelihoods, commonly used as a decision variable.

Reliability-weighted Template Matching (RTM) A template matching model that accounts for the contrast-normalization filtering of the target and the background.

Reliability-weighted, Whitened TemplateMatching (RWTM) A template matching model that accounts for the contrast-normalization filtering and whitening (in spatial frequency) of the target and the background.

Retinal Ganglion Cell (RGC) A neuron near the vitreous border in the retina that transmits visual information.

Signal Detection Theory (SDT) A mathematical framework to quantify and classify choices for detection and discrimination.

Uncertain, Eye-filtered, Reliability-weighted Template Matching (UERTM) A template matching model that accounts for the eye and contrast-normalization filtering, and whitening (in spatial frequency) of the target and the background.

Whitened Template Matching (WTM) A template matching model that accounts for the whitening (in spatial frequency) of the target and the background.

Works Cited

[1] Wilson S. Geisler, Wilson S. Geisler, Jeffrey S. Perry, and Jeffrey S. Perry. Statistics for optimal point prediction in natural images. *Journal of Vision*, 2011. doi: 10.1167/11.12.14.

[2] David M. Green and John A. Swets. *Signal Detection Theory and Psychophysics*. Signal Detection Theory and Psychophysics. John Wiley, Oxford, England, 1966.

[3] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, July 1968. ISSN 2168-2887. doi: 10.1109/TSSC.1968.300136.

[4] Marcus E. Raichle and Debra A. Gusnard. Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10237–10239, August 2002. ISSN 0027-8424. doi: 10.1073/pnas.172399499.

[5] Ibraheem Rehman, Bita Hazhirkarzar, and Bhupendra C. Patel. Anatomy, Head and Neck, Eye. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2024.

[6] Daniel J. Felleman and David C. Van Essen. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1):1–47, January 1991. ISSN 1047-3211. doi: 10.1093/cercor/1.1.1-a.

[7] H. K. Hartline. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology-Legacy Content*, 121(2):400–415, January 1938. ISSN 0002-9513. doi: 10.1152/ajplegacy.1938.121.2.400.

[8] Stephen W. Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1):37–68, January 1953. ISSN 0022-3077. doi: 10.1152/jn.1953.16.1.37.

[9] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243, March 1968. ISSN 0022-3751.

[10] Russell L. De Valois, Duane G. Albrecht, and Lisa G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(5):545–559, January 1982. ISSN 0042-6989. doi: 10.1016/0042-6989(82)90113-4.

[11] Robert Shapley and Christina Enroth-Cugell. Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3:263–346, January 1984. ISSN 0278-4327. doi: 10.1016/0278-4327(84)90011-7.

[12] Matteo Carandini and David J. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, January 2012. ISSN 1471-0048. doi: 10.1038/nrn3136.

[13] Duane G Albrecht and Wilson S Geisler. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual neuroscience*, 7 (6):531–546, 1991.

[14] David J. Heeger. Nonlinear model of neural responses in cat visual cortex. In *Computational Models of Visual Processing*, pages 119–133. The MIT Press, Cambridge, MA, US, 1991. ISBN 978-0-262-12155-2.

[15] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual neuroscience*, 9(2):181–197, August 1992. ISSN 0952-5238. doi: 10.1017/S0952523800009640.

[16] Odelia Schwartz and Eero P. Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, August 2001. ISSN 1546-1726. doi: 10.1038/90526.

[17] Stephen Sebastian, Stephen Sebastian, Jared Abrams, Jared Abrams, Wilson S. Geisler, and Wilson S. Geisler. Constrained sampling experiments reveal principles of detection in natural scenes. *Proceedings of the National Academy of Sciences of the United States of America*, 2017. doi: 10.1073/pnas.1619487114.

[18] D.J. Tolhurst, J.A. Movshon, and A.F. Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23(8):775–785, January 1983. ISSN 00426989. doi: 10.1016/0042-6989(83)90200-6.

[19] Loren W. Nolte and David Jaarsma. More on the detection of one of M orthogonal signals. *Journal of the Acoustical Society of America*, 1967. doi: 10.1121/1.1910360.

[20] Richard G Swensson and Philip F Judy. Detection of noisy visual targets: Models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics*, 29(6):521–534, 1981.

