VISUAL DECISION MAKING IN THE PRESENCE

OF STIMULUS AND MEASUREMENT

CORRELATIONS

―――――――――

A Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

―――――――――

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

―――――――――

By

Manisha Bhardwaj

December 2013

VISUAL DECISION MAKING IN THE PRESENCE OF STIMULUS AND MEASUREMENT CORRELATIONS

Manisha Bhardwaj

APPROVED:

Dr. Krešimir Josić   (Committee Chair)
Department of Mathematics, University of Houston

Dr. Matthew Nicol
Department of Mathematics, University of Houston

Dr. Ilya Timofeyev
Department of Mathematics, University of Houston

Dr. Wei Ji Ma
Center for Neural Science and
Department of Psychology, New York University

_____

_____
Dean, College of Natural Sciences and Mathematics

Acknowledgements

I owe my deepest gratitude and intellectual debt to my advisor, Dr. Krešimir Josić, for his constant thrust of guidance, support, and patience for me. This dissertation would have been impossible without his enthusiastic encouragement and persistent help. His insightful comments, and unsurpassed knowledge contributed enormously in shaping my scientific and mathematical thoughts. Frequent and regular hours of discussions with him have deeply influenced, and nurtured my academic and personal career. I am particularly grateful to him for his patient assistance in developing my technical writing and presentation skills. The financial support provided by him, and graciously offered several car rides are greatly acknowledged.

I extend my profound thanks to Dr. Wei Ji Ma for providing his insightful contributions and critical recommendations for the projects that are part of this dissertation. His useful critiques, meticulous comments, and area expertise greatly benefited the quality of my research work. His upcoming book, and the course notes have served as excellent references for the introduction chapter in the dissertation. I am also thankful to him for offering the laboratory resources in running the experiments that are part of the dissertation.

My committee members Dr. Matthew Nicol and Dr. Ilya Timofeyev deserve a special note of thanks, for generously serving on my defense committee, proofreading this dissertation, and providing useful suggestions for its improvement.

To my parents Satvir Singh Bhardwaj and Pushpa Rani Bhardwaj

VISUAL DECISION MAKING IN THE PRESENCE OF STIMULUS AND MEASUREMENT CORRELATIONS

———————————

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

———————————

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

———————————

By

Manisha Bhardwaj

December 2013

Abstract

Our brains process sensory information to infer the state of the world. However, the input from our senses is noisy, which may lead to errors in perceptual judgements. A number of theoretical studies have modeled perception as a process of probabilistic inference which involves making decisions based on uncertain evidence. Bayesian optimality is a general principle of probabilistic inference that has been successfully used to build quantitative models of perception. In addition, several experimental studies show that human behavior is close to Bayes-optimal, in the sense that humans make best possible decisions given the uncertain sensory input on various visual perceptual tasks such as visual search, sameness judgement, and change detection. However, the impact of structured stimuli on decision-making remains largely unexplored. Moreover, the sensory measurements can themselves be strongly correlated to produce a structured representation of the stimulus input. These measurement correlations can interact with the structure of the external input in many possible ways, and should not be considered in isolation.

In this work, we focus on visual search task to examine how visual perception is affected by structured input. We also theoretically analyze the impact of measurement correlations on the decisions of an ideal observer. We analyze the responses of subjects on a target detection experiment where the stimulus orientations were generated with varying strength of correlations across different experimental sessions. We fit several models to the experimental data using maximum-likelihood parameter estimation. We use rigorous model selection to determine

how human observers take into account stimulus correlations in detecting a target. We subsequently describe how the relation between measurement, and stimulus correlations effects the performance of an ideal Bayesian observer in a family of target detection tasks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Our decisions are based on sensory measurements, and prior knowledge of the surroundings. The physical observations made by our eyes, ears, skin, and other sensory organs are transmitted to the brain. The brain integrates, and interprets this information to draw inferences about the state of the world.

However, our sensory observations are typically incomplete, and imperfect. They may not always reflect the true state of the world. Noisy, and imprecise measurements can be difficult to interpret, and may lead to incorrect inferences. Thus, the question arises: how does the brain infer the state of the world from the inadequate, and uncertain sensory observations? Theoretical neuroscientists have hypothesized that our brain performs specific probabilistic computations to process the partially informative observations, and makes an inference about the state of the world [65, 76, 150, 114, 86, 89, 96, 78] . Thus, perception can be thought of as a form of probabilistic inference [145, 61, 32]. The results of several experimental

studies indicate that the brain interprets sensory information probabilistically to make the best possible guess about the state of the world [40, 64, 10, 79, 80, 50, 51, 90, 142, 94, 71, 119]. These studies suggest that the brain computes the probability of many interpretations, and chooses the one with that is most likely. Such a strategy of framing the best possible perceptual inference is known as *optimal probabilistic inference* or *Bayesian inference*. The Bayesian theory of perception is based on the assumption that the brain finds the option that has the maximum probability of being correct given incomplete, and imperfect sensory information.

Earlier studies provide concrete evidence that humans, and other animals perform probabilistic inference in a number of idealized situations. However, it is not always clear how these results translate to more realistic situations. In general, we can expect that Bayesian models describing human behavior can be fairly complex, and may require extensive, and elaborate analytic computations. Often our observation time is short, and we need to make decisions in a short time. Further, most of our decisions involve processing of information from multiple sources. Given the constraints of biologically feasible computations, does our brain really make the best possible use of the information? If not, does it use some approximate strategies - and if so, when do such strategies fail? What are the possible models that best describe the computations performed by the brain?

To examine these questions, we study human behavior in an experimental setting (Chapter 2). The purpose of this study is to determine whether humans behave optimally in a fairly difficult perceptual task. We develop the theory for an

optimal Bayesian observer (Section 2.2), and provide several alternative suboptimal models (Chapter 4) that could possibly explain the observer's responses. These models encompass a range of assumptions about the observer's behavior. We compare these models (Chapter 5) using Bayesian, and Akaike model comparisons (described in Chapter 3) to find the model that is most consistent with the experimental data.

Further, we theoretically analyze the performance of an ideal observer on a family of target detection tasks (Chapters 6 and 7). We present our analysis under certain assumptions about the parameters that determine the external structure of the task, and those that govern the structure of observer's measurements.

In this chapter, we give an overview of the key concepts of Bayesian inference. We begin by describing the different components of Bayes' rule, and how they can be interpreted in a perceptual task. We then describe the fundamentals of Bayesian modeling, and their applications to psychophysical studies. Further, we provide the detailed explanation of Bayesian modeling using a simple target detection example. We conclude with a summary of the work presented in subsequent chapters of the dissertation.

1.1 Perception as Bayesian inference

Several experimental studies have provided evidence for Bayesian inference in perception [50, 40, 64, 80, 10, 79, 51, 90, 142, 95, 71, 141]. These studies also suggest that humans are capable of optimally integrating the available information. For instance, the experiments performed by Ernst and Banks [40] have provided an evidence that human observers can optimally combine visual, and haptic information to make a decision about the height of a raised object. More recently, several studies have established that humans are near optimal in finding a target among distractors [90, 94, 95]. The Bayesian framework presented in [74] indicates that humans not only integrate information based on the content present in the stimuli, but also based on the relevance of the task. These, and many other similar studies are designed to investigate whether human behavior can be described in probabilistic terms.

Probabilistic computations can be helpful in routine life activities. For example, we try to predict the possibility of rain on a cloudy day based on weather conditions. We use information of ongoing vehicle, and pedestrian traffic to make a decision about crossing a busy road. These activities require us to integrate different sensory information, and make an informed decision. Generally, there is more than one possible choice or decision. Evidence suggests that we assign probability to the different options, and base our decisions on them [79, 40, 89, 50]. According to Bayesian theory, an optimal observer computes probabilities for each possible event given the available information, and makes a decision based on the most

probable state of the world. The theory relies on the assumption that the observer achieves this using Bayes' computation [12, 84].

1.1.1 Bayes' theorem

In the Bayesian framework, Bayes' theorem is used to compute the subjective belief about the state of the world based on accumulated evidence. The theorem depends on the computation of *conditional probabilities*. Conditional probabilities reflect the directional dependence between two events. For example, if *A* corresponds to the event of rain on a particular day, and *B* to the presence of clouds in the sky, then $P(A|B)$ indicates the probability that it will rain *given* a cloudy sky. We note that in general, $P(A|B) \neq P(B|A)$, and the two probabilities have different interpretations.

The Bayesian theory relies on the assumption that the brain combines the sensory measurements with our prior belief of the world via Bayes' formula,

$$
\begin{aligned}
P(\text{world state}|\text{data}) &= \frac{P(\text{data}|\text{world state})P(\text{world state})}{P(\text{data})} \\
&= \frac{P(\text{data}|\text{world state})P(\text{world state})}{\sum_k^{\text{Total states}} P(\text{data}|k^{\text{th}}\text{world state})P(k^{\text{th}}\text{world state})}.
\end{aligned} \tag{1.1}
$$

According to the Bayesian model of perception, we infer the probability of the world state given our sensory information, $P(\text{world state}|\text{data})$. We do so by using the probability of making a sensory measurement given a particular state of the world, $P(\text{data}|\text{world state})$. Importantly, we assume that this second probability is known to the observer, and is part of their of perceptions model thereof.

The denominator in Eq. (1.1) is a normalization constant, and ensures that the sum of the posterior probabilities of different world states is one. Frequently, we are interested in the expression in the numerator which establishes a direct proportional relation of the inferred probability to the sensory measurements, and the associated prior belief. Thus, we often consider the following unnormalized version of Bayes' equation,

$$P(\text{world state}|\text{data}) \propto P(\text{data}|\text{world state})P(\text{world state}). \qquad (1.2)$$

We note that given a particular observation, here "world state" is the variable, and "data" is a constant. This formula combines our prior belief with available evidence to infer the state of the world. In the following section, we provide an intuitive interpretation of each term in the Bayes' formula. We follow the ideas, and examples presented in [89].

1.2 Elements of Bayes' theorem

The Bayes' theorem consists of the following components: the likelihood function, the prior probability distribution, and the posterior probability distribution function.

The likelihood Function

Formally, $P(\text{data}|\text{world state})$ is written as $L(\text{world state}|\text{data})$, and is known as the *likelihood function* over possible world states given the sensory data. We note

that this is a function of the world state. It is not a probability distribution over the world states. Instead, it represents the likelihood of each state of the world given available data, and summarizes the degree to which the sensory data favor one world state interpretation over the other. The accuracy of this selection is dependent on the quality of the observation.

The shape and nature of the likelihood function depend on the quality of the sensory input. The function has a clearly defined peak in case of high quality sensory input. It is usually flat, or has multiple peaks when the information is inadequate or ambiguous. As an example, consider that we are walking on a street on a foggy day. Our visual information about incoming vehicle will be less accurate, and informative given unclear visible conditions. However, the sound, and moving vehicle noise can be informative in such a case. Therefore, our likelihood function here will be concentrated around the world state favored by our auditory sensory information rather than visual one.

In principle, there are many factors that could affect the likelihood function. Weather, distance, and other physical conditions can influence the sensory measurements, and hence the shape of the likelihood function. Moreover, the quality of sensory measurement varies across observers. Visual, auditory, and other sensory capabilities can impact the quality of an observer's sensory measurement, and hence the likelihood function.

The prior probability distribution

Prior knowledge about the world, $P(\text{world state})$ has significant impact on our inference process. It represents our beliefs or expectations about the world, and determines the probability of each hypothesized state. Similar to the likelihood function, prior probability varies over hypothesized world states. Prior knowledge can evolve over time as we gather new information. Prior belief is subjective as it depends on the observer's experiences. Hence, each observer may have a different prior distribution.

In our previous example, we can use our prior knowledge about the street, and traffic conditions there while walking under unclear visible conditions. If the street has a curve or possible diversions, we can combine the prior information about such conditions with our current sensory observation while walking.

The posterior probability distribution

Decisions, and inferences are based on the *posterior probability* of the world state, denoted by $P(\text{world state}|\text{data})$. The *posterior probability distribution function* represents the probability of each possible world state given our observational data, and prior beliefs. To compute this probability, we use the likelihood function described above. Since the posterior probability is a combination of the likelihood function, and prior information, the nature of the posterior distribution depends on these two factors. A sharp, and peaked likelihood function results in a peaked, and informative posterior probability distribution. On the other hand, in case of

flat likelihoods, posterior distribution resembles the prior distribution function, and the observer does not gain any new information from sensory data.

Continuing with our example of walking on a street under foggy conditions, the likelihood function of a car approaching us could have a broader shape since our visual information is less informative. In such a case, a prior information about the traffic conditions on the street could be helpful in making a decision.

1.3 Bayesian modeling of perception

In this section, we discuss in detail the mathematical modeling of perception using Bayesian framework. Bayesian methods have been used to describe the process of perceptual inference, and explain decisions of humans, and animals on simple tasks [40, 90, 142, 64, 10, 79, 80]. Bayesian theory assumes that humans use Bayesian inference to update their belief about the state of the world. They do so by updating the posterior probability based on new sensory information. Also, this inference process is continuous, and iterative since we incorporate our current state belief as prior information in making a new decision. There are numerous examples of routine life activities that can be explained using Bayesian inference. We discuss some of them below.

1.3.1 Visual and auditory perception

Visual and auditory perception are commonly studied examples of Bayesian inference. Recent experimental studies have focused on understanding visual perception through Bayesian modeling [40, 10, 119, 74, 135, 90, 142, 95, 34, 38, 39, 107, 7, 143]. These studies examine human behavior on simple visual tasks. We discuss some examples described further in [89].

Visual perception is of utmost importance to humans. We process multiple pieces of information contained in visual scenes to make decisions about the state of the world. One example is recognizing a friend in a crowd. If an observer is trying to find a friend in a large crowd from a distance, the visual information will have some degree of uncertainty. An ideal observer would compute a likelihood that each person in the crowd is the friend. As the observer gets closer to the friend, the quality of the sensory data improves, and the likelihood function gets more peaked around the friend. Further, the observer uses some prior knowledge about the friend - for example, if the friend likes to wear black, then people wearing black will be assigned larger prior probability.

Our brain also handles tremendous amount of auditory information every day. We are exposed to numerous types of sounds, noise, and music in our environment, such as, music play, phone ring, alarm sound, vehicle horn, human speech, etc. Similar to visual perception, our auditory perception can be described as a process of Bayesian inference. For instance, when a song is played, we may try to guess its name. An ideal observer would compute a likelihood function over all

known songs. The observer would also use prior information about the songs, or the type of music played by a radio station. In that case, the posterior distribution will be concentrated on frequently played songs.

These examples show how Bayesian computations could explain decisions in our daily routines. We now discuss how to design experiments to investigate whether human observers do employ Bayesian computations, and how we model the collected data on the experiment. Here we give a brief description of *psychophysical tasks* that are extensively used to study perceptual behavior. We also illustrate the mathematical steps involved in the Bayesian modeling of such perceptual tasks.

1.3.2 Psychophysical studies

Psychophysical studies are used to analyze how animals process information from the physical world. These tasks are frequently designed to understand how animals integrate information contained in the stimuli to make a decision. The difficulty of these tasks usually depends on the characteristics of the presented stimuli. For example, an observer may be asked to discriminate whether a line is to the left or right of vertical. If the line is really close to vertical, the task can be difficult.

Observers are usually required to perform many iterations (*trials*) of the task. The trials can differ from each other if the characteristics of the stimuli are randomly chosen on each trial. The recorded responses of the observer can then be analyzed using Bayesian models. A wide range of psychophysical experiments

have been analyzed using Bayesian inference approach [79, 80, 10, 40, 141, 50, 51, 94, 74, 135, 90, 64, 142, 71, 95] .

Bayesian modeling of perception consists of the following three steps:

1. describing the generative model,

2. specifying the inference process, and

3. computing the observer's estimate distributions.

These steps are the structural components of any perceptual Bayesian model, and characterize the behavior of a model observer or a *subject* on the task. Below, we discuss each step in detail.

1.3.3   Step 1: the generative model

The generative model is a probabilistic model that describes the generation of the observer's sensory data. It mathematically describes the complete structure of the task. It is a forward directed graph, with nodes representing the random variables characterized in the design of the task, and the directed edges indicate the dependencies between variables. Each node has an associated probability distribution, and the directed edge determines the influence of one variable on another, which is expressed in terms of *conditional probabilities*. At least one node in the model corresponds to the variable of interest describing the state of the world, and another variable is the observer's sensory data. The observer infers the latent variable of the world state from the stimulus shown in the task.

We denote the feature, or characteristic of a stimulus by s. The feature is sometimes itself referred to as the stimulus. Depending on the problem, the variable of interest, which we denote by W, could be different from the stimulus, s itself. It is frequently assumed that the observer makes a noisy measurement, x, of the presented stimulus. The measurement is also sometimes referred as the *internal representation* or the observation of the stimulus. Thus, the generative model contains the variable of interest (if different from the stimulus), the stimulus, and its measurement. Figure 1.1(A) shows the graphical representation of the generative model, and the probability distributions associated with each node.

1.3.3.1 Distributions in the generative model

The probability distributions in the generative model can be determined from the experimental design of the task. The *world state distribution* or prior distribution, denoted by $p(W)$ represents the distribution of probabilities over the world states in absence of any sensory information. This distribution could either be discrete or continuous depending on the associated random variable. The *stimulus distribution* is a function of the world state variable, W, and is denoted by $p(s|W)$. This distribution is completely specified by the design of the experiment. When a stimulus itself represents the state of the world, we have $p(s|W) = p(s)$.

Measurements or sensory data are usually noisy. This noise could come from many sources: the random variability due to intrinsic stochastic processes, limitations of our sensing capability, and other unknown sources. We need to make an assumption about the noise in the generative model. Even in response to the

(A) W $p(W)$ **(B)** s $p(s|W)$

s $p(s|W)$ x $p(x|s)$

x $p(x|s)$ $\hat{W}$ $p(W|x)$
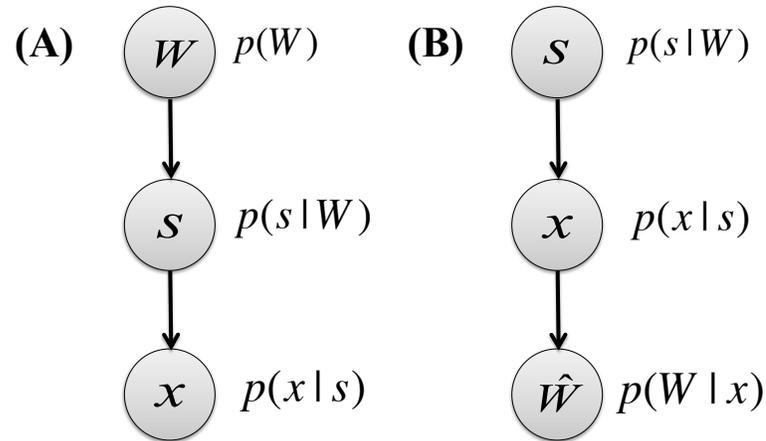
Figure 1.1: **Bayesian modeling of a perceptual task - defining the generative model, and deriving the inference process. (A) The generative model.** The first step in Bayesian modeling is to define the generative model. This figure outlines the graphical representation of the generative model we will be using throughout the dissertation. The nodes represent the variables involved in the task, and arrows determine the influence of one node on another. This influence is mathematically described in terms of conditional probabilities. The observer infers the (hidden) state of the world, W from the stimulus, s presented in the task by making a measurement, x of the stimulus. **(B) Inference process.** The second step in Bayesian modeling is to derive the inference process of an observer. That is, to understand the mathematical process by which the observer infers W based on the measurement, x. This step involves inverting the generative model, and marginalizing over intermediate variable, s to compute a decision criterion, and making a choice, $\hat{W}$ about the state of the world.

same stimulus, the observer's measurements vary randomly over the course of experimental trials. The distribution of the measurement, x given the stimulus, s, is a conditional distribution, denoted by $p(x|s)$. It represents the probability with which a stimulus results in a measurement, x. Frequently, we use the Gaussian distribution to model measurement noise,

$$p(x|s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-s)^2}{2\sigma^2}}. \tag{1.3}$$

The standard deviation, σ, of the Gaussian function reflects the uncertainty or noise in the measurement. A higher (lower) value of σ reflects the low (high) quality of the measurement, and is associated with a wider (narrower) measurement distribution. The inverse of the variance, $\frac{1}{\sigma^2}$, is commonly known as the *precision*, or *reliability* of the measurement.

1.3.3.2   Used prior distributions in the experiment

The world state distribution reflects an observer's prior belief about the state of the world. Subjects can make incorrect assumptions about the prior state of the world. Given the set up of a psychophysics experiment, it might be difficult for an observer to correctly determine the world state distribution. Subjects could use a prior based on the experiences in natural world, but this could potentially be very different from the experimental world state distribution. For example, subjects could have a higher prior probability for vertically, and horizontally aligned objects based on natural experiences while the objects in an experiment could have an equal probability for any orientation. Also, a prior of light coming from above

would be stronger than any other direction for similar reasons. Priors based on common experiences are more likely to have higher probability. Therefore, sensible assumptions need to be made about a subject's prior in the experiment. In most practical cases, a flat or uniform prior is assumed that assigns equal probability to each outcome for the state of the world, that is, $p(W) = $ constant. Such a prior simplifies the inference computations. Alternatively, we can also determine the subject's prior from experimental data. This practice is commonly used in the case of binary world state variables.

In summary, the distributions $p(W)$, $p(s|W)$, and $p(x|s)$ completely define the generative model of the task, and constitute a major component of Bayesian modeling of perceptual inference.

1.3.4   Step 2: the inference or perception process

In the next step of Bayesian modeling, we specify a model to determine how an ideal observer makes decisions. In our computations, we use certain assumptions about the observer's measurements (specified in the generative model). As discussed earlier, the inference process involves computation of posterior probability distribution given the likelihood function, and the prior distribution. It essentially involves the "inversion" of the generative model in order to perform computation about the world state, W given the sensory data, x. Given the posterior distribution, denoted by $p(W|x)$, the observer makes a single estimate of the world state.

Commonly, the observer follows the *maximum-a-posteriori estimation (MAP)* to obtain the point estimate that has the highest posterior probability. MAP estimation is one of the most common ways of reading out the posterior distribution as it finds the most probable option. The prior distribution $p(W)$, the likelihood function $L(W|x)$ or $p(x|W)$, and the posterior distribution $p(W|x)$ are key components in the inference process. Figure 1.1(B) describes a general scheme of the inference model in the Bayesian modeling.

We can also use the likelihood function to make the best guess of the stimulus by maximizing the function over hypothesized world states. This estimate is called as the *maximum-likelihood estimate (MLE)* of W, and is denoted by $\hat{W}_{\text{ML}}$. Mathematically, we write

$$\hat{W}_{\text{ML}} = \arg\max_{W} L(W|x). \tag{1.4}$$

1.3.4.1 Marginalization

We note that the observer is interested in determining W, and not s. Hence, the generative model includes a variable that is not of our interest, but it provides important information for computing the posterior distribution function. Such a situation is dealt with the *marginalization* process, where the information about the intermediate variable is averaged out. It is a commonly applied technique in a Bayesian model, that integrates or sums the values of all such auxiliary variables to obtain the desired probability distribution over the parameter of interest. Although auxiliary variables, such as, the stimulus are not of primary interest, but

they play a critical role in the generative model, and must be accounted for via marginalization to obtain an accurate perception.

In the present case, we need to compute the posterior distribution $p(W|x)$, and not $p(s|x)$. We can obtain it using the Bayes' formula described in Eq. (1.2)

$$p(W|x) \propto p(x|W)p(W), \tag{1.5}$$

The distribution $p(x|W)$ is not specified in our Bayesian model, instead we have the information about the noise distribution, $p(x|s)$. Thus, we marginalize the information over the intermediate variable s to obtain the required distribution $p(x|W)$ as follows

$$p(x|W) = \int p(x|s, W)p(s|W)ds, \qquad \text{if } s \text{ is continuous,}$$

$$= \sum_i p(x|s = s_i)p(s = s_i|W), \quad \text{if } s \text{ is discrete.}$$

The above marginalization step, also shown in Figure 1.2 links the world state variable W, to the measurement x, via the intermediate variable, the stimulus s.

We further note that the shape of the posterior distribution is preserved under the normalization constant in Eq. (1.5). After computing the posterior distribution, the observer then reads the maximum-a-posteriori estimate, denoted by $\hat{W}_{\text{MAP}}$, by maximizing the posterior distribution

$$\hat{W}_{\text{MAP}} = \arg\max_W p(W|x). \tag{1.6}$$

The MAP estimate, $\hat{W}_{\text{MAP}}$ is also the mode of the posterior distribution, and reflects the observer's estimate of the world state of interest.
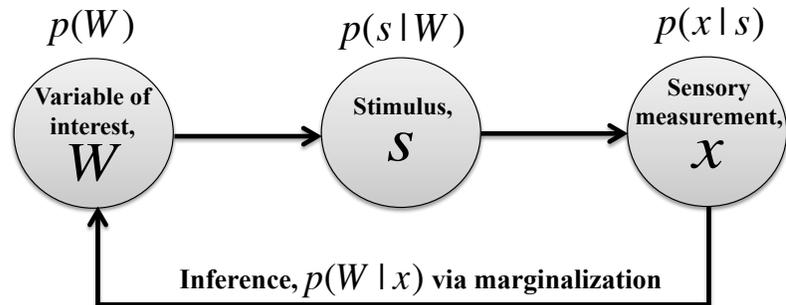
Figure 1.2: **The process of marginalization in Bayesian modeling.** The generative model usually contains auxiliary variables that are not of primary interest, but they may have necessary information about the state of the world, W. Here the stimulus, s is an intermediate variable, but it links the world state variable of interest with the measurement, x. Marginalization is a process to deal with such ancillary variables to obtain the desired expression for the likelihood or posterior probability of the world state variable of interest. It involves averaging or integrating over the possible values of the ancillary variable, and is very common in Bayesian modeling.

1.3.5   Step 3: the observer's estimate distribution

The observer's measurement x heavily depends on the sensory noise. Even under same experimental conditions, the measurements vary across trials due to different sensory noise. Therefore, x is a random variable across experimental trials. As a result, the MAP estimate $\hat{W}_{\text{MAP}}$ is also a random variable in response to a fixed stimulus, and has a probability distribution. In the final step of Bayesian modeling, we thus determine the distribution of the observer's estimates. We note that the mapping from measurement to MAP estimate is completely deterministic, and the randomness in MAP estimate is only because of variability in the measurements from trial to trial.

Since we only have access to the MAP estimates of the observer, we need to compute the probability of each possible estimate in a particular experimental condition. We can then compare the predictions of the model with the observer's behavior. Thus, we compute the distribution of MAP estimates given a fixed stimulus, say $s = s_{\text{true}}$. This is usually denoted by $p(\hat{W}_{\text{MAP}}|s_{\text{true}})$, and predicts how likely the estimate is given the fixed stimulus s_{true}.

To summarize, the Bayesian modeling of a perceptual task is based on the three steps of specifying the generative model, deriving the inference process for the observer, and evaluating the distribution of MAP estimates over many trials. This mathematical tool of quantifying the perceptual behavior is schematically shown in Figure 1.3.

Figure 1.3: **Diagram of the steps involved in Bayesian modeling of a perceptual task.** The figure presents the schematic of a Bayesian inference process to model a perceptual task. We will follow this plan for all the tasks discussed in the dissertation. The first step of specifying the generative model involves describing the probability distributions to understand how sensory data is generated from the state of the world. The observer makes an estimate of the world state based on the sensory measurement on each trial of the task. This constitutes the second step of deriving the inference process in a Bayesian model. The estimate of the observer varies across trials in response to a fixed stimulus, and follows a distribution. In the final step of Bayesian modeling of the task, this estimate distribution is computed.

So far, we have described the general process of modeling a perceptual task

using a Bayesian approach. In the dissertation, we use Bayesian models to understand visual perceptual inference. In the following section, we discuss a particular group of vision based perceptual tasks, known as *visual search tasks*. Specifically, we elaborate the process of Bayesian modeling in a simple example of such a task, and discuss the possible questions that need to be explored further. We will examine those questions in extensive details in subsequent chapters.

1.4 Visual search

Visual search is a common example of vision based perceptual task. This involves an active scan of multiple objects for a particular object or feature of interest, referred as *the target* among other objects or features, *the distractors*. Finding a friend in a large crowd or finding a particular set of keys among other similar items or locating an insect hidden in the corner are some examples of visual search from our everyday life. These examples also indicate the high importance of performing visual search in our normal life. But, doing psychophysics with natural scenes is challenging. The natural scenes are high dimensional, and highly structured, they are so rich in content that a single mathematical model may not be plausible to capture all the characteristics of a scene. Moreover, the noise in natural scenes is largely unknown, and perhaps has a complex correlation structure that is harder to capture with simple probability distributions. Further, a distinct object classification may be unavailable in case of natural scenes. Therefore, for the purposes of psychophysical studies conducted in the laboratory, quite simplified visual search

tasks are considered. These psychophysical tasks usually contain some highly distinct objects that only differ along a small number of stimulus dimensions. Clearly, these simple and fabricated tasks do not replicate natural scenes, however, they serve as a practical tool for understanding the perceptual inference computation performed by the brain.

The ability to consciously locate an object (target) among a complex array of stimuli (distractors) has been extensively studied in psychophysics over many years [104, 34, 139, 40, 38, 8, 39, 107, 108, 7, 143, 90, 94, 95]. These studies also validate the modeling of our perceptual behavior using Bayesian approach.

1.4.1 An example of a target detection task

We now discuss a specific example of visual search task, namely, a *target detection task*. Our example is similar to the one discussed by Ma et al. in [89]. We consider a simple task with only two stimuli. The observer needs to report whether a target stimulus is present in the scene. Stimulus orientation is the task-relevant feature. We elaborate the mathematical steps involved in the Bayesian modeling of the task.

1.4.1.1 Step 1: Generative model

The observer is presented with two stimuli on a visual display. These stimuli could either be bars or gratings characterized by their orientation, or ellipses characterized by their orientation or eccentricity. A target is a stimulus with a particular characteristic. We assume that a target is a stimulus with vertical orientation, and we denote its orientation by $s_T = 0$, and measure stimulus orientation relative that of a target. A distractor is defined as the stimulus having a non-target orientation. The observer reports whether a target is present in the visual display on each trial. We denote the target presence by a binary variable, T, so that $T = 1$, if a target is present, and $T = 0$ if absent. The target is present with a probability of 0.5 in each trial, and it can be present at either of the two locations. Thus, the location of the target is unknown to the observer, and hence, the observer needs to take this into account by marginalizing over both possible target locations. Figure 1.4(C) shows the three possible displays in the task. We also represent the target presence at location, i, by a binary variable, T_i, and further define the spatial location vector as $\mathbf{T} = (T_1, T_2)$. We denote the stimulus orientation of the ith stimulus by s_i.

When the target is absent, *i.e.*, $T = 0$, each stimulus orientation is drawn from a normal distribution with mean 0, and standard deviation σ_s, and we write (see notation in Appendix A)

$$s_i|T = 0 \sim \mathcal{N}(0, \sigma_s^2). \tag{1.7}$$

We denote the probability density function of the normal distribution $\mathcal{N}(0, \sigma_s^2)$ by

$f(s_i; 0, \sigma_s^2)$, where

$$f(s_i; 0, \sigma_s^2) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{s_i^2}{2\sigma_s^2}\right).$$

The assumption of the Gaussian noise is reasonable since the stimuli are not placed too close to each other on the screen. When $T = 1$, one of the stimuli is chosen as the target with uniform probability. If the target is present at location, j, for $j \in \{1, 2\}$, then $s_j = s_T = 0$, and we choose the distractor orientation according to Eq. (1.7). We assume that the observer makes independent (between locations), and noisy measurement of the stimulus, s_i, denoted by x_i, which is drawn from the following normal distribution

$$x_i|s_i \sim \mathcal{N}(s_i, \sigma_i^2). \tag{1.8}$$

The noise, σ_i determines the uncertainty in the ith measurement of the stimulus, s_i, and is known to the observer. The generative model of the task is illustrated in Figure 1.4(A).

1.4.1.2 Step 2: Inference

The optimal Bayesian observer infers target presence based on the stimulus measurements, and using the information about the generative model. The observer computes the *log posterior ratio (LPR)* of target presence from the measurements, x_1, and x_2 as

$$d(x_1, x_2) = \log \frac{p(T = 1|x_1, x_2)}{p(T = 0|x_1, x_2)}, \tag{1.9}$$

and reports "target present" if $d > 0$, and "target absent" otherwise. The variable, d is known as the Bayesian decision variable. Using the Bayes' formula, we rewrite

(A)

(B)

(C)
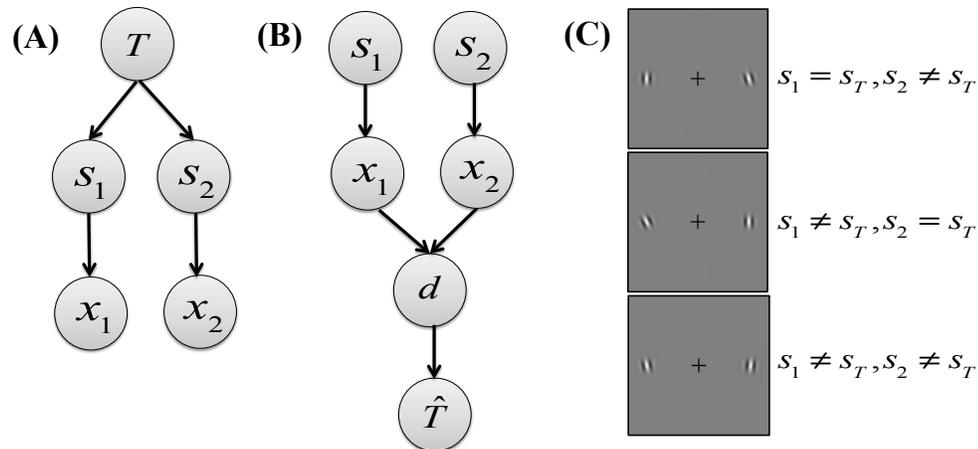
Figure 1.4: **Bayesian modeling of a simple target detection task with two stimuli. (A) The generative model.** The binary variable, T describes the target presence in a trial and determines the structure of the display. The two stimuli, s_1, and s_2 are chosen conditioned on T. When $T = 1$, one of the stimuli is a target with a vertical orientation, while the orientation of the other stimulus is chosen randomly from a normal distribution. The observer makes noisy and independent measurements, x_1, and x_2 of the two stimulus. **(B) The inference process.** The observer combines the two measurements to compute a decision variable, $d(x_1, x_2)$, and infers an estimate $\hat{T}$ of the world state variable, T. The variable $d(x_1, x_2)$ is a log posterior ratio of the probability of reporting "target present", and "target absent" given the observer's measurements. If $d > 0$, the observer reports target is present, and absent otherwise. **(C) Example displays in the task.** Since there are only two stimuli, and one target, three types of visual displays can be presented to the observer. In the first two displays, the target is present to the left and right of the cross in the center. When there is no target, both stimuli are distractors and have randomly chosen orientations. The bottom display illustrates such an example.

the above equation as the sum of *log likelihood ratio (LLR)*, and *log prior ratio*,

$$d(x_1, x_2) = \log \frac{p(x_1, x_2 | T = 1)}{p(x_1, x_2 | T = 0)} + \log \frac{p(T = 1)}{p(T = 0)}.$$

Since the target is present or absent with equal probability, we have a uniform prior over T, and therefore,

$$d(x_1, x_2) = \log \frac{p(x_1, x_2 | T = 1)}{p(x_1, x_2 | T = 0)}. \tag{1.10}$$

We compute the numerator in Eq. (1.10) by marginalizing over the two possible target locations, and stimuli, s_i in the following equations

$$
\begin{aligned}
p(x_1, x_2 | T = 1) &= \sum_{j=1}^{2} p(x_1, x_2 | T_j = 1, T = 1) p(T_j = 1 | T = 1) \\
&= \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{2} p(x_j | T_j = 1) p(x_i | T_i = 0) \\
&= \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{2} \int p(x_j | s_j) p(s_j | T_j = 1) p(x_i | s_i) p(s_i | T_i = 0) ds_j ds_i \\
&= \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{2} \int f(x_j; s_j, \sigma_j^2) \delta(s_j - 0) f(x_i; s_i, \sigma_i^2) f(s_i; 0, \sigma_s^2) ds_j ds_i \\
&= \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{2} f(x_j; 0, \sigma_j^2) f(x_i; 0, \sigma_i^2 + \sigma_s^2) \quad \text{(using Eqs. (B.1) and (B.4))} \\
&= \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{2} \frac{1}{2\pi \sqrt{\sigma_j^2 (\sigma_i^2 + \sigma_s^2)}} \exp\left(-\frac{x_j^2}{2\sigma_j^2} - \frac{x_i^2}{2(\sigma_i^2 + \sigma_s^2)} \right).
\end{aligned}
$$

When the target is absent, both stimuli are distractors. Thus, we compute the denominator in Eq. (1.10) as

$$
\begin{aligned}
p(x_1, x_2 | T = 0) &= p(x_1 | T_1 = 0)p(T_1 = 0 | T = 0)p(x_2 | T_2 = 0)p(T_2 = 0 | T = 0) \\
&= \int p(x_1 | s_1)p(s_1 | T_1 = 0)p(x_2 | s_2)p(s_2 | T_2 = 0)ds_1 ds_2 \\
&= \int \prod_{i=1}^{2} \left(f(x_i; s_i, \sigma_i^2)f(s_i; 0, \sigma_s^2)ds_i \right) \\
&= \prod_{i=1}^{2} f(x_i; 0, \sigma_i^2 + \sigma_s^2).
\end{aligned}
$$

We now substitute the above computed individual likelihoods for target present, and absent cases to obtain an expression for Bayesian decision variable, d:

$$
\begin{aligned}
d(x_1, x_2) &= \log \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{2} \frac{f(x_j; 0, \sigma_j^2)f(x_i; 0, \sigma_i^2 + \sigma_s^2)}{f(x_j; 0, \sigma_j^2 + \sigma_s^2)f(x_i; 0, \sigma_i^2 + \sigma_s^2)} \\
&= \log \frac{1}{2} \sum_{j=1}^{2} \frac{f(x_j; 0, \sigma_j^2)}{f(x_j; 0, \sigma_j^2 + \sigma_s^2)} \\
&= \log \frac{1}{2} \sum_{j=1}^{2} \left(\frac{\sigma_j^2 + \sigma_s^2}{\sigma_j^2} \right)^{1/2} \exp \left(\underbrace{-\frac{x_j^2}{2\sigma_j^2}}_{I} - \underbrace{\frac{x_j^2}{2(\sigma_j^2 + \sigma_s^2)}}_{II} \right). \quad (1.11)
\end{aligned}
$$

The above expression gives us the decision variable on the task. An ideal Bayesian observer performing the above described target detection task makes a decision based on this decision variable. The decision variable, $d(x_1, x_2)$ depends on the precision of the measurement, and also on the external variability of the distractor, σ_s^2. Each exponent term in the above expression provides an evidence towards j^{th} stimulus being a target: (I) if the jth measurement is close to the vertical orientation, this term corresponds to an increased likelihood for the jth stimulus being

the target, while (II) the second term decreases such a likelihood. The appropriate scaling of the two measurements by inverse of the external, and internal noise variances determine the correct likelihood for the jth stimulus being the target or a distractor.

1.4.1.3 Step 3: MAP estimate distribution

We denote the observer's MAP estimate of T by $\hat{T}$. When $d > 0$, the MAP estimation is to report "target present", and "target absent" otherwise. The distribution of the MAP estimate describes the behavior of the Bayesian observer across many trials, and involves computing the probabilities that the observer will report "target present" when the target is actually present, and when it is absent. That is, we need to compute $p(\hat{T} = 1|T = 1)$, and $p(\hat{T} = 1|T = 0)$. These probabilities are also known as the *hit* and *false-alarm rates*, and are computed given a fixed set of stimulus, s_1 and s_2:

$$p(\hat{T}|s_1, s_2) = \int p(\hat{T}|x_1, x_2)p(x_1|s_1)p(x_2|s_2)dx_1dx_2$$
$$= \int \delta_{\hat{T}, sgn(d(x_1, x_2))} p(x_1|s_1)p(x_2|s_2)dx_1dx_2.$$

Here δ represents the Kronecker delta function. As the decision variable, d computed in Eq. (1.11) is a non-linear function of x_1, and x_2, the above integral is analytically intractable, and hence needs to be approximated using Monte Carlo simulations in practice. This completes the final step in the Bayesian modeling of the above described target detection task.

1.4.1.4 A suboptimal model

Eq. (1.11) describes the decision variable for an optimal Bayesian observer to perform the task. However, it is not necessary that the observer will follow this rule. Specifically, the observer may not be optimal in making a decision on the task, and might use some other decision strategy to make a decision. The decision variable $d(x_1, x_2)$ computed in Eq. (1.11) is clearly non-linear, and implicitly depends on other parameters that characterize the structure of the task. For instance, the external variance of the distractor stimulus σ_s^2, and the noise in making a measurement σ_i^2 clearly affect this variable. It is generally assumed that the observer is aware of the internal noise with which the measurement is made, but he might not know the external variance that determines the structure of the task. In such a case, the observer would use an incorrect assumption about the generative model in making a decision. In the worst case, the observer could also make a guess on each trial without using any information about the task.

Therefore, the following questions frequently emerge in analyzing the responses of the observer on such tasks: What model does the observer follow in making a decision on the task? How do we infer the parameters of a model that fit the subject's data? And in the case of multiple models, how can we compare all models to determine the best model describing the experimental data? We will examine these questions in great length in Chapters 3 to 5, and present detailed analysis for a particular target detection task described in Chapter 2.

We now consider an alternative to the optimal model derived in Section 1.4.1.2.

Let us assume that the observer is not optimal, and instead makes a decision based on the minimum of the two measurements of the stimuli. If the minimum of the measurements is below some threshold, θ, the observer reports "target present", and absent otherwise. We denote the decision variable for such a model by $d_{\min}(x_1, x_2)$, and define

$$d_{\min}(x_1, x_2) = \min(|x_1|, |x_2|). \tag{1.12}$$

The observer compares whether $d_{\min}(x_1, x_2) < \theta$ to make an estimate $\hat{T} = 1$. We note that the observer's decisions based on the decision variable $d_{\min}(x_1, x_2)$ of the threshold model are suboptimal. The threshold models have been used in many earlier studies to model subjects' responses on psychophysical tasks [108, 143, 8, 39, 52, 90].