[21] Marilyn L Shaw. Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology*, 14(3):353–409, July 1982. ISSN 0010-0285. doi: 10.1016/0010-0285(82)90014-7.

[22] Denis G. Pelli. Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of The Optical Society of America A-optics Image Science and Vision*, 1985. doi: 10.1364/josaa.2.001508.

[23] M. Michel and W. S. Geisler. Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision*, 11(1): 18–18, January 2011. ISSN 1534-7362. doi: 10.1167/11.1.18.

[24] Yelda Semizer and Melchi M. Michel. Intrinsic position uncertainty impairs overt search performance. *Journal of Vision*, 17(9):13, August 2017. ISSN 1534-7362. doi: 10.1167/17.9.13.

[25] Anqi Zhang, Eric S. Seemiller, and Wilson S. Geisler. Phase-dependent asymmetry of pattern masking in natural images explained by intrinsic position uncertainty. *Journal of Vision*, 23(10):16, September 2023. ISSN 1534-7362. doi: 10.1167/jov.23.10.16.

[26] Horace Barlow. Possible Principles Underlying the Transformations of Sensory Messages. *Sensory Communication*, 1, January 1961. ISSN 9780262518420. doi: 10.7551/mitpress/9780262518420.003.0013.

[27] Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583): 607–609, June 1996. ISSN 1476-4687. doi: 10.1038/381607a0.

[28] Wilson S. Geisler. Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59(Volume 59, 2008):167–192, 2008. ISSN 1545-2085. doi: 10.1146/annurev.psych.58.110405.085632.

[29] Wilson S. Geisler. Contributions of ideal observer theory to vision research. *Vision Research*, 51(7):771–781, April 2011. ISSN 00426989. doi: 10.1016/j.visres.2010.09.027.

[30] Jost B. Jonas, Ulrike Schneider, and Gottfried O. H. Naumann. Count and density of human retinal photoreceptors. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 230(6):505–510, October 1992. ISSN 1435-702X. doi: 10.1007/BF00181769.

[31] Robert A. Frazor and Wilson S. Geisler. Local luminance and contrast in natural images. *Vision Research*, 46(10):1585–1598, May 2006. ISSN 0042-6989. doi: 10.1016/j.visres.2005.06.038.

[32] Valerio Mante, Robert A. Frazor, Vincent Bonin, Wilson S. Geisler, and Matteo Carandini. Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience*, 8(12):1690–1697, December 2005. ISSN 1546-1726. doi: 10.1038/nn1556.

[33] Jussi T. Lindgren, Jarmo Hurri, and Aapo Hyvärinen. Spatial dependencies between local luminance and contrast in natural images. *Journal of Vision*, 8 (12):6, September 2008. ISSN 1534-7362. doi: 10.1167/8.12.6.

[34] Daniel L Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5(4):517–548, January 1994. ISSN 0954-898X. doi: 10.1088/0954-898X_5_4_006.

[35] M. G. A. Thomson. Beats, kurtosis and visual coding. *Network: Computation in Neural Systems*, 12(3):271, August 2001. ISSN 0954-898X. doi: 10.1088/0954-898X/12/3/303.

[36] David J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12):2379–2394, December 1987. ISSN 1520-8532. doi: 10.1364/JOSAA.4.002379.

[37] Alex P. Pentland. Fractal-Based Description of Natural Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):661–674, November 1984. ISSN 1939-3539. doi: 10.1109/TPAMI.1984.4767591.

[38] Jinggang Huang and D. Mumford. Statistics of natural images and models. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 1, pages 541–547 Vol. 1, June 1999. doi: 10.1109/CVPR.1999.786990.

[39] Ann B. Lee, David Mumford, and Jinggang Huang. Occlusion Models for Natural Images: A Statistical Study of a Scale-Invariant Dead Leaves Model. *International Journal of Computer Vision*, 41(1):35–59, January 2001. ISSN 1573-1405. doi: 10.1023/A:1011109015675.

[40] Martin J. Wainwright, Eero P. Simoncelli, and Alan S. Willsky. Random Cascades on Wavelet Trees and Their Use in Analyzing and Modeling Natural Images. *Applied and Computational Harmonic Analysis*, 11(1):89–123, July 2001. ISSN 1063-5203. doi: 10.1006/acha.2000.0350.

[41] Neville Drasdo, C. Leigh Millican, Charles R. Katholi, and Christine A. Curcio. The length of Henle fibers in the human retina and a model of ganglion receptive

field density in the visual field. *Vision Research*, 47(22):2901–2911, October 2007. ISSN 00426989. doi: 10.1016/j.visres.2007.01.007.

[42] Andrew B. Watson. A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision*, 14(7):15, June 2014. ISSN 1534-7362. doi: 10.1167/14.7.15.

[43] Giovanni Montesano, Giovanni Ometto, Ruth E. Hogg, Luca M. Rossetti, David F. Garway-Heath, and David P. Crabb. Revisiting the Drasdo Model: Implications for Structure-Function Analysis of the Macular Region. *Translational Vision Science & Technology*, 9(10):15, September 2020. ISSN 2164-2591. doi: 10.1167/tvst.9.10.15.

[44] Christine A. Curcio, Kenneth R. Sloan, Robert E. Kalina, and Anita E. Hendrickson. Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4):497–523, February 1990. ISSN 0021-9967, 1096-9861. doi: 10.1002/cne.902920402.

[45] Howard E. Egeth, Robert A. Virzi, and Hadley Garbart. Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, 10(1):32–39, 1984. ISSN 1939-1277. doi: 10.1037/0096-1523.10.1.32.

[46] Anne Treisman and Janet Souther. Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3):285–310, 1985. ISSN 1939-2222. doi: 10.1037/0096-3445.114.3.285.

[47] A. Treisman and S. Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1):15–48, January 1988. ISSN 0033-295X. doi: 10.1037/0033-295x.95.1.15.

[48] John Duncan and Glyn W. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96(3):433–458, 1989. ISSN 1939-1471. doi: 10.1037/0033-295X.96.3.433.

[49] Jeremy M. Wolfe, Kyle R. Cave, and Susan L. Franzel. Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3):419–433, 1989. ISSN 1939-1277. doi: 10.1037/0096-1523.15.3.419.

[50] Ignace Th.C. Hooge and Casper J. Erkelens. Adjustment of fixation duration in visual search. *Vision Research*, 38(9):1295–IN4, May 1998. ISSN 00426989. doi: 10.1016/S0042-6989(97)00287-3.

[51] Harold H Greene. The Control of Fixation Duration in Visual Search. *Perception*, 35(3):303–315, March 2006. ISSN 0301-0066. doi: 10.1068/p5329.

[52] John Palmer. Attentional effects in visual search: Relating search accuracy and search time. In *Visual Attention*, Vancouver Studies in Cognitive Science, Vol. 8., pages 348–388. Oxford University Press, New York, NY, US, 1998. ISBN 978-0-19-512692-1 978-0-19-512693-8.

[53] Jacob Nachmias and Richard V. Sansbury. Grating contrast: Discrimination may be better than detection. *Vision Research*, 14(10):1039–1042, October 1974. ISSN 0042-6989. doi: 10.1016/0042-6989(74)90175-8.

[54] F. W. Campbell and J. J. Kulikowski. Orientational selectivity of the human visual system. *The Journal of Physiology*, 187(2):437–445, 1966. ISSN 1469-7793. doi: 10.1113/jphysiol.1966.sp008101.

[55] J. Rovamo and V. Virsu. An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, 37(3):495–510, November 1979. ISSN 1432-1106. doi: 10.1007/BF00236819.

[56] Marisa Carrasco, Denise L. Evert, Irene Chang, and Svetlana M. Katz. The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, 57(8):1241–1261, November 1995. ISSN 1532-5962. doi: 10.3758/BF03208380.

[57] Marisa Carrasco and Brian McElree. Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences*, 98(9):5363–5367, April 2001. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.081074098.

[58] Jiri Najemnik and Wilson S. Geisler. Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3):4, March 2008. ISSN 1534-7362. doi: 10.1167/8.3.4.

[59] R. Calen Walshe and Wilson S. Geisler. Efficient allocation of attentional sensitivity gain in visual cortex reduces foveal sensitivity in visual search. *Current Biology*, 32(1):26–36.e6, January 2022. ISSN 09609822. doi: 10.1016/j.cub.2021.10.011.