Figure 1.5 compares the performance of an observer as a function of the standard deviation of the stimulus distribution, σ_s, for the optimal, and threshold models. Here we assume that the observer has equal precision for both measurements, that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We consider the performance of the observer based on the following models: (i) optimal model, (ii) threshold model, with $\theta = \frac{1}{2}\sigma$, (iii) threshold model with $\theta = \sigma$, and (iv) threshold model, with $\theta = \frac{3}{2}\sigma$. Thus, we consider models where the threshold parameter, θ depends on the precision of the observers' measurements. We note that the performance increases for all models as the standard deviation of distractor orientations, σ_s increases. This is expected since the task becomes easier as the orientations of the distractors deviates relative to that of the target. Also, we observe that the performance predicted by threshold models is lower as compared to the optimal model when the threshold parameter,
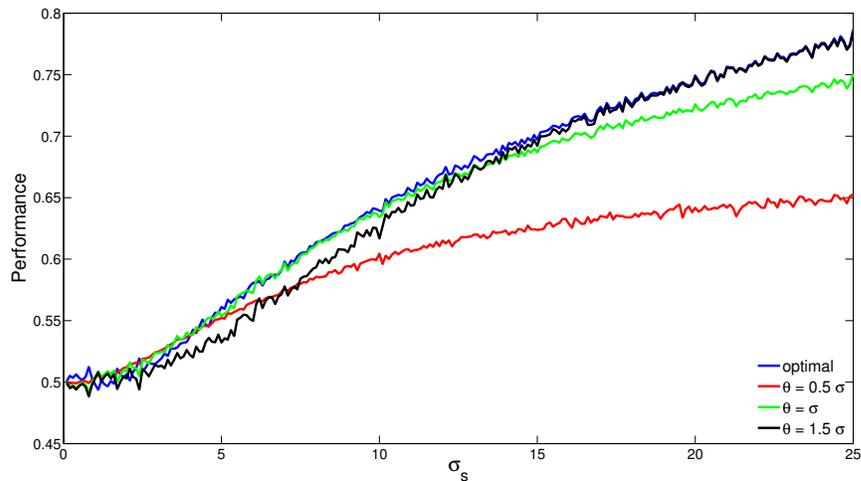
Figure 1.5: **Performance of an observer in a simple target detection example based on different models.** The performance of an observer as a function of the standard deviation of the distractor orientations, σ_s based on different models. The optimal model has the maximum performance at all values of σ_s than other threshold models. A lower performance is observed at low standard deviations for all models since it becomes to difficult to detect a target on the task among distractors that have relatively similar orientations to that of the target. As the standard deviation of the distractor orientations increases, the task becomes relatively easier, and the performance increases in the case of all models. The model with high threshold parameter, θ predicts a similar performance of the observer as in the optimal model. This indicates that it is difficult to choose a model that is most consistent to describe the observer's behavior.

θ is small. However, for a sufficiently large threshold value (here $\theta = \frac{3}{2}\sigma$), the observer behavior is similar to the optimal model at large value of σ_s. We find that all models closely predict the performance as that of the optimal model at low values of standard deviation between distractors. At low values of σ_s, the distractors are more likely to have orientations close to that of the target (see Eq. (1.7)). In such a case, it will be difficult for the observer to make a decision, and will have lower performance as predicted by both optimal, and threshold models.

Therefore, the decisions of an observer depend on the model parameters, and the precision of the measurements. In order to understand the responses of the observer, we need to estimate these model parameters, and also determine the precision level of observer's measurements. Based on the parameter estimates, we make predictions of the observer behavior based on different plausible models. These models can be very close in their predictions (for example, optimal, and threshold model with $\theta = \frac{3}{2}\sigma$ in Figure 1.5), and it may be difficult to find the best model that is consistent with the observer's behavior.

1.4.2 Generalizations

In general, visual search task comprises of group of tasks: *target detection* - determining whether a target is present or not in a scene; *target localization* - finding the location of the target when the target is always present, and *target discrimination or classification* - classifying the target to one of the pre-defined categories. We only focus on the target detection tasks in the dissertation, and explore different

parameter relations in these tasks.

The target detection task described in Section 1.4.1 is an extremely simple example of a target (or visual) search task with only two stimuli. However, in general, we make decisions in presence of a large number of distractors. For instance, in order to find a friend in a large crowd, we need to carefully scan each individual, and undergo an identification process. Also, the distractor orientations may not always be independent. They could possess an unknown complex structure, and we must account for such structures to infer correct decisions.

Several research studies have considered the visual search tasks with a reasonably large number of stimuli [8, 71, 41, 107, 94, 127, 39, 95, 142, 90, 141, 129]. Some of them have focused on analyzing performance as a function of set-size [107, 8, 39, 127, 41, 128, 94, 95]. Furthermore, the experimental studies done by Ma et al. [90, 142, 94, 95] have explored the behavior of the subjects on a search task with two types of distractors: when all distractors have identical orientation, and when the distractors possess different independent orientations.

In addition to the possibility of varying set size, and structural orientations of the distractors, the task could also have multiple targets in the visual display. Such a possibility has not been explored in scientific studies. We thus examine this possibility with complete mathematical details in Chapter 7, and further analyze the impact of different parameter correlations on the performance of an optimal observer on the task. In real life, there are several examples where multiple targets are present, and our brain needs to process the information to find at least one of them. For instance, we may need to find a blue marker in a box of black and blue

markers. There could be more than one blue marker present in the box among black markers that serve as distractors.

Thus far, we have described the principles of Bayesian approach used in building perceptual models. These concepts are fundamentals for the work presented in the dissertation. We now present the outline of the dissertation, and the main questions examined in the following chapters.

1.5 Outline of the dissertation

The dissertation is split into two parts. Each part is based on the application of the Bayesian approach to understand how humans make decisions.

Visual search for a single target among distractors, with a single relevant feature has been studied extensively. These studies have largely focused on two types of distractors: distractors with identical orientations, and with independent random orientations. Therefore, either distractors are exactly alike, or they differ from each other across trials. In a target present trial with identical distractors, the target would be an odd-ball stimulus, and hence can be detected easily. While if the distractors have independent orientations, there would be hardly any structure in the scene that could possibly help the observer in finding a target. The experimental studies [90, 95] showed that human decision-making behavior is consistent with optimal Bayesian models in the case of both types of distractors. But, these two categories of distractors represent the two extremes: from high structure in

the scene to none. However, natural scenes possess more complex structure, and the objects in the scene can be correlated with each other in many possible ways. Therefore, this raises the question how humans make decisions in response to scenes with weak structures? In particular, do humans take into account the weak correlations among objects in the scene, and do they infer near-optimal decisions in such a case?

We examine these questions in Chapters 2 to 5. We consider a target detection task with N stimuli, and one target. The distractors are assigned orientations that have different amount of correlations. The varying amount of correlations among distractor orientations allows us to introduce structure in the visual scene. The structure of these correlations must be taken into account to make optimal decision about target presence. In Chapter 2, we derive the mathematical theory of the Bayes-optimal model to infer correct decisions on this task. We also perform a psychophysical experiment based on the design of the task, and analyze the collected human subjects' data in Chapters 4, and 5.

We explore whether humans are optimal in inferring correct correlation strength among distractor orientations. We test several Bayesian models that could possibly explain subjects' behavior on the experiment, and fit them to the data. The fitting of a model requires finding the model parameters that fit to the subjects' responses. Further, we need techniques to compare any two models to determine the better one. We describe the maximum-likelihood parameter estimation, and model comparison techniques in details in Chapter 3. We use these techniques to fit the different parameters of the models explained in Chapter 4, and find the

model that best describes our experimental data in Chapter 5.

For the purposes of our analysis in Chapters 2 and 5, we have assumed the noise in the measurements to be independent, and normally distributed. Both the assumptions about independence, and Gaussianity can be questioned. In particular, there is some evidence that neural correlations can be present at long distances in visual field suggesting that sensory measurements will be correlated [36, 28, 27, 118]. However, much of the visual search studies lack the assumption about noise correlations in the sensory measurements. We thus focus on the effects of such an assumption on the performance in categorical, and global perceptual judgements. To make a correct decision in this case, the observer needs to take into account not only the correlations between the measurements, but also the statistical structure of the stimuli.

In the second part of the dissertation (Chapters 6 and 7), we explore the joint effects of measurement, and the stimulus correlations in a family of visual search tasks . To investigate how the interaction between two different types of correlations: the stimulus, and measurements can affect the decisions of an observer, we include the assumption of correlated sensory noise in the target detection task introduced in Chapter 2. Thus, in Chapter 6, we study a target detection task with a single target, and the sensory measurements of the observer are correlated following a multivariate normal distribution. We provide complete details of the analytical computations that an ideal observer follows to make a correct decision on this task. We analyze the impact of statistical structure of the scene along with measurement correlations on the performance of an ideal observer here.

Further, in Chapter 7, we continue with our examination, and analysis in the case of multiple targets in the detection task. Multiple targets having identical orientation would introduce more statistical structure in the scene as compared to a single target. Even if the distractors possess independent orientations, having multiple targets increase the chance of detecting a target in a pool of stimuli. Therefore, we inspect how the external structure present in the stimuli communicates with the internal structure of the observer's measurements, and affects the decisions of the observer. In particular, we analytically, and numerically analyzed how the performance of an ideal observer behaves as a function of different parameters that determine the structure in the external scene, and measurements of the observer. We find that the performance changes by a considerable amount in the case of multiple targets, while it remains unchanged when only a single target is present. Therefore, the decisions on these visual search tasks are greatly influenced by the relationship between the correlations present in the stimulus, and measurements.

Stimulus correlations in a target detection task

Bayesian models have been successfully used to study perceptual behavior. Many studies have used these quantitive models to show that humans perform near-optimally on simple perceptual tasks. That is, the behavior of the observers on these tasks was successfully explained by optimal Bayesian models which assumes that the observers make best possible decisions given the uncertain, and noisy sensory measurements.

A number of recent studies have also analyzed human behavior in simplified scenes containing multiple objects. Human behavior was found to be close to Bayes-optimal, on different visual search [90, 94, 95], sameness judgement [142], and change detection tasks [71]. However, many of the visual search studies have

only considered two types of distractors: *homogeneous*, and *heterogeneous* distractors [138, 34, 151, 101, 143, 119, 90, 94, 102, 95]. Homogeneous condition refers to having identical distractors, while heterogeneous condition have independent randomly oriented distractors. These studies considered the orientations of homogeneous distractors to be same across all trials, while they varied the orientation of distractors across experimental trials in the heterogeneous condition. Thus, orientations of the distractors on homogeneous condition were predictable from trial to trial.

Recent studies done by Mazyar et al. [95] studied the human behavior on a target detection task under the violation of trial-to-trial predictability in the case of homogeneous distractors. They used Gaussian distribution to randomly draw the orientation of distractors in the case of homogeneous condition. They also performed experiments to study human decisions in response to heterogeneous distractors, and found that humans were near-optimal in detecting a target in the case of both types of distractors.

However, their studies were also limited to homogeneous, and heterogeneous distractors. By contrast, visual stimuli in natural scenes possess a complex, and higher-order structure. The orientations of the objects in natural scenes are correlated to different extent with each other. It is therefore important to examine how visual perception is affected by structured input. Specifically, to understand how differently correlated input affect our decisions.

We examine these questions in a psychophysical task. We study the decisions of human observers on a target detection experiment under the effect of structured

input. We introduce structure in visual scenes by varying the amount of correlations between distractor orientations. We note that homogeneous condition refers to perfectly correlated (identical) distractors, while heterogeneous distractors are uncorrelated. In our study, we interpolated between the heterogeneous, and the homogeneous conditions, and studied the intermediate regimes of partial correlations. We are interested in understanding how human observers make decisions in response to differently correlated (uncorrelated, partially, and perfectly correlated) stimuli. The intermediate regime of correlations can be challenging from an observer's point of view since the stimuli are only partially correlated, and only introduce a weak structure in the scenes. The observer needs to take into account the strength, and structure of the stimulus correlations to make an optimal decision.

Therefore, we examine the following questions in our study: Do humans take into account the strength, and structure of the stimulus correlations? And if they do, can they make near-optimal decisions? We note that visual search is one particular example of a task where these questions are relevant. They are more generally applied to a variety of perceptual tasks. We provide answers to these questions for the target detection task in Chapter 4.

We begin this chapter with the description of the model for the target detection task. We continue with the derivation of the inference process for an optimal Bayesian observer. An ideal observer makes decisions according to the derived decision rule to infer target presence on the task. Later in the chapter, we present the details of the task based experiments we have conducted.

2.1 Generative model

We consider the following target detection task: An observer is presented with N stimuli. The observer reports whether a vertical target stimulus is present or absent among a group of distractors. Stimuli are characterized by their orientations. We denote the target orientation by $s_T = 0$. We represent target presence by the binary variable, T, so that $T = 1$ if the target is present, and $T = 0$ if absent. This notation is consistent with studies done by Ma et al. [90, 94, 95]. Target presence at location i is similarly represented by a binary variable, T_i. We also denote the spatial location vector by $\mathbf{T} = (T_1, T_2, \cdots, T_N)$. In each trial, the target is present with a $1/2$ probability.

We denote the orientations of the stimuli by $\mathbf{s} = (s_1, s_2, \cdots, s_N)$. When $T = 0$, the target is absent, and all stimuli are distractors. Therefore, $\mathbf{T} = (0, 0, \cdots, 0)$, and we write

$$p(\mathbf{T}|T = 0) = \delta(\mathbf{T} - \mathbf{0}_N).$$

Here a subscript denotes the length of a vector, so that the vector $\mathbf{0}_N$ has N components. In this case, the orientations of the stimuli are drawn from an N-dimensional multivariate normal distribution with mean vector, $\mathbf{s_D} = (s_D, s_D, \cdots, s_D)$ and covariance, $\Sigma_\mathbf{s}$, and we write

$$\mathbf{s}|T = 0 \sim \mathcal{N}(\mathbf{s_D}, \Sigma_\mathbf{s}). \tag{2.1}$$

We denote the probability density function of the multivariate normal distribution, $\mathcal{N}(\mathbf{s_D}, \Sigma_{\mathbf{s}})$ by $f(\mathbf{s}; \mathbf{s_D}, \Sigma_{\mathbf{s}})$, where

$$f(\mathbf{s}; \mathbf{s_D}, \Sigma_{\mathbf{s}}) = \frac{1}{\sqrt{(2\pi)^N |\Sigma_{\mathbf{s}}|}} \exp\left(-\frac{1}{2}(\mathbf{s} - \mathbf{s_D})^{\mathrm{T}} \Sigma_{\mathbf{s}}^{-1} (\mathbf{s} - \mathbf{s_D})\right). \tag{2.2}$$

The $N \times N$ covariance matrix, $\Sigma_{\mathbf{s}}$, contains identical diagonal entries, σ_s^2 (variances), and identical off-diagonal entries, $\rho_s \sigma_s^2$ (covariances):

$$\Sigma_{\mathbf{s}} = \begin{bmatrix} \sigma_s^2 & \rho_s \sigma_s^2 & \cdots & \rho_s \sigma_s^2 \\ \rho_s \sigma_s^2 & \sigma_s^2 & \cdots & \rho_s \sigma_s^2 \\ \vdots & & \ddots & \vdots \\ \rho_s \sigma_s^2 & \rho_s \sigma_s^2 & \cdots & \sigma_s^2 \end{bmatrix}. \tag{2.3}$$

We let the pairwise correlation coefficient, ρ_s, vary between 0 and 1. When $T = 1$, one of the N possible location is chosen with equal probability, and the stimuli at that location is assigned the target orientation. If $\mathbf{1}_j$ represents the N-dimensional vector having j^{th} entry as 1, and rest zeros, then

$$p(\mathbf{T}|T = 1) = \frac{1}{N} \sum_{j=1}^{N} \delta(\mathbf{T} - \mathbf{1}_j). \tag{2.4}$$

When the target is present at location j, for some fixed $j \in \{1, 2, \cdots, N\}$, we have

$$p(s_j | T_j = 1) = \delta(s_j - s_T). \tag{2.5}$$

In such a case, the distractors will be present at all locations but the jth one. Therefore, we can decompose the Eq. (2.4) into target, and distractors location:

$$p(T_j = 1 | T = 1) = \frac{1}{N} \delta(T_j - 1), \text{ and } p(\mathbf{T}_{\backslash j} = \mathbf{0}_{N-1} | T = 1) = \delta(\mathbf{T}_{\backslash j} - \mathbf{0}_{N-1}).$$

The vector $\mathbf{T}_{\backslash j}$ is obtained by removing the jth location from the spatial location vector $\mathbf{T}$. In the target present trial, the orientations of the $N-1$ distractors are denoted by $\mathbf{s}_{\backslash \mathbf{j}}$ (see notation in Appendix A). They are drawn from an $(N-1)$-dimensional multivariate normal distribution with $(N-1)$-dimensional mean vector $\mathbf{s}_{\mathbf{D}_{\backslash \mathbf{j}}}$, and covariance $\Sigma_{\mathbf{s}\backslash j}$ (conditioned on j^{th} stimuli being the target). We write

$$\mathbf{s}_{\backslash \mathbf{j}}|T = 1 \sim \mathcal{N}(\mathbf{s}_{\mathbf{D}_{\backslash \mathbf{j}}}, \Sigma_{\mathbf{s}\backslash j}), \tag{2.6}$$

where the $(N-1) \times (N-1)$ covariance matrix, $\Sigma_{\mathbf{s}\backslash j}$ is obtained by removing the jth row and the jth column of $\Sigma_{\mathbf{s}}$.

We assume that an observer makes a measurement, x_i, of the presented stimulus, s_i, for $i \in \{1, 2, \cdots, N\}$. It is commonly assumed [90, 94, 95, 142] that these measurements are noisy but unbiased, and are normally distributed. Therefore, we assume that at each location i,

$$x_i|s_i \sim \mathcal{N}(s_i, \sigma_i^2). \tag{2.7}$$

Moreover, we consider the measurement noise to be independent between locations. Hence, for the vector of measurements, $\mathbf{x} = (x_1, x_2, \cdots, x_N)$, we write

$$\mathbf{x}|\mathbf{s} \sim \mathcal{N}(\mathbf{s}, \Sigma_{\mathbf{x}}) = \prod_{i=1}^{N} \mathcal{N}(s_i, \sigma_i^2). \tag{2.8}$$

Here $\Sigma_{\mathbf{x}}$ is an $N \times N$ diagonal matrix with entries $\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2$ on the diagonal. The optimal-observer model of the task is illustrated in Figure 2.1.
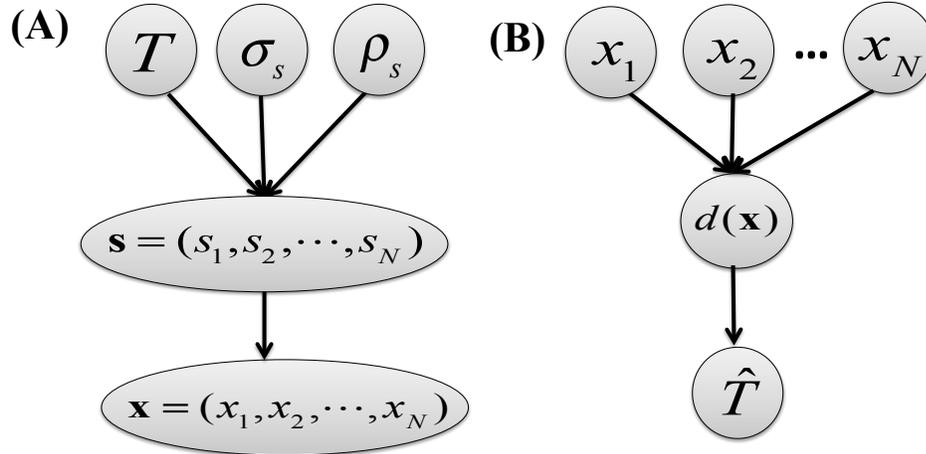
Figure 2.1: **Statistical structure of relevant task variables in the optimal-observer model for a target detection task with stimulus correlations. (A) Generative model.** The nodes represent the variables in the task, and arrows indicate conditional dependencies between them. The binary variable, T represents target presence for $T = 1$, and absent when $T = 0$. The standard deviation, σ_s, and the pairwise correlation coefficient, ρ_s determine the structure of the stimulus, $\mathbf{s} = (s_1, s_2, \cdots, s_N)$ in the task. An observer makes a measurement, x_i, of each presented stimulus s_i. These measurements are assumed to be noisy and independent between locations. **(B) Inference process.** The optimal observer infers T by "inverting" the generative model. The observer computes a decision variable, $d(\mathbf{x})$ based on the measurements, $\mathbf{x}$, and it is given by the log-posterior ratio between the two possibilities, $\log \left(p(T = 1|\mathbf{x})/p(T = 0|\mathbf{x}) \right)$. The sign of $d(\mathbf{x})$ gives the optimal estimate of T, and it is denoted by $\hat{T}$.

2.2 Inference process

The observer infers target presence based on the stimulus measurements, $\mathbf{x}$, and knowledge of the process that generated the stimulus, also called the generative model (Figure 2.1(A)). Specifically, an optimal observer computes the probability of $T = 0$, and the probability of $T = 1$, given $\mathbf{x}$, and chooses the option with highest probability. This is equivalent to computing the log posterior ratio,

$$d(\mathbf{x}) = \underbrace{\log \frac{p(T = 1|\mathbf{x})}{p(T = 0|\mathbf{x})}}_{\text{log posterior ratio}} , \qquad (2.9)$$

and reporting "target present" when $d(\mathbf{x}) > 0$ and "target absent" otherwise. The relation, $d(\mathbf{x}) > 0$ to make an estimate is known as the *Bayesian decision rule*, and $d(\mathbf{x})$ itself is referred as the *Bayesian decision variable*. Here $d(\mathbf{x}) = 0$ represents the *decision boundary*, and 0 is also called as the *decision criterion*.

By applying Bayes' theorem in the above equation we obtain

$$d(\mathbf{x}) = \underbrace{\log \frac{p(T = 1|\mathbf{x})}{p(T = 0|\mathbf{x})}}_{\text{log posterior ratio}} = \underbrace{\log \frac{p(\mathbf{x}|T = 1)}{p(\mathbf{x}|T = 0)}}_{\text{log likelihood ratio}} + \underbrace{\log \frac{p(T = 1)}{p(T = 0)}}_{\text{log prior ratio}} . \qquad (2.10)$$

Here $p(T = 1)$ denotes the observer's prior belief that the target is present. Based on the above equation, the observer reports "target present" when the log-likelihood ratio is greater than the negative log prior ratio, *i.e.*,

$$\log \frac{p(\mathbf{x}|T = 1)}{p(\mathbf{x}|T = 0)} > -\log \frac{p(T = 1)}{p(T = 0)}.$$

Also, it is easy to see that any change in prior results in shifting of the decision criterion, thus prior has a large effect in the inference process of the observer. The

optimal observer uses a uniform prior on T, and compute the log-likelihood ratio, and hence, the required decision variable. We denote the log-likelihood ratio for the task by $L_{\mathrm{ST}}(\mathbf{x})$, and the decision variable by $d_{\mathrm{ST}}(\mathbf{x})$.

2.2.1 The log-likelihood ratio

The observer needs to marginalize over intermediate variables, $\mathbf{T}$, and $\mathbf{s}$ to compute the log-likelihood ratio. The marginalization process is described in Section 1.3.4.1, and illustrated with an example in Section 1.4.1. In this case, we compute the log-likelihood ratio in the following manner:

$$L_{\mathrm{ST}}(\mathbf{x}) = \log \frac{p(\mathbf{x}|T=1)}{p(\mathbf{x}|T=0)} = \log \frac{\sum_{\mathbf{T}} p(\mathbf{x}|\mathbf{T}, T=1)p(\mathbf{T}|T=1)}{\sum_{\mathbf{T}} p(\mathbf{x}|\mathbf{T}, T=0)p(\mathbf{T}|T=0)}$$

$$= \log \frac{\sum_{\mathbf{T}} \left(p(\mathbf{x}|\mathbf{T}, T=1)\frac{1}{N}\sum_{j=1}^{N}\delta(\mathbf{T}-\mathbf{1}_j) \right)}{\sum_{\mathbf{T}} p(\mathbf{x}|\mathbf{T}, T=0)\delta(\mathbf{T}-\mathbf{0}_N)}$$

$$= \log \frac{1}{N} \frac{\sum_{j=1}^{N} \left(\sum_{\mathbf{T}} p(\mathbf{x}|\mathbf{T}, T=1)\delta(\mathbf{T}-\mathbf{1}_j) \right)}{\sum_{\mathbf{T}} p(\mathbf{x}|\mathbf{T}, T=0)\delta(\mathbf{T}-\mathbf{0}_N)}$$

$$= \log \frac{1}{N} \sum_{j=1}^{N} \frac{p(\mathbf{x}|\mathbf{T}=\mathbf{1}_j)}{p(\mathbf{x}|\mathbf{T}=\mathbf{0}_N)}$$

$$= \log \frac{1}{N} \sum_{j=1}^{N} \frac{\int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|\mathbf{T}=\mathbf{1}_j)d\mathbf{s}}{\int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|\mathbf{T}=\mathbf{0}_N)d\mathbf{s}}. \tag{2.11}$$

Here δ is the generalized Kronecker delta function defined on $\mathbb{R}^N$:

$$\delta(\mathbf{z}) = \begin{cases} 1, & \text{if and only if } z_i = 0 \text{ for all } 1 \leq i \leq N, \\ \\ 0, & \text{otherwise.} \end{cases}$$

We further simplify the above expressions by decomposing the vector $\mathbf{s}$ into target stimulus, s_j, and distractors, $\mathbf{s}_{\backslash \mathbf{j}}$. Similarly, we break the vector $\mathbf{x}$ into a target measurement, x_j, and distractors measurements, $\mathbf{x}_{\backslash \mathbf{j}}$. Similar to $\Sigma_{\mathbf{s}\backslash j}$, we define matrix $\Sigma_{\mathbf{x}_{\backslash \mathbf{j}}}$ obtained by removing the jth row and column of matrix $\Sigma_{\mathbf{x}}$, so that

$$\mathbf{x}_{\backslash \mathbf{j}} | \mathbf{s}_{\backslash \mathbf{j}} \sim \mathcal{N}(\mathbf{s}_{\backslash \mathbf{j}}, \Sigma_{\mathbf{x}_{\backslash \mathbf{j}}}). \tag{2.12}$$

Thus, we obtain

$$L_{\text{ST}}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \frac{\int p(x_j|s_j)p(s_j|T_j = 1)p(\mathbf{x}_{\backslash \mathbf{j}}|\mathbf{s}_{\backslash \mathbf{j}})p(\mathbf{s}_{\backslash \mathbf{j}}|\mathbf{T}_{\backslash j} = \mathbf{0}_{N-1})ds_j \, d\mathbf{s}_{\backslash \mathbf{j}}}{\int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|T = 0)d\mathbf{s}}$$

$$= \log \frac{1}{N} \sum_{j=1}^{N} \frac{\int f(x_j; s_j, \sigma_j^2)\delta(s_j - s_T)f(\mathbf{x}_{\backslash \mathbf{j}}; \mathbf{s}_{\backslash \mathbf{j}}, \Sigma_{\mathbf{x}_{\backslash \mathbf{j}}})f(\mathbf{s}_{\backslash \mathbf{j}}; \mathbf{s}_{\mathbf{D}_{\backslash \mathbf{j}}}, \Sigma_{\mathbf{s}\backslash j})ds_j \, d\mathbf{s}_{\backslash \mathbf{j}}}{\int f(\mathbf{x}; \mathbf{s}; \Sigma_{\mathbf{x}})f(\mathbf{s}; \mathbf{s}_{\mathbf{D}}, \Sigma_{\mathbf{s}})d\mathbf{s}}.$$

We now apply the product and integral rules for normal distributions in Eqs. (B.3) and (B.4), and denote

$$\mathbf{C} = \Sigma_{\mathbf{s}} + \Sigma_{\mathbf{x}}, \text{ and } \mathbf{C}_{\backslash \mathbf{j}} = \Sigma_{\mathbf{s}\backslash j} + \Sigma_{\mathbf{x}\backslash j}. \tag{2.13}$$

The matrix $\mathbf{C}_{\backslash \mathbf{j}}$ can also be obtained by removing the jth row and column of matrix $\mathbf{C}$. In the case of positive definite covariance matrices, *i.e.*, for $\rho_s \neq 1$, we integrate

and obtain the following expression for the log-likelihood ratio,

$$
L_{ST}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \frac{f(x_j; s_T, \sigma_j^2) f(\mathbf{x}_{\backslash \mathbf{j}}; \mathbf{s}_{\mathbf{D}_{\backslash j}}, \mathbf{C}_{\backslash j})}{f(\mathbf{x}; \mathbf{s}_{\mathbf{D}}; \mathbf{C})}
$$

$$
= \log \frac{1}{N} \sum_{j=1}^{N} \sqrt{\frac{|\mathbf{C}|}{\sigma_j^2 |\mathbf{C}_{\backslash j}|}} \exp\left(-\frac{(x_j - s_T)^2}{2\sigma_j^2} - \frac{1}{2}(\mathbf{x}_{\backslash \mathbf{j}} - \mathbf{s}_{\mathbf{D}_{\backslash j}})^{\mathsf{T}} \mathbf{C}_{\backslash j}^{-1} (\mathbf{x}_{\backslash \mathbf{j}} - \mathbf{s}_{\mathbf{D}_{\backslash j}}) \right.
$$

$$
\left. + \frac{1}{2}(\mathbf{x} - \mathbf{s}_{\mathbf{D}})^{\mathsf{T}} \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}_{\mathbf{D}}) \right). \tag{2.14}
$$

We further reduce the above expression by computing the determinant, and inverse of matrices $\mathbf{C}$, and $\mathbf{C}_{\backslash j}$. The matrices $\mathbf{C}$, and $\mathbf{C}_{\backslash j}$ have rank 1, therefore, we use the matrix determinant lemma, and Sherman-Morrison formula to compute their determinants and inverses, respectively.

2.2.1.1   Determinants and inverses of matrices C and $\mathbf{C}_{\backslash j}$

We decompose the matrix $\mathbf{C} = \begin{bmatrix} \sigma_s^2 + \sigma_1^2 & \rho_s \sigma_s^2 & \cdots & \rho_s \sigma_s^2 \\ \rho_s \sigma_s^2 & \sigma_s^2 + \sigma_2^2 & \cdots & \rho_s \sigma_s^2 \\ \vdots & & \ddots & \vdots \\ \rho_s \sigma_s^2 & \rho_s \sigma_s^2 & \cdots & \sigma_s^2 + \sigma_N^2 \end{bmatrix}$, as

$$
\mathbf{C} = \underbrace{\begin{bmatrix} \sigma_s^2(1 - \rho_s) + \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_s^2(1 - \rho_s) + \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_s^2(1 - \rho_s) + \sigma_N^2 \end{bmatrix}}_{D} + \rho_s \sigma_s^2 \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{J},
$$

and compute the determinant of $\mathbf{C}$ using the matrix determinant lemma in Appendix B.2,

$$|\mathbf{C}| = |D + \rho_s \sigma_s^2 J| = |D + \rho_s \sigma_s^2 \mathbf{1} \mathbf{1}^{\mathrm{T}}| = (1 + \mathbf{1}^{\mathrm{T}} D^{-1} \mathbf{1}) |D|,$$

where $\mathbf{1}$ is a column vector of ones.

Since D is a diagonal matrix, $|D| = \prod_{i=1}^{N} \left(\sigma_s^2 (1 - \rho_s) + \sigma_i^2 \right)$, and

$$(D^{-1})_{k,l} = \begin{cases} \frac{1}{\sigma_s^2 (1 - \rho_s) + \sigma_k^2}, & \text{if } k = l, \\[2ex] 0, & \text{otherwise.} \end{cases}$$

We also define

$$w_i = \frac{1}{\sigma_i^2}, \quad \tilde{w}_i = \frac{1}{\sigma_s^2 (1 - \rho_s) + \sigma_i^2}, \quad \tilde{W} = \sum_{i=1}^{N} \tilde{w}_i, \text{ and } \tilde{W}_{\backslash j} = \sum_{i \neq j}^{N} \tilde{w}_i, \quad (2.15)$$

and obtain

$$|\mathbf{C}| = \left(1 + \rho_s \sigma_s^2 \, \tilde{W} \right) \prod_{i=1}^{N} \frac{1}{\tilde{w}_i}. \quad (2.16)$$

Similarly, we compute

$$|\mathbf{C}_{\backslash \mathbf{j}}| = \left(1 + \rho_s \sigma_s^2 \, \tilde{W}_{\backslash j} \right) \prod_{i \neq j}^{N} \frac{1}{\tilde{w}_i}. \quad (2.17)$$

Next, we compute the inverse of matrix $\mathbf{C}$ using Sherman-Morrison Formula described in Appendix B.2. Specifically, we obtain

$$\mathbf{C}^{-1} = D^{-1} - \frac{D^{-1} \mathbf{1} \, \mathbf{1}^{\mathrm{T}} D^{-1}}{1 + \mathbf{1}^{\mathrm{T}} D^{-1} \mathbf{1}} = \begin{bmatrix} \tilde{w}_1 - \alpha \tilde{w}_1^2 & -\alpha \tilde{w}_1 \tilde{w}_2 & \cdots & -\alpha \tilde{w}_1 \tilde{w}_N \\ -\alpha \tilde{w}_2 \tilde{w}_1 & \tilde{w}_2 - \alpha \tilde{w}_2^2 & \cdots & -\alpha \tilde{w}_2 \tilde{w}_N \\ \vdots & & & \vdots \\ -\alpha \tilde{w}_N \tilde{w}_1 & -\alpha \tilde{w}_N \tilde{w}_2 & \cdots & \tilde{w}_N - \alpha \tilde{w}_N^2 \end{bmatrix},$$

$$(2.18)$$

49

with

$$\alpha = \frac{\rho_s \sigma_s^2}{1 + \rho_s \sigma_s^2 \; \tilde{W}} = \frac{1}{\frac{1}{\rho_s \sigma_s^2} + \tilde{W}}. \tag{2.19}$$

The inverse of matrix $\mathbf{C}_{\backslash \mathbf{j}}$ has the same form as $\mathbf{C}^{-1}$ in Eq. (2.18) except that α is replaced by $\alpha_{\backslash j}$ with

$$\alpha_{\backslash j} = \frac{\rho_s \sigma_s^2}{1 + \rho_s \sigma_s^2 \; \tilde{W}_{\backslash j}} = \frac{1}{\frac{1}{\rho_s \sigma_s^2} + \tilde{W}_{\backslash j}}. \tag{2.20}$$

2.2.1.2 Bayesian decision variable

We continue simplifying the log-likelihood ratio (Eq. (2.14)) in order to obtain an analytically tractable expression for the decision variable, $d_{\mathrm{ST}}(\mathbf{x})$ (Eq. (2.9)). First, we compute the required ratio in Eq. (2.14) of determinants of the two matrices

$$\frac{|\mathbf{C}|}{\sigma_j^2 |\mathbf{C}_{\backslash \mathbf{j}}|} = \frac{w_j |\mathbf{C}|}{|\mathbf{C}_{\backslash \mathbf{j}}|} = \frac{w_j \left(1 + \rho_s \sigma_s^2 \; \tilde{W}\right) \prod\limits_{i}^{N} \frac{1}{\tilde{w}_i}}{\left(1 + \rho_s \sigma_s^2 \; \tilde{W}_{\backslash j}\right) \prod\limits_{i \neq j}^{N} \frac{1}{\tilde{w}_i}} = \frac{w_j \left(1 + \rho_s \sigma_s^2 \; \tilde{W}\right)}{\tilde{w}_j \left(1 + \rho_s \sigma_s^2 \; \tilde{W}_{\backslash j}\right)} = \frac{w_j \; \alpha_{\backslash j}}{\tilde{w}_j \; \alpha}.$$

Next, we compute the exponent terms in Eq. (2.14):

$$(\mathbf{x} - \mathbf{s_D})^{\mathsf{T}} \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s_D}) = \sum_{i=1}^{N} (\tilde{w}_i - \alpha \tilde{w}_i^2)(x_i - s_D)^2 - \alpha \sum_{i \neq k}^{N} \tilde{w}_i \tilde{w}_k (x_i - s_D)(x_k - s_D),$$

$$(\mathbf{x}_{\backslash \mathbf{j}} - \mathbf{s_{D_{\backslash j}}})^{\mathsf{T}} \mathbf{C}_{\backslash \mathbf{j}}^{-1} (\mathbf{x}_{\backslash \mathbf{j}} - \mathbf{s_{D_{\backslash j}}}) = \sum_{i \neq j}^{N} (\tilde{w}_i - \alpha_{\backslash j} \tilde{w}_i^2)(x_i - s_D)^2$$

$$- \alpha_{\backslash j} \sum_{\substack{i \neq k \\ i,k \neq j}}^{N} \tilde{w}_i \tilde{w}_k (x_i - s_D)(x_k - s_D),$$

and combine them to obtain

$$- (\mathbf{x}_{\backslash \mathbf{j}} - \mathbf{s}_{\mathbf{D}_{\backslash j}})^{\mathsf{T}} \mathbf{C}_{\backslash \mathbf{j}}^{-1} (\mathbf{x}_{\backslash \mathbf{j}} - \mathbf{s}_{\mathbf{D}_{\backslash j}}) + (\mathbf{x} - \mathbf{s}_{\mathbf{D}})^{\mathsf{T}} \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}_{\mathbf{D}})$$

$$= (\tilde{w}_j - \alpha \tilde{w}_j^2)(x_j - s_T)^2 - 2\tilde{w}_j \alpha (x_j - s_D) \sum_{i \neq j}^{N} \tilde{w}_i (x_i - s_D)$$

$$+ (\alpha_{\backslash j} - \alpha) \sum_{i,k \neq j}^{N} \tilde{w}_i \tilde{w}_k (x_i - s_D)(x_k - s_D)$$

$$= \frac{\tilde{w}_j \, \alpha}{\alpha_{\backslash j}} (x_j - s_T)^2 - 2\tilde{w}_j \alpha (x_j - s_D) \sum_{i \neq j}^{N} \tilde{w}_i (x_i - s_D) + \tilde{w}_j \alpha \alpha_{\backslash j} \left(\sum_{i \neq j}^{N} \tilde{w}_i \, (x_i - s_D) \right)^2.$$

We substitute the above expressions in Eq. (2.14) to compute the log-likelihood ratio, and therefore, obtain the following expression for the decision variable

$$d_{\mathrm{ST}}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \left(\frac{w_j (1 + \rho_s \sigma_s^2 \tilde{W})}{\tilde{w}_j (1 + \rho_s \sigma_s^2 \tilde{W}_{\backslash j})} \right)^{1/2} \exp \left(-\frac{1}{2} \underbrace{w_j (x_j - s_T)^2}_{I} \right.$$

$$+ \frac{1}{2} \underbrace{(\tilde{w}_j - \alpha \tilde{w}_j^2)(x_j - s_D)^2}_{II} - \underbrace{\tilde{w}_j \alpha (x_j - s_D) \sum_{i \neq j}^{N} \tilde{w}_i (x_i - s_D)}_{III}$$

$$\left. + \frac{1}{2} \underbrace{(\alpha_{\backslash j} - \alpha) \sum_{i,k \neq j}^{N} \tilde{w}_i \tilde{w}_k (x_i - s_D)(x_k - s_D)}_{IV} \right). \tag{2.21}$$

The above equation gives the non-linear decision variable $d_{\mathrm{ST}}(\mathbf{x})$ in terms of the stimulus measurement, and model parameters: the total number of stimuli, N, the variability, and covariability between distractors orientations determined by σ_s^2, and ρ_s. These parameters govern the statistical structure of the visual stimuli. The observer must infer these model parameters in order to make an optimal decision on the task. We assume that the observer is aware of the noise with which the measurement is made.

2.2.2 Interpretation of the decision variable

We note that the decision variable computed in Eq. (2.21) depends in an intricate manner on the model parameters that describe the structure of the stimulus, and its measurements. Although the expression is complex, each term in the exponent has an intuitive interpretation. We can think of the different terms as different pieces of evidence about whether the jth stimulus is a target: (I) if the jth measurement is close to the target orientation, this term is larger (less negative), corresponding to an increased likelihood that the jth stimulus is the target; (II) the second term decreases as the jth measurement approaches the mean distractor orientation, this corresponds to a decreased likelihood that the jth stimulus is the target; (III) the third term compares the jth measurement to the sample distractor mean; if it is large, it is less likely that the jth stimulus is the target; (IV) the fourth term can be rewritten in terms of sample covariance of potential distractor measurements, and in that case a large covariance increases the likelihood that the jth stimulus is the target. Therefore, different terms in Eq. (2.21) contribute towards finding the target.

2.3 MAP estimate distribution

We denote the observer's MAP estimate of T by $\hat{T}$. The optimal observer responds $\hat{T}$ based on the sign of the decision variable $d_{\text{ST}}(\mathbf{x})$ computed in Eq. (2.21). The probability of the optimal observer responding $\hat{T}$ given a fixed stimuli $\mathbf{s}_{\text{fixed}}$ is

denoted by $p(\hat{T}|\mathbf{s}_{\text{fixed}})$. As illustrated in [142], we compute this probability by marginalizing over the hypothesized observations $\mathbf{x}$ generated by $\mathbf{s}$:

$$p(\hat{T}|\mathbf{s}_{\text{fixed}}) = \int p(\hat{T}|\mathbf{x})p(\mathbf{x}|\mathbf{s}_{\text{fixed}})d\mathbf{x} = \int \delta_{\hat{T},\text{sgn}(d_{\text{ST}}(\mathbf{x}))}\ p(\mathbf{x}|\mathbf{s}_{\text{fixed}})d\mathbf{x}. \qquad (2.22)$$

This means that the probability of reporting $\hat{T} = 1$ is evaluated by averaging over the observations that are drawn from the distribution $\mathbf{x}|\mathbf{s}_{\text{fixed}}$, and have $d_{\text{ST}}(x) > 0$. But we note that the decision variable $d_{\text{ST}}(\mathbf{x})$ is a non-linear function of $\mathbf{x}$, and hence, the above expression is analytically intractable. Therefore, we use numerical approximations for our computational purposes. In particular, we apply *Monte Carlo* method that is described in Section 3.2.2.

2.4 Suboptimal models

The computations involved in evaluating the exponent terms in Eq. (2.21) are complex, and require the complete information about the generative model. But in general, an observer may not learn, and use the correct generative model in making decisions. The observer can either use an incorrect assumption about the correlation strength, ρ_s, or may not use an equal odd prior for T. In that case, their inferences will be suboptimal. Therefore, we need to investigate what inference models are used by the observers to make their decisions, and what model parameters have been assumed by them to infer their responses. We study a range of suboptimal models along with the optimal one in Chapter 4. For mathematical purposes, we consider two special variants of the suboptimal models below.

These are the cases of homogeneous, and heterogeneous distractors. The experimental studies [90, 95] have well characterized the human behavior on both these conditions, however, they assumed different probability distributions in the generative model.

2.4.1 Heterogeneous model, $\rho_s = 0$

In this model, we assume that the observer do not learn any information about the stimulus correlations, and make decisions assuming that no structure is present in the scenes. This amounts to the observer using $\rho_s = 0$ in making decisions, therefore, the decision variable in Eq. (2.21) reduces to the following simplified expression in such a case

$$
d_{\mathrm{ST}}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} (1 + w_j \, \sigma_s^2)^{1/2} \exp \left(-\frac{1}{2} \, w_j \, (x_j - s_T)^2 + \frac{1}{2} \, \tilde{w}_j \, (x_j - s_D)^2 \right)
$$

$$
= \log \frac{1}{N} \sum_{j=1}^{N} \left(\frac{\sigma_j^2 + \sigma_s^2}{\sigma_j^2} \right)^{1/2} \exp \left((s_T - s_D) \frac{x_j - \frac{(s_T + s_D)}{2}}{\sigma_j^2 + \sigma_s^2} - \frac{\sigma_s^2 (x_j - s_T)^2}{2 \, \sigma_j^2 \, (\sigma_j^2 + \sigma_s^2)} \right).
$$

$$(2.23)$$

This condition reflects a suboptimal decision, and we would expect the observer following this model to have a low performance.

2.4.2 Homogeneous model, $\rho_s = 1$

We also consider another extreme possibility: the observer may assume that the stimuli are always maximally correlated, and may make decisions using $\rho_s = 1$.

That is, the distractors always have a common orientation. We note that when $\rho_s = 1$, the covariance matrix $\Sigma_\mathbf{s}$ (defined in Eq. (2.3)) is singular. Therefore, the conditions to use Eqs. (B.3) and (B.4) are violated. Hence, we cannot use Eq. (2.21) here.