[60] Sabine Kastner Ungerleider and Leslie G. Mechanisms of Visual Attention in the Human Cortex. *Annual Review of Neuroscience*, 23(1):315–341, March

2000. ISSN 0147-006X, 1545-4126. doi: 10.1146/annurev.neuro.23.1.315.

[61] Jeffrey Moran and Robert Desimone. Selective Attention Gates Visual Processing in the Extrastriate Cortex. *Science*, 229(4715):782–784, August 1985. doi: 10.1126/science.4023713.

[62] Pascal Fries, John H. Reynolds, Alan E. Rorie, and Robert Desimone. Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. *Science*, 291(5508):1560–1563, February 2001. doi: 10.1126/science.1055465.

[63] James W. Bisley and Michael E. Goldberg. Neuronal Activity in the Lateral Intraparietal Area and Spatial Attention. *Science*, 299(5603):81–86, January 2003. doi: 10.1126/science.1077395.

[64] D. J. Simons and C. F. Chabris. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28(9):1059–1074, 1999. ISSN 0301-0066. doi: 10.1068/p281059.

[65] Trafton Drew, Melissa L.-H. Võ, and Jeremy M. Wolfe. The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers. *Psychological Science*, 24(9):1848–1853, September 2013. ISSN 0956-7976, 1467-9280. doi: 10.1177/0956797613479386.

[66] Samuel G. Robson and Jason M. Tangen. The invisible 800-pound gorilla: Expertise can increase inattentional blindness. *Cognitive Research: Principles and Implications*, 8(1):33, May 2023. ISSN 2365-7464. doi: 10.1186/s41235-023-00486-x.

[67] M. I. Posner and Y. Cohen. Components of visual orienting. *Attention and Performance X*, 32:531–556, 1984.

[68] Raymond M. Klein. Inhibition of return. *Trends in Cognitive Sciences*, 4(4): 138–147, April 2000. ISSN 1364-6613. doi: 10.1016/S1364-6613(00)01452-2.

[69] Jiri Najemnik and Wilson S. Geisler. Simple summation rule for optimal fixation selection in visual search. *Vision Research*, 49(10):1286–1294, June 2009. ISSN 00426989. doi: 10.1016/j.visres.2008.12.005.

[70] Yunhui Zhou and Yuguo Yu. Human visual search follows a suboptimal Bayesian strategy revealed by a spatiotemporal computational model and experiment. *Communications Biology*, 4(1):34, January 2021. ISSN 2399-3642. doi: 10.1038/s42003-020-01485-0.

[71] Claus Bundesen. A theory of visual attention. *Psychological Review*, 97(4): 523–547, 1990. ISSN 1939-1471. doi: 10.1037/0033-295X.97.4.523.

[72] Claus Bundesen, Thomas Habekost, and Søren Kyllingsbæk. A Neural Theory of Visual Attention: Bridging Cognition and Neurophysiology. *Psychological Review*, 112(2):291–328, 2005. ISSN 1939-1471. doi: 10.1037/0033-295X.112.2.291.

[73] Jeremy M. Wolfe. Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, 28(4):1060–1092, February 2021. ISSN 1531-5320. doi: 10.3758/s13423-020-01859-9.

[74] R. A. Kinchla. Detecting target elements in multielement arrays: A confusability model. *Perception & Psychophysics*, 15(1):149–158, January 1974. ISSN 0031-5117, 1532-5962. doi: 10.3758/BF03205843.

[75] John Palmer, Cynthia T. Ames, and Delwin T. Lindsey. Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology:*

Human Perception and Performance, 19(1):108–130, 1993. ISSN 1939-1277. doi: 10.1037/0096-1523.19.1.108.

[76] Miguel P. Eckstein. The Lower Visual Search Efficiency for Conjunctions Is Due to Noise and not Serial Attentional Processing. *Psychological Science*, 9(2):111–118, March 1998. ISSN 0956-7976, 1467-9280. doi: 10.1111/1467-9280.00020.

[77] Jiri Najemnik and Wilson S. Geisler. Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–391, March 2005. ISSN 1476-4687. doi: 10.1038/nature03390.