Instead, we independently compute the decision variable in this case. We note that the covariance matrix $\Sigma_\mathbf{s}$ reduces to a single entry σ_s^2 for $\rho_s = 1$, and the multivariate normal distribution to a one-dimensional Gaussian distribution with mean, s_D, and variance, σ_s^2. This results in having all distractors with an identical orientation, and we denote that common orientation by s. Thus, the common distractor orientation follows the Gaussian distribution

$$s|T \sim \mathcal{N}(s_D, \sigma_s^2).$$

As before, we compute the log-likelihood ratio by marginalizing over the intermediate variables $\mathbf{T}$, and $\mathbf{s}$ as in Eq. (2.11) to obtain

$$L_{\mathrm{ST}}(\mathbf{x}) = \log \frac{p(\mathbf{x}|T=1)}{p(\mathbf{x}|T=0)} = \log \frac{1}{N} \sum_{j=1}^{N} \frac{\int p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}|\mathbf{T}=\mathbf{1}_j) d\mathbf{s}}{\int p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}|\mathbf{T}=\mathbf{0}_N) d\mathbf{s}}.$$

We now use the assumption about the independence (between locations) of measurement noise, and use the product form in Eq. (2.8) for the probability distribution $P(\mathbf{x}|\mathbf{s})$. Also, we decompose the measurement vector $\mathbf{x}$ into a target measurement x_j, and a common distractor measurement, x_i for $i \neq j$. This gives

us

$$L_{\text{ST}}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \frac{\int p(x_j|s_j = s_T)p(s_j|T_j = 1)ds_j \left(\prod_{i \neq j}^{N} p(x_i|s_i = s) \right) p(s|T)ds}{\int \left(\prod_{i=1}^{N} p(x_i|s_i = s) \right) p(s|T)ds}$$

$$= \log \frac{1}{N} \sum_{j=1}^{N} \frac{\int f(x_j; s_j, \sigma_j^2)\delta(s_j - s_T)ds_j \left(\prod_{i \neq j}^{N} f(x_i; s, \sigma_i^2) \right) f(s; s_D, \sigma_s^2)ds}{\int \left(\prod_{i=1}^{N} f(x_i, s, \sigma_i^2) \right) f(s; s_D, \sigma_s^2)ds}.$$

We use Eqs. (B.2), and (B.4) to further compute the expression for the log-likelihood ratio

$$L_{\text{ST}}(\mathbf{x}) = \log \frac{1}{N} \times$$

$$\sum_{j=1}^{N} \frac{\dfrac{f(x_j; s_T, \sigma_j^2)}{\sqrt{(2\pi)^{(N-1)} \left(\frac{1}{\sigma_s^2} + \sum_{i \neq j}^{N} \frac{1}{\sigma_i^2} \right)} \sigma_s \prod_{i \neq j}^{N} \sigma_i} \exp \left(-\frac{1}{2} \left(\sum_{i \neq j}^{N} \frac{x_i^2}{\sigma_i^2} + \frac{s_D^2}{\sigma_s^2} - \dfrac{\left(\sum_{i \neq j}^{N} \frac{x_i}{\sigma_i^2} + \frac{s_D}{\sigma_s^2} \right)^2}{\frac{1}{\sigma_s^2} + \sum_{i \neq j}^{N} \frac{1}{\sigma_i^2}} \right) \right)}{\dfrac{1}{\sqrt{(2\pi)^N \left(\frac{1}{\sigma_s^2} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)} \sigma_s \prod_{i=1}^{N} \sigma_i} \exp \left(-\frac{1}{2} \left(\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} + \frac{s_D^2}{\sigma_s^2} - \dfrac{\left(\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + \frac{s_D}{\sigma_s^2} \right)^2}{\frac{1}{\sigma_s^2} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \right) \right)}.$$

We simplify above expression to obtain

$$
L_{ST}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \sqrt{\left(\frac{\frac{1}{\sigma_s^2} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}}{\frac{1}{\sigma_s^2} + \sum_{i \neq j} \frac{1}{\sigma_i^2}} \right)}
$$

$$
\times \exp\left(-\frac{(x_j - s_T)^2}{2\sigma_j^2} + \frac{x_j^2}{2\sigma_j^2} - \frac{1}{2} \frac{\left(\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + \frac{s_D}{\sigma_s^2} \right)^2}{\frac{1}{\sigma_s^2} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} + \frac{1}{2} \frac{\left(\sum_{i \neq j} \frac{x_i}{\sigma_i^2} + \frac{s_D}{\sigma_s^2} \right)^2}{\frac{1}{\sigma_s^2} + \sum_{i \neq j} \frac{1}{\sigma_i^2}} \right). \quad (2.24)
$$

Assuming the observer uses a uniform prior over T, the above equation represents the decision variable under the assumption of heterogeneous distractors. Using the variables defined in Eqs. (2.15), and (2.19) we rewrite the following compact expression for the decision variable in the case of $\rho_s = 1$

$$
d_{ST}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \sqrt{(1 + w_j \alpha_{\backslash j})} \times
$$

$$
\exp\left(-\frac{1}{2} \left(w_j(x_j - s_T)^2 - w_j x_j^2 + \alpha \left(\sum_{i=1}^{N} w_i x_i + \frac{s_D}{\sigma_s^2} \right)^2 - \alpha_{\backslash j} \left(\sum_{i \neq j} w_i x_i + \frac{s_D}{\sigma_s^2} \right)^2 \right) \right).
$$

$$
(2.25)
$$

The above decision variable characterizes the decision-making behavior of an observer that assumes the distractors to be always perfectly correlated. The observer making decisions based on this variable will be suboptimal since the correct correlation strength ρ_s is ignored, and always assumed as 1.

The homogeneous, and heterogeneous models explained above are two particular suboptimal models. We consider a range of other possible suboptimal models

in Chatper 4. We now provide the details of the experiment designed on the target detection task. We performed the experiment to collect human subjects data on the task.

2.5 Experimental Methods

We conducted an experiment based on the design of the target detection task described in Section 2.1. The experiment was performed under the supervision of Dr. Wei Ji Ma in the Theoretical Systems Neuroscience laboratory at the Department of Neuroscience, Baylor College of Medicine, Houston, Texas, USA.

The aim of our experimental study was to determine whether human observers use the structures present in the visual scenes to infer their decisions. In the case that they do, we want to examine whether they are able to infer the correct correlation strength ρ_s that is used to generate the experimental displays. If not, we test several suboptimal models in Chapter 4, and find the one that best explain the responses of the subjects on the experiment. In the following section, we provide the details of the experiment.

2.5.1 Subjects

Eleven subjects (6 male, 5 female) participated in the experiment. All subjects had normal or corrected-to-normal acuity and gave informed consent.

2.5.2   Apparatus and stimuli

Stimuli were generated in Matlab using the Psychophysics Toolbox [23, 109], and were presented on a 21″ LCD monitor with a refresh rate of 60 Hz. Subjects viewed the displays from a distance of approximately 60 cm. The background luminance was 33.1 cd/m^2. A set of 4 stimuli were shown on each trial. On target-present trials, the stimulus set consisted of one target, and 3 distractors while on target-absent trials, it contained 4 distractors. A target was present in exactly half of the trials. Each stimulus was a Gabor patch (or a Gabor filter) with a spatial frequency of approximately 2.67 cycles/deg, a standard deviation of 0.26 deg, and a peak luminance of 136 cd/m^2. A Gabor patch is a sine wave multiplied by a Gaussian function (additional details can be found in Appendix C). Figure 2.2(A) shows one example of a Gabor patch. Stimuli were placed on a circle centered at the fixation cross with a radius of 3.2 degrees of visual angle. The position of the first stimulus was chosen at random on each trial, and other stimuli were placed in a way so that the angular distance between two adjacent stimuli was always $45°$. The target (s_T) and mean distractor orientation (s_D) were set to vertical, and used to define the origin. The standard deviation of the distractor distribution σ_s was fixed at $15°$ while the correlation coefficient ρ_s was varied to be $0, \frac{1}{3}, \frac{2}{3}$, and 1 across different experimental sessions.

2.5.3   Sessions and blocks

The experiment was split into four different sessions. The correlation coefficient was fixed at $0, 1/3, 2/3$, or 1 within a session. The order of the sessions was randomized across subjects. Each session consisted of one practice block of 50 trials and 6 testing blocks of 150 trials each, and lasted for about 50 minutes. A 30 sec break was provided between blocks. After each block, performance on that block was revealed to the subject along with the scores of the other subjects who had completed the same session. Each subject completed a total of 3600 test trials. All subjects were instructed about the experiment at the beginning of their first session with a demo consisting of 10 practice trials.

2.5.4   Procedure

Testing trials: Each test trial began with the display of a fixation cross at the center of the screen (0.5 sec), followed by the stimulus display containing 4 stimuli (0.1 sec), and followed by a screen with the fixation cross until the subject responded. The subject reported whether a target was present or absent through a key press. Feedback was provided by subsequently coloring the fixation cross green (correct) or red (incorrect) during the (0.75 sec) inter-trial period. The experiment, and time course are shown in Figure 2.2(B).

Practice trials: Each practice trial was identical to a test trial, except that it was followed by a feedback screen showing the original stimulus display with a blue
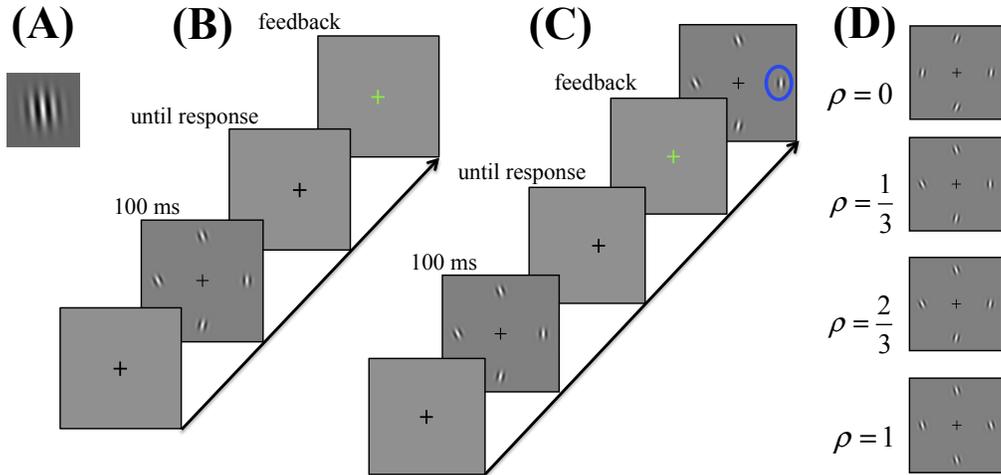
Figure 2.2: **Target detection experiment procedure. (A) Gabor filter.** Subjects were presented with 4 stimuli on each trial. Each stimulus was a Gabor patch. The figure illustrates an example of a Gabor patch from the experiment. **(B) Time course of a test trial.** The experiment started with a display of fixation cross in the center followed by the stimulus display for 0.1 sec. Subjects reported through a key press whether a vertical stimulus was present in the display. After their response, a display screen was shown with a green (correct) or red (incorrect) fixation cross to provide feedback. **(C) Time course of a practice trial.** Each experimental session started with 50 practice trials. The procedure of a practice trial was same as a test trial, except that an additional feedback screen was shown for 2 sec at the end of the trial. The extra display contained the original stimulus with a blue circle marking the target stimulus if it was present. **(D) Sample displays of different experimental conditions.** The experiment was divided in four different sessions. Each session was characterized by the unique value of the correlation coefficient $\rho_s \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$ used in generating stimuli. The order of the sessions was randomized across subjects. This figure shows example displays from each experimental session.

circle identifying the target stimulus, when present (Figure 2.2(C)). The data from practice trials was excluded for the analysis, and results purposes.

Chapter 3

Model fitting and model comparison

Our perception about the true state of the world is based on our sensory information along with the prior knowledge of the world state. But it is not clear what computations the brain performs to combine the information it receives, and makes a decision. Numerous theoretical, and experimental studies provide evidence that our perception can be described as a process of probabilistic inference. In particular, Bayesian models are applied to study human behavior on various perceptual tasks. Several experimental studies have found that humans are near-optimal in simple visual perceptual tasks [64, 40, 8, 39, 95, 107, 7, 101, 90, 10, 79, 144, 94]. We note that the observers needs to have complete knowledge of the underlying generative model of the task in order to make the best possible decisions. However, we may expect that on a complex task, or a task with a large number of latent variables, it may not always be possible for the observers to determine the correct generative model. Therefore, the relevant question here is what model of

the world observers use to make their decisions, and what are the possible techniques that we can use to compute the predictions of a model for the experimental data.

In this chapter, we discuss some commonly used techniques of fitting a model to the responses of subjects on a psychophysical task. We describe the maximum-likelihood parameter estimation method to find the model parameters that describe the best fit to an experimental data. Further, we discuss the criteria that can be used to compare models, and find the one that best describes the data. We present these techniques in a generalized form here, but discuss their applications, and possible issues in the context of our experimental study (described in Chapter 2). Specific to our experiment, we begin this chapter with a detailed description of the experimental data we have collected. Further, we elaborate the types of psychometric curves we use to represent the subjects' responses from our experiment. In the end, we describe how we use different model comparison criteria specific to the models used to predict our experimental data.

3.1   Experimental data

The set up of our experiment has been described in Section 2.5. Subjects were presented with a set of 4 stimuli on each experimental trial. The orientations of the stimuli were drawn randomly across trials. Also, the stimuli were placed at random locations. As a consequence, each subject was presented with a unique set of stimuli over the course of the experiment. We thus recorded the collection

of stimulus orientations presented to the subject. We then recorded whether any stimulus is a target in the trial, that is, we noted if $T = 1$, or $T = 0$. Furthermore, we documented the response of the subject in each trial as either 1 or 0. Finally, we measured the exact duration of the stimulus presentation on the display screen. We note that this duration is actually a constant (100 ms) in the experiment. In summary, for each subject we recorded the following information on each trial:

1. a set of 4 stimulus orientations, $\mathbf{s}$,

2. the information about the target presence variable T,

3. the subject's response or the MAP estimate $\hat{T}$,

4. whether $\hat{T}$ matches T or not,

5. the duration of the stimulus presentation.

In the following section, we discuss the methods of obtaining model predictions for an experimental data, and to use these predictions for fitting the subject's responses. In our analysis, we evaluate the predictions of a model for each individual instead of relying on average (over subjects) statistics. Therefore, our model fitting, and model comparison processes are computed based on individual responses.

3.2 Model predictions

In the experiment, we record the responses of a subject - whether the subject responded "target present" or "absent" on each trial. However, we are not aware of the sensory measurements with which the subject made a decision on a trial. Thus, as part of the modeling process, we need to consider what measurements would have led the subject to make a particular response on the trial. Hence, we must use a reasonable assumption about the distribution of the measurements. Moreover, in general, we have no means of measuring what parameters the subject would have used to make the decision. For instance, the sensory noise with which the subject made the measurement is unknown to us, and we need to estimate it in order to understand the behavior of the subject. Such parameters are sometimes referred as *free parameters* of the model, and can be estimated from the data. Therefore, we consider two important issues here: to find the maximum-likelihood estimates of model parameters based on subject's data, and to make predictions of the model for the data given those parameters.

Let us consider a model M having a parameter θ. The parameter θ can either be a scalar or vector quantity given the model. We assume that we have access to a subject's responses on K trials in the experiment. We denote the subject's response by a binary variable r_i, and the set of presented stimuli by $\mathbf{s}_i$ on the ith trial. We are interested in evaluating the model prediction for the response r_i on i^{th} trial given the model parameter θ. We denote $p(r_i|\mathbf{s}_i, M, \theta)$ as the probability of the subject response r_i given the stimuli on the ith trial under the model M with hypothesized

parameter value $\boldsymbol{\theta}$. This is also known as the *response probability* under model M with parameter $\boldsymbol{\theta}$.

3.2.1 Computing response probabilities

We would like to compute the response probability $p(r_i|\mathbf{s}_i, M, \boldsymbol{\theta})$ on the ith trial under model M. This is similar to evaluating the distribution of the MAP estimate in Eq. (2.22) under the decision variable of the model M. The cases where such an evaluation is not analytically possible, we use Monte Carlo method [58, 11, 56, 68, 120, 15] for a numerical approximation of the involved integral. The integral is approximated by the sum which converges to the correct value as the number of measurement samples increases:

$$
\begin{aligned}
p(r_i|\mathbf{s}_i, M, \boldsymbol{\theta}) = \int p(r_i|\mathbf{x})p(\mathbf{x}_i|\mathbf{s}_i, M, \boldsymbol{\theta})d\mathbf{x} &\approx \sum_{\mathbf{x}_i} p(r_i|\mathbf{x})p(\mathbf{x}_i|\mathbf{s}_i, M, \boldsymbol{\theta}) \\
&= \sum_{\mathbf{x}_i} \delta_{r_i, \text{sgn}(d_M(\mathbf{x}_i))} p(\mathbf{x}_i|\mathbf{s}_i, M, \boldsymbol{\theta}). \quad (3.1)
\end{aligned}
$$

Here $\mathbf{x}_i$ is the hypothesized measurement of the subject on the ith trial, and $d_M(\mathbf{x})$ represents the Bayesian decision variable under the model M with parameter $\boldsymbol{\theta}$.

3.2.2 Monte Carlo algorithm

We now describe the Monte Carlo algorithm we implemented for computing the response probabilities under the assumptions of a model M.

1. The first step is to describe the model M, and its parameter(s) $\boldsymbol{\theta}$.

2. We then derive the Bayesian decision variable $d_M(\mathbf{x})$ under the hypothesis of the model M.

3. We fix a value of the model parameter $\boldsymbol{\theta}$.

4. For this fixed hypothesized value of the parameter $\boldsymbol{\theta}$, we draw hypothesized measurements $\mathbf{x}_i$ of the subject following Eq. (2.8) on the ith trial in response to the presented stimulus $\mathbf{s}_i$. The stimuli $\mathbf{s}_i$ are used from the experimental data, and not generated during this step.

5. Using the hypothesized measurements $\mathbf{x}_i$, we evaluate the decision variable $d_M(\mathbf{x}_i)$ of the model M. Further, we compute the prediction for the subject's response, denoted by $\hat{r}_i$ on the ith trial based on the decision rule for the model M.

6. We then compare the model predicted response $\hat{r}_i$ with the subject's true response r_i on the ith trial. A match between the two results in an increased probability of the response $p(r_i|\mathbf{s}_i, M, \boldsymbol{\theta})$.

7. We repeat steps 4 to 6 with R samples of measurements $\mathbf{x}_i$, and thus compute R values of model predicted responses $\hat{r}_i$. We subsequently match each of them with the corresponding actual response r_i of the subject.

8. We obtain an approximation of the probability of response $p(r_i|\mathbf{s}_i, M, \boldsymbol{\theta})$ by averaging over the number of correct matches in R samples.

9. Next, we pick a different value of $\boldsymbol{\theta}$, and evaluate the response probability on the ith trial for a different parameter value by following steps 4 to 8.

Steps 3 to 9 are repeated for every experimental trial to obtain the likelihood function of the subject's actual responses under the model M with parameter θ.

3.3   Maximum-likelihood estimation

Our aim is to find the value of the model parameter θ that maximizes the response probabilities over all trials. In other words, we are interested in finding the model parameter at which the predictions of the model provide the best possible explanation for the behavior of subjects in an experiment. This amounts to finding the maximum-likelihood estimate of θ. Below, we discuss the parameter estimation method illustrated in [89, 90, 94]. The likelihood function of a parameter value θ is defined as the probability of the data given the model M with parameter θ:

$$L_M(\theta) = p(\text{data}|M, \theta).$$

For simplification purposes, it is generally assumed that the noise in the observer's responses is independent across trials. Thus, we can write the probability of the data given the model, and its parameters as a product of probabilities over trials:

$$L_M(\theta) = \prod_{i=1}^{K} p(r_i|\mathbf{s}_i, M, \theta). \tag{3.2}$$

Here K denotes the total number of trials in the experiment. Maximizing the above parameter likelihood is equivalent to maximizing its logarithm $\log L_M(\theta)$,

$$\log L_M(\theta) = \sum_{i=1}^{K} \log p(r_i|\mathbf{s}_i, M, \theta). \tag{3.3}$$

The logarithm prevents possible numerical issues which could arise because of a very small probability that could mistakenly be treated as zero, and could lead

to potential errors in the computation of the product in Eq. (3.2). We denote the maximum-likelihood estimate (MLE) of parameter θ by $\hat{\theta}$, and the corresponding maximum value of the likelihood function as $L_M^* = \max_{\theta} L_M(\theta) = L_M(\hat{\theta})$.

In the event that an analytical expression is unavailable for the response probability $p(r_i|\mathbf{s}_i, M, \theta)$ under the model M, and hence, for the log-likelihood function in Eq. (3.3), we evaluate the function using numerical methods. Therefore, most of our data analysis practices rely on obtaining an accurate approximation of the log-likelihood function. As expected, the accuracy of this evaluation depends on the size of the data, and the number of measurement samples used in the numerical approximation. Further, we need to find an appropriate numerical algorithm to find the maximum of the numerically evaluated stochastic log-likelihood function $L_M(\theta)$. We use suitable optimization algorithms to find the maximum-likelihood estimate of the parameter θ that maximizes the log-likelihood function defined in Eq. (3.3).

In our target detection task (Chapter 2), we do not have any analytical approximation of the log-likelihood function for any model. Therefore, we numerically evaluate the function for a model M at all hypothesized values of the model parameter θ. But the evaluation of our function $\log L_M(\theta)$ depends on the decision variable $d_{\text{ST}}(\mathbf{x})$ computed in Eq. (2.21), and thus it is a non-linear, non-smooth stochastic function. To maximize this stochastic function, we tested the following optimization algorithms, and chose the one that was most suitable for our purposes.

3.4 Optimization techniques

In order to find the maximum of our non-linear stochastic objective function $L_M(\boldsymbol{\theta})$, we tried several different optimization algorithms, and compared the obtained results. We checked the common practice of grid search method, and the standard techniques such as *genetic algorithm*, and *pattern search*. However, most of our models (described in Chapter 4) were high-dimensional in parameter space, and thus optimization algorithms required tremendous amount of computational power, and time to produce results. This restricted us to only use grid search method for our results purposes. In the following section, we briefly describe the three algorithms we had tested for our data.