[78] M. P. Eckstein. Visual search: A retrospective. *Journal of Vision*, 11(5): 14–14, December 2011. ISSN 1534-7362. doi: 10.1167/11.5.14.

[79] Stephen Sebastian, Stephen Sebastian, Eric Seemiller, Eric Seemiller, Wilson S. Geisler, and Wilson S. Geisler. Local reliability weighting explains identification of partially masked objects in natural images. *Proceedings of the National Academy of Sciences of the United States of America*, 2020. doi: 10.1073/pnas.1912331117.

[80] Anqi Zhang and Wilson S. Geisler. Detection of targets in filtered noise: Whitening in space and spatial frequency. *JOSA A*, 39(4):690–701, April 2022. ISSN 1520-8532. doi: 10.1364/JOSAA.447391.

[81] G. J. Burton and Ian R. Moorhead. Color and spatial structure in natural scenes. *Applied Optics*, 26(1):157–170, January 1987. ISSN 2155-3165. doi: 10.1364/AO.26.000157.

[82] J. H. van Hateren. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A*, 171(2):157–170, September 1992. ISSN 1432-1351. doi: 10.1007/BF00188924.

[83] D. J. Tolhurst, Y. Tadmor, and Tang Chao. Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, 12(2):229–232, 1992. ISSN 1475-1313. doi: 10.1111/j.1475-1313.1992.tb00296.x.

[84] Dawei W Dong and Joseph J Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–358, January 1995. ISSN 0954-898X. doi: 10.1088/0954-898X_6_3_003.

[85] A. van der Schaaf and J. H. van Hateren. Modelling the Power Spectra of Natural Images: Statistics and Information. *Vision Research*, 36(17):2759–2770, September 1996. ISSN 0042-6989. doi: 10.1016/0042-6989(96)00002-8.

[86] David J. Field and Nuala Brady. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37(23): 3367–3383, December 1997. ISSN 0042-6989. doi: 10.1016/S0042-6989(97)00181-8.

[87] Mitchell G. A. Thomson and David H. Foster. Role of second- and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America A*, 14(9):2081, September 1997. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.14.002081.

[88] Michael A. Webster and Eriko Miyahara. Contrast adaptation and the spatial structure of natural images. *JOSA A*, 14(9):2355–2366, September 1997. ISSN 1520-8532. doi: 10.1364/JOSAA.14.002355.

[89] C. A. Párraga, G. Brelstaff, T. Troscianko, and I. R. Moorehead. Color and luminance information in natural scenes. *Journal of the Optical Society of America A*, 15(3):563, March 1998. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.15.000563.

[90] Erik Reinhard, Peter Shirley, and Tom Troscianko. Natural Image Statistics for Computer Graphics. *Univ. Utah Tech Report*, 2001.

[91] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3):391, May 2003. ISSN 0954-898X. doi: 10.1088/0954-898X/14/3/302.

[92] Ron O. Dror, Alan S. Willsky, and Edward H. Adelson. Statistical characterization of real-world illumination. *Journal of Vision*, 4(9):11, September 2004. ISSN 1534-7362. doi: 10.1167/4.9.11.

[93] J. P. Rolland and H. H. Barrett. Effect of random background inhomogeneity on observer detection performance. *Journal of the Optical Society of America A*, 9(5):649, May 1992. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.9.000649.

[94] A. E. Burgess. Statistically defined backgrounds:performance of a modified nonprewhitening observer model. *JOSA A*, 11(4):1237–1242, April 1994. ISSN 1520-8532. doi: 10.1364/JOSAA.11.001237.

[95] Arthur E. Burgess and Arthur E. Burgess. Comparison of non-prewhitening and Hotelling observer models. *null*, 1995. doi: 10.1117/12.206837.

[96] Ramona W. Bouwman, Ramona W. Bouwman, Ruben E. van Engen, R E van Engen, R.E. van Engen, Mireille J. M. Broeders, Mireille J. M. Broeders, Ger-

ard J. den Heeten, G. J. den Heeten, David R. Dance, David R. Dance, David R. Dance, Kenneth C. Young, Kenneth C. Young, Wouter J. H. Veldkamp, and Wouter J. H. Veldkamp. Can the non-pre-whitening model observer, including aspects of the human visual system, predict human observer performance in mammography? *Physica Medica*, 2016. doi: 10.1016/j.ejmp.2016.11.109.