3.4.1 Exhaustive or grid search

Exhaustive or brute-force is one of the simplest possible methods of optimizing an objective function by manually evaluating it on a predefined set of parameter space. The possible parameter space is identified, and systematically divided into possibly a large set of discrete values known as the *grid space*. The objective function is then evaluated at all points of the pre-defined grid space, and the point at which the function attains the global maximum is regarded as the maximum-likelihood estimate of the parameter. The accuracy of this method greatly depends on the choice of the parameter space, and the grid spacing.

Most of our models (described in Chapter 4) have a high-dimensional parameter vector θ. This resulted in a multi-dimensional grid spacing for θ, and also a large number of grid points. Also, the number of evaluations of the log-likelihood function has an exponential increase with respect to finer discretization of the grid space. Moreover, our likelihood function is evaluated using Monte Carlo method (see Eqs. (3.1), and (3.3)), and require a large number of measurement samples to guarantee convergence. As a consequence, finding maximum-likelihood parameter estimates for our models requires a large amount of computational resources, and time. Furthermore, the precision of the results depends on the convergence of the log-likelihood evaluation, and the discretization of the grid space.

However, we still use grid search method over other optimization algorithms to obtain our results (in Chapter 4). It is because this method can be run in parallel for each grid point as the evaluation of the log-likelihood function for our models is independent between grid points. Therefore, we can numerically evaluate the function at multiple grid points at the same time, and combine the results together. To be able to run in parallel is the most crucial feature of this exhaustive search that led us to use this method for our analysis purposes. Also, we obtained consistent results with this approach. We performed several tests on synthetic data sets to ensure the accuracy, and performance of the algorithm.

3.4.2   Genetic algorithm and pattern search

Genetic algorithm [99, 134, 49] belongs to a larger class of optimization search algorithms, known as evolutionary algorithms that work on the principle of natural selection. An initial population of the parameter θ is (randomly) chosen, and the objective function is evaluated for each individual in the parameter population. The value of the objective function usually represents the fitness of the individual. A fixed number of fittest members of the parameter population are stochastically chosen to be evolved at the next iteration. The parameter is usually characterized by specific features or properties, which are then mutated or altered to obtain a new population or generation of the parameter at the next iteration. The process is repeated until a desired fitness or maximum value of the objective function is achieved, or predefined maximum number of generations is reached.

Genetic algorithm is suitable for both constrained, and unconstrained optimization problems. It is specifically used for discontinuous, non-differentiable stochastic functions to obtain a global extremum, however, it has a slow convergence rate, and requires sufficiently large number of iterations to converge.

On the other hand, pattern search [19, 63, 137, 85] converges quickly to the solution. The search is initiated by evaluating the objective function at an initial value of the parameter. Then, successively neighboring parameter points are found for which the value of the objective function increases. The new parameter points are found using a mesh around the previous ones, and the search continues until a maximum number of iterations is reached or mesh size is too small.

Pattern search belongs to the family of direct search algorithms, and is applied for discontinuous or non-differentiable objective functions.

We tested both genetic algorithm, and pattern search to optimize the log-likelihood function of our models. In the case of simple models having less than 4 parameters, the results were consistent with those obtained using the grid search method. However, the algorithms did not work for high dimensional models. The involved computations makes it difficult for the algorithms to be run in parallel, and in general, they require a huge amount of computational time, and resources. Though these methods are certainly better, and more precise than grid search algorithm, we were unable to obtain results using them for our models. Hence, our analysis is only based on the results obtained using grid search optimization algorithm.

Thus far, we discussed the procedure to obtain the predictions of a model given the data. We summarized the estimation method for model parameters, and different optimization techniques that can be used for numerical estimation. We now focus on the visualization of our experimental data. We described the format of our collected data on the experiment in Section 3.1, but the question is how we could plot the data for analysis purposes. In the following section, we discuss the different types of *psychometric curves* we have used to analyze our data. Psychometric curves are frequently used in psychophysics to study the responses of subjects on an experiment. The predictions of a model are generated for these curves using model fitting techniques (discussed in Section 3.2). The error between a subject's actual psychometric curve, and the model predicted curve provides a measure of

goodness of a model fit.

3.5 Psychometric curves

Psychometric curves [147, 73] are extensively used in psychophysics to represent the summary of subjects' responses on an experiment. They describe the behavior of subjects over a range of stimulus values. The curve is determined by the number of alternative choices in the task, for example, a psychophysical task can have binary choice, two-alternative forced choice (2AFC), or n-alternative choices. We are mainly interested in the curves having binary choices as our target detection task has a yes/no paradigm.

For our analysis purposes, we examine the behavior of subjects on our experiment using three different types of psychometric curves. We note that our experiment was divided into four sessions (details in Section 2.5), and each session was characterized by the unique value of correlation coefficient ρ_s used to generate the stimuli. The purpose of our study is to determine whether subjects learn the structure present in the scenes, and infer the correct correlation strength to make their decisions. Therefore, correlation coefficient ρ_s serves as the physical stimulus parameter of our interest, and we specify our results in terms of this parameter. We consider three types of psychometric curves based on different characterizations of stimuli as a function of ρ_s.

3.5.1 Types of psychometric curves

We consider the following three types of psychometric curves to analyze the responses of subjects on our target detection experiment described in Chapter 2:

(I) **Hit and false-alarm rates:** In a target detection task, the probability of reporting "target present" when the target is present is known as the *hit rate*, or *detection rate*, or *true positive rate*, and is denoted by $P(\hat{T} = 1|T = 1)$. Whereas, the probability of reporting "target present" when the target is absent is known as the *false-alarm rate* or *false-positive rate*, and is denoted by $P(\hat{T} = 1|T = 0)$. The origin of these terminologies lies in the *signal detection theory* [53, 93, 148, 69, 97]. Further, we can obtain *miss rate* or *false negative rate*, and *correct rejection rate* or *true negative rate* by subtracting the hit, and false-alarm rates, respectively from 1.

We plot the hit, and false-alarm rates as a function of correlation strength ρ_s in four different experimental conditions. We also compare the performance of subjects as the strength of ρ_s varies in the experiment.

(II) We also plot the **proportion of "target present" responses as a function of minimum difference between the target and a distractor's orientation**. We plot these curves separately for the target-present, and target-absent trials.

In a target absent trial, as the distractor orientation gets closer to the target orientation, the difference between the two reduces, and it would be difficult for an observer to discriminate the distractor from the target. The distractor would appear as a target in such a case, and the observer is more likely to

report "target present" even though there is no target. On the other hand, as the minimum difference increases, the distractors will have significantly different orientations than the target, and the observer may be able to detect the absence of target easily.

While on a target present trial, a comparatively small orientation difference between the target, and a distractor could provide more evidence to the observer in responding "target present". This is because more than one object in the display would have orientation closer to the target, and the observer could possibly make a decision based on either one of them. In the case of large minimum difference between the target, and a distractor, the observer must make a decision based on the measurement of the target.

(III) Further, we analyze the behavior of subjects using a plot of **proportion of "target present" responses as a function of sample standard deviation between distractor orientations**. Again, we separate the data on target-present, and target-absent trials. In terms of interpretation, these curves are closely related to type (II) curves with x-axis as the minimum orientation difference between target, and a distractor. Also, the curve follows a similar shape as its counterpart in type (II) except in the case of perfect correlations. In the case of $\rho_s = 1$, all distractors are identical, and hence the standard deviation between any pair of distractor orientation is equal. Thus, all data points lie in a single bin on target absent trials, and we only obtain a single point in the plot. For instance, left panel in Figure 4.1(C), and other similar figures in Chapter 4 have such curves.

3.5.2 Predictions using synthetic data

We want to fit the psychometric curves of subjects on the experiment with a hypothesized model M. We use the maximum-likelihood estimate $\hat{\theta}$ of model parameter θ to generate the predictions of the model for a subject's data. We assume that $\hat{\theta}$ represents the subject's parameter, and can be used to completely describe the behavior of the subject. We use this estimate to generate the stimuli, and the corresponding hypothesized measurements of the subject. These measurements are then used to make hypothesized responses of the subject based on the decision variable $d_M(\mathbf{x})$ of the underlying model. This constitutes the model predicted data set for the subject, and is also referred as the *synthetic*, or *fake data set*. The synthetic data is generated to replicate the subject's behavior based on the assumptions of the model M. These data sets are then used to make model generated psychometric curves. The psychometric curves predicted by the model with the parameter estimate of $\hat{\theta}$ are compared with the actual psychometric curves of the subject on the experiment. If the model predicted curves are well fitted to the subject's true responses, then the model provides a good explanation of the behavior of the subject on the experiment. If the two curves are significantly different, then the underlying model lacks the assumptions that could reproduce the subject's responses.

It is possible that the responses of the subjects are best described by multiple models, or behavior of different individuals is explained by different models. We find such issues in Chapters 4, and 5. But, we use summary statistics to draw

conclusions. We generate the model predictions for each subject, and take an average over all subjects. These average model predictions are then matched with averaged true psychometric curves of the subjects. The results based on summary statistics may not reflect a complete picture in the event subjects follow different strategies or models to make decisions. However, it is difficult to build a mathematical model to test such a possibility.

3.5.3 Error measures

The model provides a prediction for the subject's data based on the maximum-likelihood estimate of parameter θ. The subject's psychometric curves are fitted using the model predicted curves. To quantify the difference between both curves, we measure two types of statistical errors: (i) the *root-mean-square error (RMSE)*, and (ii) the R^2 *statistic* or *coefficient of determination*. RMSE is defined as the square root of the mean square error between the subject's data, y_i, and the model predicted curves, $\hat{y}_i$:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{D_N} (y_i - \hat{y}_i)^2}{D_N}}.$$

Here D_N denotes the total number of data points in the psychometric curve.

Statistic R^2 is another measure of determining how well the model fits to the

experimental data, and it is defined as:

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{D_N}(y_i - \hat{y}_i)^2}{\sum\limits_{i=1}^{D_N}(y_i - \bar{y})^2},$$

where $\bar{y} = \dfrac{1}{D_N}\sum\limits_{i=1}^{D_N} y_i$ is the sample mean of subject's data. The range of R^2 depends on the type of regression used; however, a negative value can occur in case a non-linear function is fitted to the data [29].

3.6 Model comparison

The fitting of model predicted psychometric curves to a subject's responses determines the goodness of the model to the experimental data, and we measure this goodness in terms of the statistical error between the two curves. Frequently, we consider multiple models with different assumptions about the behavior of the subject. Models generally differ in terms of the assumptions about their parameters, and the dimensionality of the parameter space. We fit each hypothesized model to the data using a similar fitting procedure. It is possible that more than one model provide a good explanation for the subject's behavior. They could have equally well predicted curves that match the experimental data, and the error is comparable for both models.

How do we compare models to find the one that best describes the data? Psychometric curves do not help in discriminating two models if both predict similar

fitting curves for the data. Instead, we compare models on standard criteria that also weigh models based on their dimensionality. A general model with large number of parameters will always provide a fit that is at least as good as a special model obtained from the general case. This could mainly be because of additional free parameters in the general model. Therefore, we consider criteria that take this issue into account, and penalize a model for the number of free parameters it has. In the following section, we briefly discuss three criteria that are commonly used to compare models in pscyhophysics. We also discuss the limitations we encountered while comparing our models (described in Chapter 5) based on these criteria.

3.6.1 Bayesian model comparison (BMC)

Bayesian model comparison (BMC) [92, 89, 146] is a fundamental method of model selection. The model with the highest posterior probability (probability of the model given the data) is selected. The posterior probability of a model M given the data can be computed using Bayes' theorem

$$p(M|\text{data}) = \frac{p(\text{data}|M)p(M)}{p(\text{data})}.$$

The ratio of posterior probabilities is computed for the two models M_1, and M_2 that need to be compared

$$\frac{p(M_1|\text{data})}{p(M_2|\text{data})} = \frac{p(\text{data}|M_1)p(M_1)}{p(\text{data}|M_2)p(M_2)}.$$

An equal prior is chosen for both models, $p(M_1) = p(M_2) = \frac{1}{2}$ since no model is favored over the other. This reduces the comparison to the ratio of *model likelihoods*, $p(data|M)$. The ratio is also called a *Bayes' factor*. The *model likelihood* is computed [89, 94] by averaging the model likelihood under a hypothesized value of the parameter in the parameter space

$$p(\text{data}|M) = \int p(\text{data}|M, \boldsymbol{\theta}) p(\boldsymbol{\theta}|M) d\boldsymbol{\theta}. \tag{3.4}$$

The above expression is rewritten in terms by taking logarithm

$$\log p(\text{data}|M) = \log L_M^* + \log \int e^{\log L_M(\boldsymbol{\theta}) - \log L_M^*} p(\boldsymbol{\theta}|M) d\boldsymbol{\theta}. \tag{3.5}$$

Bayesian model comparison computed using the above equation is based on the entire model likelihood function $L_M(\boldsymbol{\theta})$, instead of only of its maximum value L_M^*. Further, it penalizes models for additional free parameters. Therefore, BMC is a principal method to compare models.

However, there are several issues in the practical implementation of BMC. Eq. (3.5) is based on the integral evaluation of the parameter likelihood function $L_M(\boldsymbol{\theta})$ over the entire parameter space. This integral does not have a closed form expression in case of analytically intractable model likelihood. Thus in practice, the integral is approximated by Riemann sums. The accuracy of such an approximation depends on many factors. For instance, the parameter grid spacing, and number of samples used in the evaluation can potentially change the results to a large extent. Further, the sum may not converge to the true value, and could suffer numerical issues. High-dimensional models are most sensitive to these problems,

and precaution must be taken when using Bayesian model comparison for complex models.

Most of the models that we considered to explain our experimental data are high-dimensional in parameter space. Thus, we faced computational issues using BMC to compare our models. In general, the integral in Eq. (3.5) can be evaluated using *Markov Chain Monte Carlo (MCMC)* method [47, 115, 92, 103, 18]. However, we also encountered problems in implementing this method to apply BMC in our case. Lack of an analytical approximation for the model likelihood function $L_M(\theta)$ created most of the numerical problems for us. Due to computational inefficiency of our models, we use criteria that use maximum model likelihoods to compare models. We describe two such comparison measures below.

3.6.2 Bayesian information criterion (BIC)

Bayesian information criterion (BIC) [123] is a model selection criterion that is based on the maximum value of model likelihoods L_M^*. It computes the goodness of a model from maximum model likelihood while penalizing the model for extra free parameters. The penalty term increases with the number of free parameters in the model. BIC is mainly computed using the following formula

$$\text{BIC} = -2L_M^* + f \log K, \tag{3.6}$$

where f is the total number of free parameters in the model, and K represents the total number of experiment trials. The above formula is based on a Laplace's approximation of the integral in Eq. (3.5). Bhat et al. [17] have provided a detailed

derivation of the same.

The penalty term in BIC increases with increasing number of free parameters, f, and it scales with the size of the data. Thus, it more strongly penalizes models with large number of free parameters as compared to the *Akaike information criterion (AIC)*. Therefore, the model with the lower value of BIC is preferred when using Bayesian information criterion to compare models. A lower value of BIC for a model implies fewer free parameters, better fit or both.

3.6.3 Akaike information criterion (AIC)

Similar to BMC, and BIC, Akaike information criterion(AIC) [3, 20, 21, 25] is also a relative measure of a statistical model quality. It is closely related to BIC in terms of the criterion, and the penalty term. However, the penalty term is weaker as compared to BIC. The following equation is generally used to compute Akaike information criterion:

$$\text{AIC} = -2L_M^* + 2f. \tag{3.7}$$

Under this criterion of model selection, the model with the minimum value of AIC is preferred over others, and it also penalizes models for over fitting.

The derivation of the criterion in Eq. (3.7) can be found in [3]. It is mainly based on the information theoretic concept, and computing the *Kullback-Leibler (KL) divergence* [81, 25, 18] between a model prediction, and the data. The model with the minimum value of KL distance, or having minimized information loss is selected.

We note that both BIC, and AIC are only based on maximum model likelihoods. Thus, this requires finding the maximum-likelihood estimates, however, we do not need to compute any integrals here. Therefore, to overcome our numerical difficulties, we compare our models using BIC, and AIC in Chapter 5, and make conclusions based on the results obtained from these comparisons.

3.7 Discussion

In this chapter, we have presented a brief description of the tools that are frequently used in fitting models to the psychophysics experimental data. Estimating maximum-likelihood parameters of a model is a crucial step in order to generate predictions of the model for a subject's responses. However, the possible numerical issues must be diagnosed that are specific to the task, and the data. A suitable, and efficient optimization algorithm must be applied to estimate model parameters. Lack of a better approximation of the underlying model likelihood, or an inefficient method of obtaining this approximation can limit the application of an optimization algorithm to the data.

The parameter estimates are then used to generate the model predictions for the psychometric curves of the subject's responses. These curves are analyzed to determine the goodness of the model fit to the data, and how well the model explains the behavioral response of the subject on the experiment. Eventually, different models are compared on an appropriate criterion to find the model that best describes the data.

Chapter 4

Data analysis I: model fitting

Recent experimental studies have found that humans behave in Bayes-optimal manner in detecting a target among homogeneous, and heterogeneous distractors [90, 94, 95]. But both these conditions are extremes, and do not represent the possible realistic intermediate structure present in natural scenes. This provides the motivation for our experimental study to interpolate between the homogeneous, and heterogeneous conditions by varying the strength of correlation between distractor orientations. The correlation coefficient ρ_s defined in the generative model in Section 2.1 controls the amount of correlations between pairs of distractors. A higher correlation between distractors introduces more similarity in their orientations, and hence gives more structured stimuli. To examine the effect of structured input on visual perception, we designed a target detection task with distractors having correlated orientations (details in Section 2.1), and performed an experiment based on the task to determine how human subjects make

decisions in response to structured stimuli. The information about the recorded experimental data is provided in Section 3.1.

We provide several plausible models here, and use rigorous model comparison to decide which one explains the data best. Further, we interpret the behavior of subjects based on the assumptions of the best fitting model, and its parameters. We mainly focus on understanding the impact of stimulus correlations on subjects' responses. We note that the correlation coefficient ρ_s is an experimental parameter that controls the amount of correlations between distractor orientations. It is also the relevant stimulus feature of our interest. We controlled the statistical structure in visual scenes by varying the strength of stimulus correlations ρ_s. We chose 4 different levels of ρ_s - $0, \frac{1}{3}, \frac{2}{3}$, and 1 to be tested in the experiment. The condition $\rho_s = 0$ corresponds to heterogeneous distractors, and $\rho_s = 1$ with perfectly correlated (or homogeneous) distractors. The intermediate value of $\rho_s \in \{\frac{1}{3}, \frac{2}{3}\}$ interpolates between the two extreme conditions, and introduces a partial structure in the scenes. The experiment was divided in four different sessions (details in Section 2.5.3), and a unique value of ρ_s was used to generate the stimuli in each session. Therefore, each session was characterized by a unique statistical structure present in the displays, and represents a different condition on the experiment. We thus refer to these sessions as *experimental conditions*, and the difference between these conditions is determined by the experimental value of ρ_s used in that session.

Our objective is to examine the subjects' behavior on different experimental conditions, and determine whether subjects use correct or incorrect assumption

about the experimental value of ρ_s in different conditions. In addition, we are not aware of the precision levels with which subjects made their measurements. It is evident from Eq. (2.21) that precision (inverse of measurement noise σ_i^2) of a measurement can greatly affect the accuracy of a subject's decisions. In order to generate model predictions for the data, we also need to estimate the precision parameters that regulate the subject's measurements. Therefore, our models have two important category of parameters – the subject's assumption about ρ_s, and the uncertainty in the subject's measurements.

We present a variety of models in this chapter that differ in their assumption about ρ_s, and the encoding precision. We consider a range of assumptions about both parameters in our models. Each model is individually tested on the experimental data, and model predications are generated for subjects' psychometric curves. We use the methods, and procedures described in Chapter 3 for finding parameter estimates, and fitting a model to the data. We evaluate the error (RMSE), and goodness of fit (R^2) between model predicted curves, and the data. Finally, we discuss why we need to compare models.

4.1 Psychometric curves of subjects data

We first examine the experimentally obtained psychometric curves that are based on subjects' responses. We consider three types of psychometric curves: (I) hit and false-alarm rates, (II) proportion of "target present" responses as a function

of minimum difference between the target, and any distractor, and (III) propor-tion of "target present" responses as a function of sample standard deviation be-tween stimuli orientations. We separately analyze type (II), and (III) curves in target-present, and target-absent trials. A detailed description of these curves is presented in Section 3.5.1. Figure 4.1 shows the mean responses of subjects on different types of psychometric curves.

4.1.1 Hit and false-alarm rates

The left panel in Figure 4.1(A) displays the average behavior of subjects in terms of hit (black), and false-alarm rates (red curve). The mean proportion of subjects responding "target present" on target-present, and target-absent trials are plotted in each experimental condition ($\rho_s \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$). The hit, and false-alarm rates show a similar trend in the first three experimental conditions, while they have an expected increase (hit rate), and decrease (false-alarm rate) in the case of homoge-neous distractors. Increased pairwise correlations between distractor orientations result in more structure among the stimuli, which should make it easier to single out the target, if present. Therefore, the hit rate increase (or false rate decrease) in case of homogeneous distractors is due to the effect of increased correlations that facilitate the target detection.