[97] Arthur E. Burgess and Arthur Burgess. Signal detection in radiology. *null*, 2018. doi: 10.1017/9781108163781.005.

[98] A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow. Efficiency of Human Visual Signal Discrimination. *Science*, 214(4516):93–94, October 1981. doi: 10.1126/science.7280685.

[99] Darren North and D.O. North. An Analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems. *null*, 1963. doi: 10.1109/proc.1963.2383.

[100] R. Brunelli and T. Poggiot. Template matching: Matched spatial filters and beyond. *Pattern Recognition*, 30(5):751–768, May 1997. ISSN 0031-3203. doi: 10.1016/S0031-3203(96)00104-5.

[101] David H. Brainard. The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–436, January 1997. ISSN 0169-1015, 1568-5683. doi: 10.1163/156856897X00357.

[102] Denis G. Pelli. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4):437–442, January 1997. ISSN 0169-1015, 1568-5683. doi: 10.1163/156856897X00366.

[103] Arthur E. Burgess. Visual signal detection with two-component noise: Low-pass spectrum effects. *Journal of the Optical Society of America A*, 16(3):694, March 1999. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.16.000694.

[104] Craig K. Abbey and Miguel P. Eckstein. Classification images for simple detection and discrimination tasks in correlated noise. *JOSA A*, 24(12):B110–B124, December 2007. ISSN 1520-8532. doi: 10.1364/JOSAA.24.00B110.

[105] Chris Bradley, Chris Bradley, Jared Abrams, Jared Abrams, Wilson S. Geisler, and Wilson S. Geisler. Retina-V1 model of detectability across the visual field. *Journal of Vision*, 2014. doi: 10.1167/14.12.22.

[106] Andrew B. Watson, Andrew B. Watson, Albert J. Ahumada, and Albert J. Ahumada. A standard model for foveal detection of spatial contrast. *Journal of Vision*, 2005. doi: 10.1167/5.9.6.

[107] Andrew B. Watson. A formula for the mean human optical modulation transfer function as a function of pupil size. *Journal of Vision*, 13(6):18, May 2013. ISSN 1534-7362. doi: 10.1167/13.6.18.

[108] Arthur E. Burgess. Visual Perception Studies and Observer Models in Medical Imaging. *Seminars in Nuclear Medicine*, 41(6):419–436, November 2011. ISSN 0001-2998. doi: 10.1053/j.semnuclmed.2011.06.005.

[109] Yani Zhang, Craig K. Abbey, and Miguel P. Eckstein. Adaptive detection mechanisms in globally statistically nonstationary-oriented noise. *Journal of the Optical Society of America A*, 23(7):1549, July 2006. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.23.001549.

[110] Harrison H Barrett, Jie Yao, JANNICK P ROLLANDt, and KYLE J MYERSt. Model observers for assessment of image quality. *Proc. Natl. Acad. Sci. USA*, 1993. doi: 10.1073/pnas.90.21.9758.

[111] Miguel P. Eckstein, Craig K. Abbey, and François O. Bochud. A Practical Guide to Model Observers for Visual Detection in Synthetic and Natural Noisy Images. *SPIE*, 1:593–628, 2000.

[112] François O. Bochud, Craig K. Abbey, and Miguel P. Eckstein. Visual signal detection in structured backgrounds III Calculation of figures of merit for model observers in statistically nonstationary backgrounds. *Journal of the Optical Society of America A*, 17(2):193, February 2000. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.17.000193.

[113] Arthur E Burgess. Mammographic structure: Data preparation and spatial statistics analysis. In *Medical Imaging 1999: Image Processing*, volume 3661, pages 642–653. SPIE, 1999.

[114] John J. Heine and Robert P. Velthuizen. Spectral analysis of full field digital mammography data. *Medical Physics*, 29(5):647–661, May 2002. ISSN 0094-2405. doi: 10.1118/1.1445410.

[115] Ingrid Reiser and Robert M. Nishikawa. Identification of simulated microcalcifications in white noise and mammographic backgrounds. *Medical Physics*, 33 (8):2905–2911, 2006. ISSN 2473-4209. doi: 10.1118/1.2210566.

[116] Hui Li, Maryellen L Giger, Olufunmilayo I Olopade, and Michael R Chinander. Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment. *Journal of digital imaging*, 21:145–152, 2008.

[117] Kathrine G. Metheany, Craig K. Abbey, Nathan Packard, and John M. Boone. Characterizing anatomical variability in breast CT images. *Medical Physics*, 35(10):4685–4694, 2008. ISSN 2473-4209. doi: 10.1118/1.2977772.

[118] Emma Engstrom, Ingrid Reiser, and Robert Nishikawa. Comparison of power spectra for tomosynthesis projections and reconstructed images. *Medical Physics*, 36(5):1753–1758, 2009. ISSN 2473-4209. doi: 10.1118/1.3116774.

[119] K. Bliznakova, S. Suryanarayanan, A. Karellas, and N. Pallikarakis. Evaluation of an improved algorithm for producing realistic 3D breast software phantoms: Application for mammography. *Medical Physics*, 37(11):5604–5617, 2010. ISSN 2473-4209. doi: 10.1118/1.3491812.

[120] Beverly A Lau, Ingrid Reiser, and Robert M Nishikawa. Issues in characterizing anatomic structure in digital breast tomosynthesis. In *Medical Imaging 2011: Physics of Medical Imaging*, volume 7961, pages 331–338. SPIE, 2011.

[121] Lin Chen, John M Boone, Anita Nosratieh, and Craig K Abbey. NPS comparison of anatomical noise characteristics in mammography, tomosynthesis, and breast CT images using power law metrics. In *Medical Imaging 2011: Physics of Medical Imaging*, volume 7961, pages 131–134. SPIE, 2011.

[122] Terry Caelli and Giampaolo Moraglia. On the detection of signals embedded in natural scenes. *Attention Perception & Psychophysics*, 1986. doi: 10.3758/bf03211490.

[123] Ann Marie Rohaly, Albert J. Ahumada, and Andrew B. Watson. Object detection in natural backgrounds predicted by discrimination performance and models. *Vision Research*, 1997. doi: 10.1016/s0042-6989(97)00156-9.

[124] Damon M. Chandler, Matthew Gaubatz, and Sheila S. Hemami. A patch-based structural masking model with an application to compression. *Eurasip Journal on Image and Video Processing*, 2009. doi: 10.1155/2009/649316.

[125] Mushfiqul Alam, Kedarnath P. Vilankar, David J. Field, Damon M. Chandler, and Damon M. Chandler. Local masking in natural images: A database and analysis. *Journal of Vision*, 2014. doi: 10.1167/14.8.22.

[126] Carlos Dorronsoro, Carlos Dorronsoro, Carlos Dorronsoro, Calen Walshe, Calen Walshe, Steve Sebastian, Steve Sebastian, Wilson S. Geisler, and Wilson S. Geisler. Separable effects of similarity and contrast on detection in natural backgrounds. *Journal of Vision*, 2018. doi: 10.1167/18.10.747.

[127] Reuben Rideaux, Rebecca K West, Thomas S A Wallis, Peter J Bex, Jason B Mattingley, and William J Harrison. Spatial structure, phase, and the contrast of natural images. *Journal of Vision*, 2022. doi: 10.1167/jov.22.1.4.

[128] R. Calen Walshe, R. Calen Walshe, Wilson S. Geisler, and Wilson S. Geisler. Detection of occluding targets in natural backgrounds. *Journal of Vision*, 2020. doi: 10.1167/jov.20.13.14.

[129] Gordon E. Legge and F. W. Campbell. Displacement detection in human vision. *Vision Research*, 1981. doi: 10.1016/0042-6989(81)90114-0.

[130] Edward H. Adelson and James R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of The Optical Society of America A-optics Image Science and Vision*, 1985. doi: 10.1364/josaa.2.000284.

[131] Anqi Zhang and Wilson S. Geisler. Optimal Visual Search with Highly Heuristic Decision Rules, September 2024.

[132] Miguel P. Eckstein, James P. Thomas, John Palmer, and Steven S. Shimozaki. A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, 62(3):425–451, April 2000. ISSN 0031-5117, 1532-5962. doi: 10.3758/BF03212096.