A similar trend is also visible in the right panel of Figure 4.1(A). The average performance of subjects is plotted in each experimental condition. The mean per-formance in first three conditions is very close to 60%, while in the case of $\rho_s = 1$,
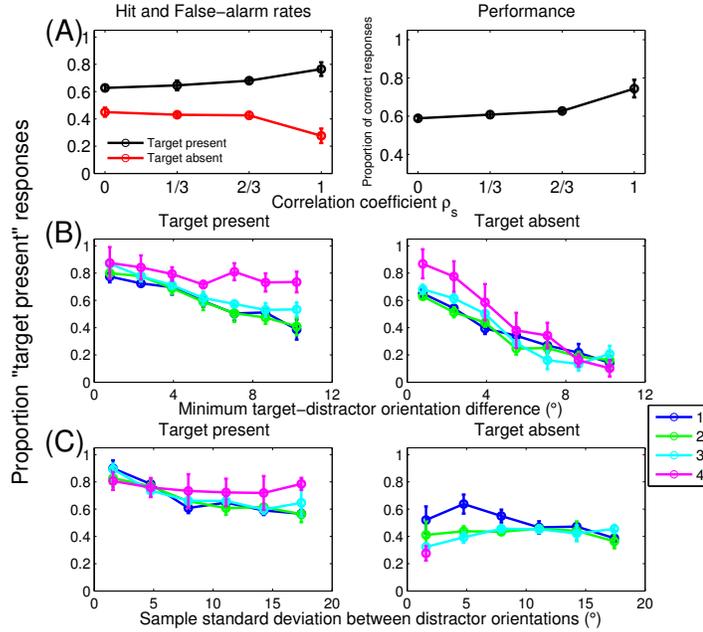
Figure 4.1: **Psychometric curves based on the experimental data.** Throughout the chapter, the error bars indicate unit standard error mean (s.e.m). **(A) Hit and false-alarm rates (left), and performance (right).** (Left) Hit (black), and false-alarm (red) rates as a function of correlation strength ρ_s used in the experimental conditions. Hit rate has similar behavior in the first three experimental conditions, while it shows a large increase in the case of homogeneous distractors. Similar trend is seen for false-alarm rate with a significant decrease only when $\rho_s = 1$. (Right) Mean subject performance in the four experimental conditions. **(B) Minimum target-distractor orientation difference.** Proportion of "target present" responses is plotted as a function of minimum difference between the target, and any distractor, separately for target-present (left), and target-absent (right) trials in the four experimental conditions. The difference is measured in degrees. Each curve corresponds to an experimental condition with $\rho_s \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. The data in first three conditions ($\rho_s = 0, \frac{1}{3},$ or $\frac{2}{3}$) show a very similar trend, while the curve for $\rho_s = 1$ is distinguishable from other three conditions in both target-present, and target-absent cases. **(C) Sample standard deviation between distractor orientations.** Proportion "target present" responses on different experimental conditions as a function of sample standard deviation between distractor orientations, in target-present (left), and target-absent (right) trials. The behavior is similar to the curves in (B) except for $\rho_s = 1$ in target-absent trials (right). Since all distractors are identical, there is a single value for the sample standard deviation, and hence a single data point.

it shows a large increase to 72.7%. This suggests that subjects might be using the enhanced structure present in visual displays to improve their decisions. However, we will investigate further to test this possibility.

4.1.2 Minimum target-distractor orientation difference

A more detailed view of the data can be seen in Figure 4.1(B). The mean proportion of "target present" responses in all four experimental conditions are plotted as a function of minimum difference between the target, and any distractor in both target-present (left), and target-absent (right) trials. These plots show that the proportion of subjects responding "target present" decreases as the minimum target-distractor orientation difference increases, both for target-present, and target-absent trials. Such a behavior is expected, since a large difference between the target, and any distractor reflects more dissimilarity of the distractors from the target, and hence it would be easier to perform the task. The curves corresponding to $\rho_s = 0, \frac{1}{3}$, and $\frac{2}{3}$ overlap, while the responses of subjects have a different behavior in the case of $\rho_s = 1$. This behavior is consistent with the trend seen in hit, and false-alarm rates in (A). Also, the decrease in the proportion of "target present" responses at larger minimum target-distractor differences is higher in target-absent trials for $\rho_s = 1$ in comparison to other experimental conditions. This is because the distractors are identical when $\rho_s = 1$, and when the difference between the target, and distractors increases, it would become easier to determine whether all stimuli are same, or there is an odd-ball (target) stimulus.

91

4.1.3 Sample standard deviation of distractor orientations

Another view of the data on different experimental conditions is presented in terms of proportion "target present" responses as a function of sample standard deviation of distractor orientations in Figure 4.1(C). Again, these are plotted separately for target-present (left), and target-absent (right) trials. These curves contain similar information as minimum target-distractor orientation difference plots in (B). Thus, the behavior on target-present trials is very similar to the left panel figure in (B). The proportion of "target present" responses decreases as the sample standard deviation of distractors increases, and hence the distractors get more dissimilar. On target-absent trials, the sample standard deviation reduces to a single value in the case of homogeneous distractors, and thus the curve has only one data point.

Thus, Figure 4.1 presents different psychometric curves to visualize the experimental data, and each curve provides a different insight about the behavior of subjects. The different plots suggest that the subjects' behavior is similar in experimental conditions with $\rho_s < 1$, while they do behave differently in the case of homogeneous distractors. We next describe different plausible models that could explain these observations, and provide a good fit for the experimentally obtained psychometric curves.

4.2 Models

In the optimal-observer model derived in Section 2.2, we assume that the observer
is aware of the correct generative model (Figure 2.1), and the associated parame-
ters such as N, σ_s^2, ρ_s, and prior over T. However, we need to test whether human
observers learn, and use the correct generative model in making decision. If they
do not, their inferences are *suboptimal*. For instance, an observer may not use
equal odds prior, or the correct value of ρ_s. Therefore, we want to analyze how
subjects made their decisions, and what parameter values they used to make their
responses. Specifically, we are interested in determining what values of ρ_s subjects
used to make their responses, and how certain they are in making their measure-
ments. We note that we can only determine this by fitting models to the data,
and making conclusions based on the best fitting model. It is always possible that
there are better models, and better explanations, however.

The subjects were pre-informed about the number of stimuli being $N = 4$ in
the experiment. Further, we assume that subjects were able to infer the correct
value of σ_s as $15°$ in the experiment, and used it to make their decisions. Though
this may not necessarily be true. It is possible that subjects did not infer the value
of σ_s correctly in the experiment, and might have used an incorrect assumption
about it. In that case, we would need models that incorporate the plausible as-
sumptions about σ_s that observers could have used. Such models will have σ_s
as a free parameter, and we would need to test possible assumptions on it. For
instance, subjects could use different values of σ_s across experimental sessions, or

possibly they could use a distribution over the values of σ_s. We can easily see that this would add another dimension of complexity to our existing complex models, and would further make our computations intractable. Thus, we save some order of complexity in our models by assuming that subjects were able to correctly infer the true value of σ_s in the experiment.

We are mainly interested in determining whether subjects use the correct or incorrect assumption about ρ_s in making decisions on different experimental conditions. We denote the true experimental value of ρ_s by $\rho_{s_{\text{true}}}$, and a subject's assumed value of ρ_s by $\rho_{s_{\text{assumed}}}$. Therefore, we want to determine whether $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}}$, or $\rho_{s_{\text{assumed}}} \neq \rho_{s_{\text{true}}}$ for a subject in each experimental condition. If the subject uses $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}}$ in all experimental conditions, the subject is *optimal* on the task. Otherwise, we refer to the condition $\rho_{s_{\text{assumed}}} \neq \rho_{s_{\text{true}}}$ as *suboptimal condition*.

To answer this question, we consider several suboptimal models. We categorize these models based on the assumptions about ρ_s, and the encoding precision:

1. **Assumption about ρ_s**: we explored whether subjects use the correct ($\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}}$) or incorrect ($\rho_{s_{\text{assumed}}} \neq \rho_{s_{\text{true}}}$) assumption about ρ_s in the generative model.

2. **Encoding precision**: we do not know how subjects made their measurements. In particular, what precision values they used for encoding stimuli. We need to determine whether subjects encode all stimuli with equal precision, or they use varying precision for stimuli across trials. Therefore, we test models that have the assumptions about subjects encoding stimuli with

equal or variable precision.

In the following sections, we discuss the above two categories of models in details, and also their further division into sub-categories.

4.2.1 Assumption about ρ_s

An ideal observer uses the correct value of ρ_s when inferring target presence. However, it is possible that the observer may not be aware of the correct correlation strength. To test which case is more likely, we consider models with ρ_s as a free parameter. Specifically, we tested three main assumptions about ρ_s:

(a) $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}}$: we assume that the observer uses the correct value of correlation strength in all experimental conditions to infer target presence, that is, $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}} \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$.

(b) $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{constant}}}$: in this condition, we consider that the observer assumes that the level of correlation among distractor orientations is constant across all conditions. In particular, we check whether

  (i) the observer completely ignores the information about correlation in all experimental conditions, and uses $\rho_{s_{\text{constant}}} = 0$. This would mean that observer uses $(0, 0, 0, 0)$ as correlation values in making decision on all experimental conditions.

  (ii) the observer may also use any other value between 0, and 1 as $\rho_{s_{\text{constant}}}$ in making decision. We thus let $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{constant}}}$ to be a free parameter in

the model.

(c) $\rho_{s_{\text{assumed}}} \neq \rho_{s_{\text{true}}}$: we further allow the possibility that the observer may use different, possibly incorrect correlation strengths across experimental conditions. We let $\rho_{s_{\text{assumed}}}$ to be a free parameter in the model, and check for the following possibilities:

(i) the observer considers the first three conditions identically, and thus uses an equal correlation strength in making decision on these conditions, while assume a different level of correlation in the fourth experimental condition. In such a case, we fit $\rho_{s_{\text{assumed}}}$ as a constant free parameter in the first three conditions, and as another free parameter in the fourth experimental condition *i.e.,* $(\alpha, \alpha, \alpha, \beta)$.

(ii) the observer uses an incorrect assumption about ρ_s in all experimental conditions. We thus fit $\rho_{s_{\text{assumed}}}$ per condition in this model, *i.e.,* $\rho_{s_{\text{assumed}}}$ has following form $(\alpha, \beta, \gamma, \delta)$. We note that this is the most general assumption about ρ_s, and all above models are special cases of this model.

Here α, β, γ, and δ represent free parameters of the model, and they are fitted per model for each subject. We also note that $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{constant}}}$, and $\rho_{s_{\text{assumed}}} \neq \rho_{s_{\text{true}}}$ are suboptimal conditions on the task.

4.2.2 Encoding precision

Along with the assumption about ρ_s, we also consider different assumptions about encoding precision in our models. Since the precision of measurements greatly impacts the accuracy of decisions, we want to determine which assumption is most consistent with the responses of the subjects. Signal detection models have typically assumed that encoding precision is constant across stimuli, and trials at a given set size [106, 127, 153, 149]. However, recent experimental studies show that the observers' measurements are of variable precision. That is, the encoding precision varies across stimuli, and trials [141, 43, 94, 95, 71, 72, 129, 131]. This variability could be attributed to the attentional fluctuations, or other factors.

We thus consider both possible models of the precision of measurements: in the first type of model, we assume that the measurements of all stimuli have equal precision. We denote the precision of the ith measurement by $J_i = \frac{1}{\sigma_i^2}$. Under the equal precision assumption, we assume that J_i is constant across stimuli, that is, $J_i = J$ for all $i = 1, 2, \cdots, N$. The constant J is a free parameter in the model. These models are known as **equal precision (EP) models**. We assume that the precision is constant across trials in the same experimental condition, however, it may vary across trials in different experimental conditions. We thus test both possibilities in our models by assuming J is constant, and variable across experimental conditions. We, therefore, consider two types of EP models with J as a single constant parameter in all conditions, and as a varying parameter across the four experimental conditions.

In the second model type of precision, we assume encoding precision varies randomly across stimuli, and trials. These models are known as **variable precision (VP) models**. We denote the precision variable by a vector $\mathbf{J} = (J_1, J_2, \cdots, J_N)$, where J_i corresponds to the precision of the ith measurement. In the VP models, we assume that the precision variable $\mathbf{J}$ with which stimuli are encoded is a random variable. To model such variability, we assume $\mathbf{J}$ follows a gamma distribution with mean $\bar{J}$, and scale parameter, τ. We sample the precision randomly for each stimulus on each trial. Therefore, the measurement is described by a doubly stochastic process, $(\bar{J}, \tau) \rightarrow \mathbf{J} \rightarrow \mathbf{x}$ [141]. Thus, the precision determines the distribution of stimulus estimate, but is itself also a random variable.

Why do we choose the gamma distribution to model the variability in precision? The proper choice of a distribution for modeling variable precision would require the marginalization over all possible ways to implement this variability. Though a full marginalization seems impossible, but the success of the VP concept can be assessed how well it performs under various specific alternative distributions. Van den Berg et al. [140] have implemented, and tested VP models with many other alternatives such as log-normal, Weibull, and log-uniform distributions. They have found that the results are consistent under changes in the assumed distribution over precision. Therefore, we consider that our results would be robust under the choice of a distribution. The gamma distribution is a two-parameter family of continuous, unimodal distributions on the positive real line, and has been successfully used for modeling variable precision. We thus consider the same choice for our models.

In our VP models, both $\bar{J}$, and τ are free parameters. We assumed the scale τ to be constant across experimental conditions. But similar to EP models, we consider models with $\bar{J}$ constant, or varying across experimental conditions.

4.2.3 Summary of models

We test the following two assumptions about encoding precision in EP (with precision parameter J), and VP (mean precision parameter $\bar{J}$) models:

- precision J (or $\bar{J}$ in VP) is independent of experimental conditions (ρ_s-independent),

- precision J (or $\bar{J}$ in VP) vary across experimental conditions, and is thus ρ_s-dependent.

In addition, we also have the following model variants based on the assumption about ρ_s for each category of EP, and VP models:

(i) $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}} = (0, \frac{1}{3}, \frac{2}{3}, 1)$, the optimal model

(ii) $\rho_{s_{\text{assumed}}} = (0, 0, 0, 0)$, *i.e.*, no correlations model

(iii) $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$, constant correlations model

(iv) $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$, constant in first three conditions, and different in $\rho_s = 1$ condition

(v) $\rho_{s_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$, the most flexible model in terms of ρ_s.

We test each of the above assumption about ρ_s for each possible combination of EP, and VP models. Therefore, in total, we consider $2 \times 2 \times 5 = 20$ models. In addition to the assumptions about ρ_s, and encoding precision, we also consider prior for T as a free parameter in our models. We assume this parameter to be constant across experimental conditions.

Table 4.1 gives a detailed summary of the category of models we consider, and their number of parameters. For reference convenience, we number models in each precision category, from EP1 to EP10, and VP1 to VP10. We note that most of the models are high-dimensional (with parameters > 3), and these parameters are associated with different correlation conditions.

Precision	Assumption on J or $\bar{J}$	Model No.	Assumption on ρ_s	No. of free parameters
EP	J condition-independent	EP1	$\rho_{s_{\text{assumed}}} = (0, \frac{1}{3}, \frac{2}{3}, 1)$	2
		EP2	$\rho_{s_{\text{assumed}}} = (0, 0, 0, 0)$	2
		EP3	$\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$	3
		EP4	$\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$	4
		EP5	$\rho_{s_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$	6
EP	J condition-dependent	EP6	$\rho_{s_{\text{assumed}}} = (0, \frac{1}{3}, \frac{2}{3}, 1)$	5
		EP7	$\rho_{s_{\text{assumed}}} = (0, 0, 0, 0)$	5
		EP8	$\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$	6
		EP9	$\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$	7
		EP10	$\rho_{s_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$	9
VP	$\bar{J}$ condition-independent	VP1	$\rho_{s_{\text{assumed}}} = (0, \frac{1}{3}, \frac{2}{3}, 1)$	3
		VP2	$\rho_{s_{\text{assumed}}} = (0, 0, 0, 0)$	3
		VP3	$\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$	4
		VP4	$\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$	5
		VP5	$\rho_{s_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$	7
		VP6	$\rho_{s_{\text{assumed}}} = (0, \frac{1}{3}, \frac{2}{3}, 1)$	6

Continued on next page

Table 4.1 – *Continued from previous page*

Precision	Assumption on J or $\bar{J}$	Model No.	Assumption on ρ_s	No. of free parameters
		VP7	$\rho_{s_{\text{assumed}}} = (0,0,0,0)$	6
VP	$\bar{J}$ condition-dependent	VP8	$\rho_{s_{\text{assumed}}} = (\alpha,\alpha,\alpha,\alpha)$	7
		VP9	$\rho_{s_{\text{assumed}}} = (\alpha,\alpha,\alpha,\beta)$	8
		VP10	$\rho_{s_{\text{assumed}}} = (\alpha,\beta,\gamma,\delta)$	10

Table 4.1: **List of models considered to explain the data obtained on the target detection experiment.** Description of models fitted to the experimental data with different assumptions about encoding precision parameter, and correlation strength. The number of free parameters per model is also listed.

Apart from these models, there could be many other possible assumptions about ρ_s such as $\rho_{s_{\text{assumed}}} = (1,1,1,1)$, $\rho_{s_{\text{assumed}}} = (0,0,0,1)$, $\rho_{s_{\text{assumed}}} = (0,0,0,\alpha)$, and $\rho_{s_{\text{assumed}}} = (\alpha,\alpha,\alpha,1)$. We also test these possibilities in our models, however, the model fits are either worse, or comparable to other general models we have considered above. Thus, we do not include these models in our analysis below.

We now present the model fits of the data for each of the model listed in Table 4.1. Each model fit is obtained by generating synthetic data based on the maximum-likelihood estimates of the model parameters for each subject, using

the same number of trials as in the subject data (averaged over 100 runs). The predictions for different psychometric curves are obtained based on these synthetic data.

4.3   Equal precision models

We first examine the fits of the EP models to the subjects' data. We present the predictions of all EP models for the psychometric curves, and discuss the consequences.

4.3.1   Condition-independent precision J

We consider the precision to be constant across experimental conditions, and test different assumptions about ρ_s. We check the model fits of EP1 to EP5 here. These models have a common assumption about precision, but they differ in the assumption about $\rho_{s_{\text{assumed}}}$, and hence have varying number of free parameters. In addition, they have a common assumption about the prior of T. In the figures below, the shaded areas represent the fits of the model to the data. The model predictions for each subject are individually obtained, and averaged over subjects.

Figure 4.2 shows the fit of EP1 model for the data. This model assumes that subjects use correct assumption about the correlation strength ρ_s in making their decisions. The fits for the hit rate, false-alarm rate, and performance illustrate a good match (RMS error 0.036, and 0.011, and R^2 of 0.94, and 0.96, respectively)
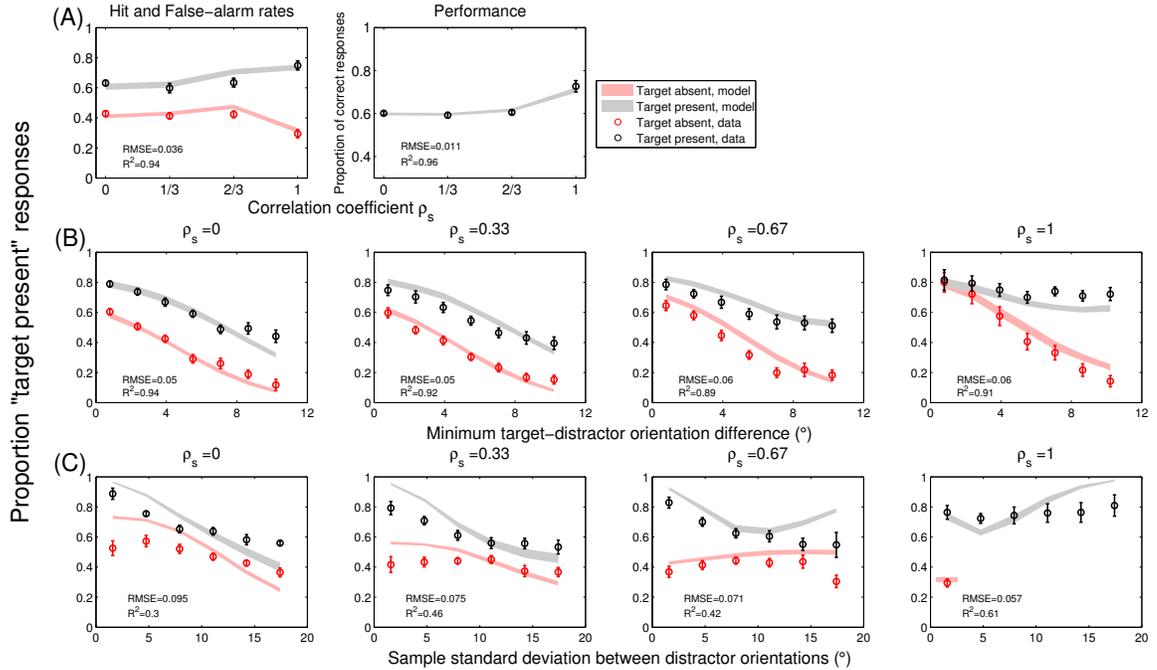
Figure 4.2: **EP model 1 (*J* condition-independent, and $\rho_{s_{\mathbf{assumed}}} = \rho_{s_{\mathbf{true}}}$) fits for the data.** Throughout the chapter, the shaded areas show the fit of the model, and the circles are averaged subject responses. The error bars, and shaded areas represent unit standard error of the mean for subjects data, and model fits, respectively. The values in the plots indicate the RMSE, and R^2 errors between the data, and model fit. **(A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

between the model predictions, and the data. Also, the model predictions have a close agreement to the data for the psychometric curves in Figure 4.2(B) with small errors (RMSE equal to 0.05, 0.05, 0.06, and 0.06 in the four experimental conditions, while R^2 equal to 0.94, 0.92, 0.89, and 0.91, respectively). However, when compared based on the measure of sample standard deviation between distractors, the model predictions fail badly, and show large deviations (Figure 4.2(C)). Specifically, the predictions are worse in the cases of $\rho_s = 0$, and $\rho_s = \frac{1}{3}$ (RMSE equal to 0.095, and 0.075; R^2 equal to 0.3, and 0.46, respectively). These curves clearly suggest that subjects do not use the assumptions of this model in their inference process.

We see a far more worse trend (RMS errors of more than 0.6, and R^2 values are even negative for some curves) in the fits of EP2 model in Figure 4.3. None of the curves are predicted by the assumption of zero correlations in this model. The model predictions are worse in the case of homogeneous distractors, $\rho_s = 1$ (Figure 4.3(C)) with the RMS errors of 0.2, and R^2 a large negative value. The worse fit of this model in all conditions clearly indicates that subjects might be using correlations in making their decisions. However, we still need to investigate what values they use in such a case.