[133] John Palmer and Elizabeth Davis. Visual search and attention: An overview. *Spatial Vision*, 17(4-5):249–255, January 2004. doi: 10.1163/1568568041920168.

[134] Marisa Carrasco. Visual attention: The past 25 years. *Vision Research*, 51 (13):1484–1525, July 2011. ISSN 00426989. doi: 10.1016/j.visres.2011.04.012.

[135] Mark W. Cannon. Contrast perception in the peripheral visual field. In *Annual Meeting Optical Society of America (1985), Paper WJ41.* Optica Publishing Group, October 1985. doi: 10.1364/OAM.1985.WJ41.

[136] Eli Peli, Jian Yang, and Robert B. Goldstein. Image invariance with changes in size: The role of peripheral contrast thresholds. *JOSA A, Vol. 8, Issue 11, pp. 1762-1774*, November 1991. doi: 10.1364/JOSAA.8.001762.

[137] Daniel L. Adams, Lawrence C. Sincich, and Jonathan C. Horton. Complete Pattern of Ocular Dominance Columns in Human Primary Visual Cortex. *Journal of Neuroscience*, 27(39):10391–10403, September 2007. doi: 10.1523/JNEUROSCI.2923-07.2007.

[138] Zhong-Lin Lu and Barbara Anne Dosher. Characterizing human perceptual inefficiencies with equivalent internal noise. *JOSA A, Vol. 16, Issue 3, pp. 764-778*, March 1999. doi: 10.1364/JOSAA.16.000764.

[139] Preeti Verghese. Active search for multiple targets is inefficient. *Vision Research*, 74:61–71, December 2012. ISSN 00426989. doi: 10.1016/j.visres.2012.08.008.

[140] Adam Kohn, Amin Zandvakili, and Matthew A Smith. Correlations and brain states: From electrophysiology to functional imaging. *Current Opinion in Neurobiology*, 19(4):434–438, August 2009. ISSN 0959-4388. doi: 10.1016/j.conb.2009.06.007.

[141] Marlene R. Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819, July 2011. ISSN 1546-1726. doi: 10.1038/nn.2842.

[142] Adam Kohn, Ruben Coen-Cagli, Ingmar Kanitscheider, and Alexandre Pouget. Correlations and Neuronal Population Information. *Annual review of neuroscience*, 39:237–256, July 2016. ISSN 0147-006X. doi: 10.1146/annurev-neuro-070815-013851.

[143] Robert Rosenbaum, Matthew A. Smith, Adam Kohn, Jonathan E. Rubin, and Brent Doiron. The spatial structure of correlated neuronal variability. *Nature Neuroscience*, 20(1):107–114, January 2017. ISSN 1546-1726. doi: 10.1038/nn.4433.

[144] Benjamin R. Cowley, Adam C. Snyder, Katerina Acar, Ryan C. Williamson, Byron M. Yu, and Matthew A. Smith. Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron*, 2020. doi: 10.1016/j.neuron.2020.07.021.

[145] Herbert Alexander Simon. *Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in Society Setting*. Wiley, 1957.

[146] Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic Decision Making. *Annual Review of Psychology*, 62(1):451–482, January 2011. ISSN 0066-4308, 1545-2085. doi: 10.1146/annurev-psych-120709-145346.

[147] Leonard J. Savage. *The Foundations of Statistics*. Wiley, 1954.

[148] Robbe L. T. Goris, J. Anthony Movshon, and Eero P. Simoncelli. Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–865, June 2014. ISSN 1546-1726. doi: 10.1038/nn.3711.

[149] Michael A. Bruno, Eric A. Walker, and Hani H. Abujudeh. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. *RadioGraphics*, 35(6):1668–1676, October 2015. ISSN 0271-5333. doi: 10.1148/rg.2015150023.

[150] Andrew J. Degnan, Emily H. Ghobadi, Peter Hardy, Elizabeth Krupinski, Elena P. Scali, Lindsay Stratchko, Adam Ulano, Eric Walker, Ashish P. Wasnik, and William F. Auffermann. Perceptual and Interpretive Error in Diagnostic Radiology—Causes and Potential Solutions. *Academic Radiology*, 26(6):833–845, June 2019. ISSN 1076-6332. doi: 10.1016/j.acra.2018.11.006.