Next, we compare the fits of EP3 model in Figure 4.4 that has the assumption of constant $\rho_{s_{\text{assumed}}}$ across experimental conditions. The model has better predictions for the data in the first three experimental conditions (RMSE < 0.07, and $R^2 > 0.64$, respectively) as compared to the condition of $\rho_s = 1$. Though the fits
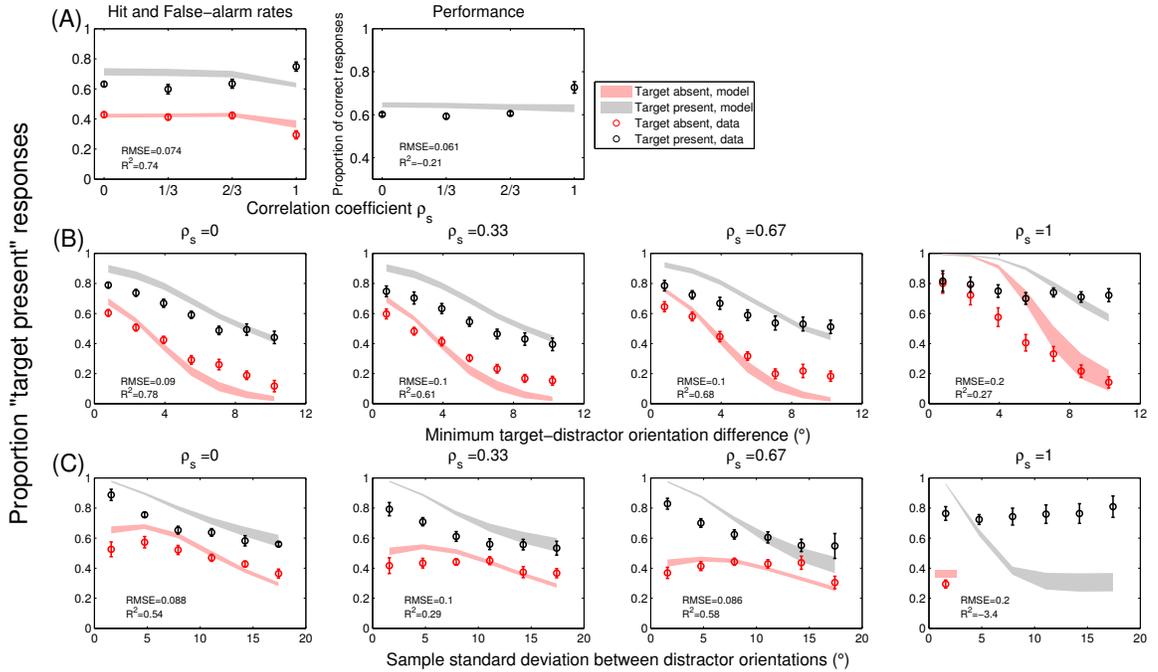
Figure 4.3: **EP model 2 (*J* condition-independent, and $\rho_{s_{\textbf{assumed}}} = (0, 0, 0, 0)$)**
fits for the data. (A) Hit and false-alarm rates. (Left) Hit, and false-alarm
rates as a function of correlation strength ρ_s used in the experimental condi-
tions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum**
target-distractor orientation difference. Proportion "target present" responses
as a function of minimum target-distractor orientation difference, separately for
target-present (gray), and target-absent (red) trials in each experimental condition
(columns). **(C) Sample standard deviation of distractor orientations.** Proportion
"target present" responses as a function of sample standard deviation of distrac-
tor orientations, separately for target-present (gray), and target-absent (red) trials
in each experimental condition (columns).

Figure 4.4: **EP model 3 (J condition-independent, and $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

are not perfect even when $\rho_s = \frac{1}{3}$, and $\rho_s = \frac{2}{3}$, particularly, the predictions in Figure 4.4(C) show disagreement between model fits, and the data at large values of sample standard deviation between distractor orientations. However, the model completely fails to explain the behavior of subjects on the homogeneous distractors condition, and has RMS errors of 0.1, and R^2 of 0.64 in the minimum target-distractor orientation plots (Figure 4.4(B)), and RMS errors of 0.15, and R^2 of -0.21 in the case of sample standard deviation plots (Figure 4.4(C)). These fittings of the model with the constant assumption of ρ_s suggest that perhaps subjects treat the case of $\rho_s = 1$ differently from the other experimental conditions, and they might be using different inference processes when correlations are not perfect, and when they are. This seems to be consistent with the responses of subjects observed in Figure 4.1. However, since this model could not completely explain the behavior even in the cases of $\rho_s < 1$, it is difficult to support this hypothesis using the predictions of the EP3 model.

To test our hypothesis of whether subjects treat the homogeneous condition differently from other experimental conditions, we check the predictions of the EP4 model in Figure 4.5. EP4 model allows the possibility of having $\rho_{s_{\mathrm{assumed}}}$ as a free parameter in the first three conditions, and separately in the fourth condition along with the assumption of constant precision parameter J. That is, $\rho_{s_{\mathrm{assumed}}}$ has the form of $(\alpha, \alpha, \alpha, \beta)$, and assumes that the observer uses a constant value of $\rho_{s_{\mathrm{assumed}}}$ in the first three experimental conditions, but treat the fourth condition differently (when $\beta \neq \alpha$ for the observer). If for an observer, β is equal to α, the fits of the EP4 model reduce to those of EP3 (with $\rho_{s_{\mathrm{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$) for that
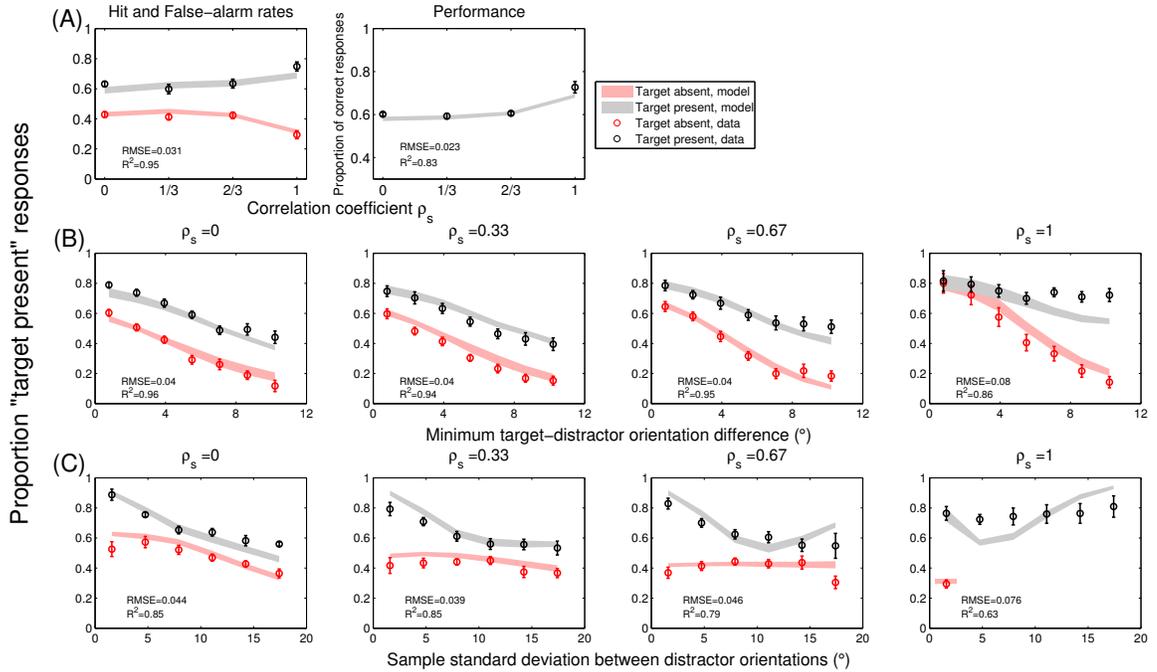
Figure 4.5: **EP model 4 (*J* condition-independent, and $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$)** **fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

observer's data.

The EP4 model does extremely well in predicting the subjects' behavior on conditions with $\rho_s < 1$. Specifically, the hit, and false-alarm rates (left panel), and performance curve (right panel) are closely fitted in Figure 4.5(A) with RMS errors of 0.031, and 0.023 (R^2 of 0.95, and 0.83), respectively. Further, the shaded curves of model are well aligned (RMS errors of 0.04, and $R^2 \geq 0.94$) with the data points in the minimum target-distractor orientation plots in Figure 4.5(B) except in the case of $\rho_s = 1$ which shows some deviations in the target-present trials (fourth panel). But, the comparison based on the sample standard deviation figures (Figure 4.5(C)) indicate that this model also fails to capture the responses of subjects on the perfect correlation condition. The model does have better fit as compared to EP3 model in Figure 4.4(C, fourth panel) however, they still have large RMS errors of 0.076, and low R^2 value of 0.63. In addition, the model predictions fail at couple of data points in the first three experimental conditions in Figure 4.5(C).

Therefore, the models EP1 to EP4 still lack the assumptions to explain the behavior of subjects on the target detection task. In order to further explore how subjects made their decisions, we test the most flexible assumption about $\rho_{s_{\text{assumed}}}$ with constant precision J across experimental conditions in EP models. We let $\rho_{s_{\text{assumed}}}$ to be a free parameter, and fit it per condition, thus it is of the form of $(\alpha, \beta, \gamma, \delta)$ form. This is our model EP5, and it represents a general model for the assumption on $\rho_{s_{\text{assumed}}}$. The model fitting curves are shown in Figure 4.6. Analyzing the model fits for each psychometric curve, we find that the predictions of this model are very close to those of EP4 model in Figure 4.5. Further, the magnitude
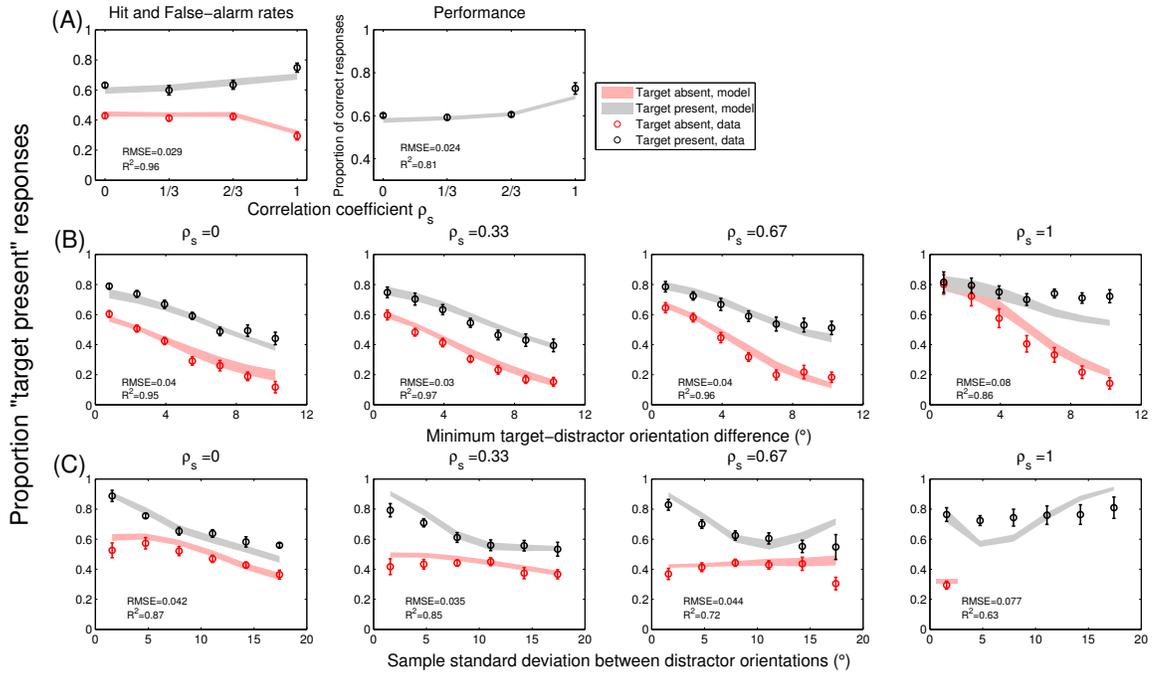
Figure 4.6: **EP model 5 (J condition-independent, and $\rho_{s_{\mathbf{assumed}}} = (\alpha, \beta, \gamma, \delta)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

of RMS errors, and the goodness of fit R^2 is similar in both cases. This suggest that the two models are close in their predictions for the data, however, none of them provide a better explanation in the perfect correlation experimental condition.

Since none of the models EP1 to EP5 could provide a good fit for the experimental data, it is difficult to conclude anything about the behavior of subjects based on these models. However, these models were based on different assumptions about $\rho_{S_{\text{assumed}}}$, including the most general one but with the common hypothesis of constant J across experimental conditions. The mis-fit of all these models to the data suggests that in addition to $\rho_{S_{\text{assumed}}}$, the assumption of precision J is also important, and that plays a crucial role in determining the predictions of the model for the data. Thus, we consider models to test the assumption whether subjects have varying precision J across experimental conditions. To be consistent, we test the similar assumptions about $\rho_{S_{\text{assumed}}}$ as in models EP1 to EP5 in the following section.

4.3.2   Condition-dependent precision J

We now consider models EP6 to EP10 which have a common assumption about the precision J constant in a particular experimental condition, but varying across conditions. We begin with testing the hypothesis of $\rho_{S_{\text{assumed}}} = \rho_{S_{\text{true}}} = (0, \frac{1}{3}, \frac{2}{3}, 1)$ in EP model 6.

We find in Figure 4.7(A), and (B) that the model has relatively good fit ($R^2 \geq$
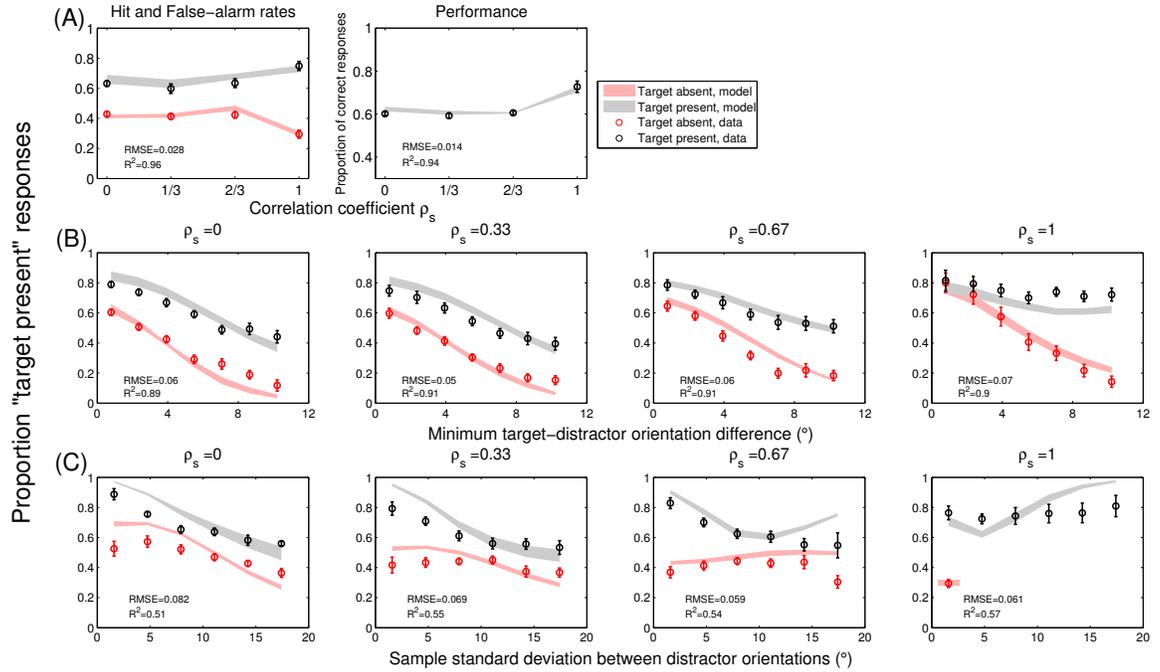
Figure 4.7: **EP model 6 (*J* condition-dependent, and $\rho_{s_\text{assumed}} = \rho_{s_\text{true}}$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

0.89) for the data as compared to models EP1 to EP5 that assume J to be constant across experimental conditions. But, the model predictions are worse for all conditions in the sample standard deviation plots (Figure 4.7(C)). The RMS errors range from a minimum of 0.059 ($\rho_s = \frac{1}{3}$) to 0.082 ($\rho_s = 0$), while R^2 values are of the orders of 0.51(when $\rho_s = 0$) to 0.57 ($\rho_s = 1$). Therefore, sample standard deviation plots provide us additional information about the model fits in addition to the minimum target-distractor orientation plots which were used for analysis in studies done by Mazyar et al.[94, 95]. Based on the poor performance of this model in Figure 4.7(C), we reject this model, and test other assumptions about the $\rho_{s_{\text{assumed}}}$ to predict subjects' responses.

We next consider the fits of EP7 model in Figure 4.8. This model assumes that observers do not learn any structural correlations present in visual scenes, and consider all conditions as heterogeneous displays ($\rho_{s_{\text{assumed}}} = (0, 0, 0, 0)$). Considering only the curves in Figure 4.8(A), the model seems to be predicting well for the data with small RMS errors of 0.035 (left panel), and 0.025 (right panel). However, when we consider the fits for the minimum target-distractor orientations in Figure 4.8(B), and the sample standard deviation plots in Figure 4.8(C), we see huge errors (RMSE of the order of 0.06 or more) between the model predictions, and the data. This clearly indicates that this model is a bad choice to predict the responses of the subjects in this experiment. Perhaps subjects do take correlations into account while making their decisions. But, we are unaware of the process they follow.

Thus, we test the assumption of constant correlations in the experiment in EP
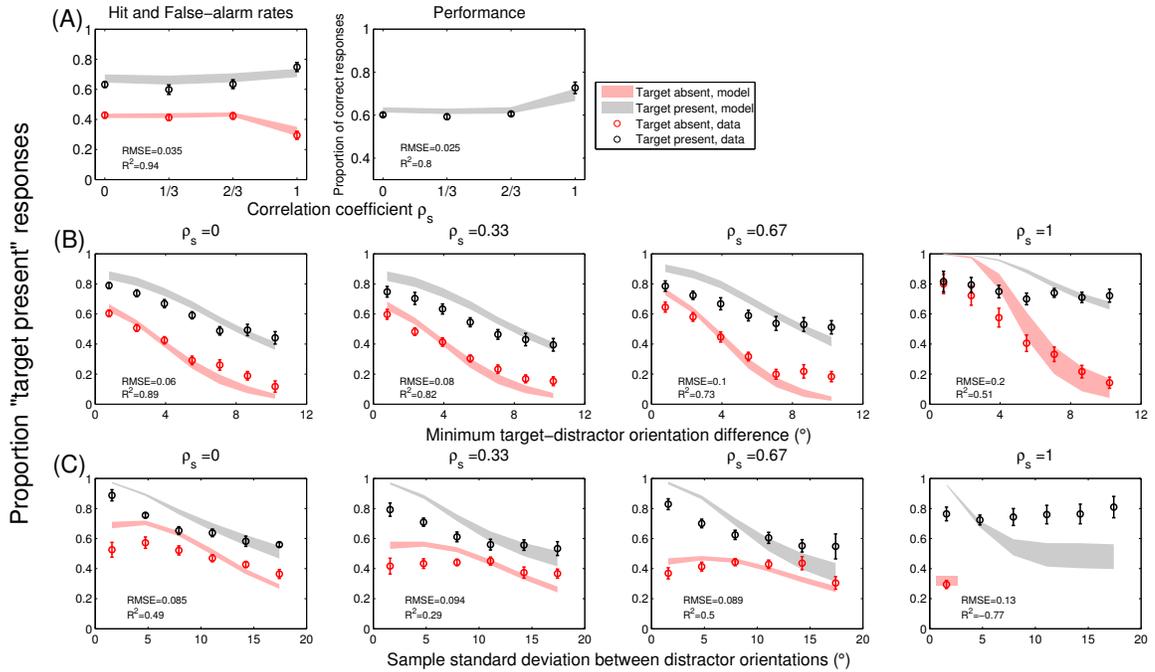
Figure 4.8: **EP model 7 (*J* condition-dependent, and $\rho_{s_{\textbf{assumed}}} = (0, 0, 0, 0)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

Figure 4.9: **EP model 8 (*J* condition-dependent, and $\rho_{S_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

model 8. The model predictions, and fits are shown in Figure 4.9. Analyzing the curves in (B), and (C), we find that the model does not reproduce the subjects' behavior on the experiment. It particularly fails in the condition of $\rho_s = 1$ with RMS error of more than 0.1, and relatively poor R^2 value. We also find that the fitting of psychometric curve in the experimental condition of $\rho_s = \frac{1}{3}$ in Figure 4.9(B, second panel) is worse as compared to EP7 model in Figure 4.8(B, second panel). This seems odd since EP7 model is a special case of EP8 model with $\alpha = 0$ in all conditions. This might be explained using the computational inefficiency we faced in obtaining the fits of our models. We obtained the maximum-likelihood model parameters using the exhaustive grid search method (Section 3.4.1), and such an estimation depends on the convergence of the likelihood function. If the likelihood function does not converge, the estimates may not represent the global maximum-likelihood parameter estimates. But, overall the model has good predictions as compared to EP7 model, particularly comparing the fits in Figure 4.9(C) to the ones in Figure 4.8(C).

We continue our analysis with EP9 model with condition-dependent precision parameter J. The model provides a better picture of the data as seen in Figure 4.10. We find better fits for the psychometric curves in (A), and (B). However, we continue to find model deviations for the sample standard deviation plots in Figure 4.10(C). In particular, the model fails to account for the data in the homogeneous condition. Though compared to other EP models (EP1 to EP8) discussed so far, the RMS error is found to be small in this case at 0.058 (comparable in case of EP1), and hence the model does make close predictions for the data but it still
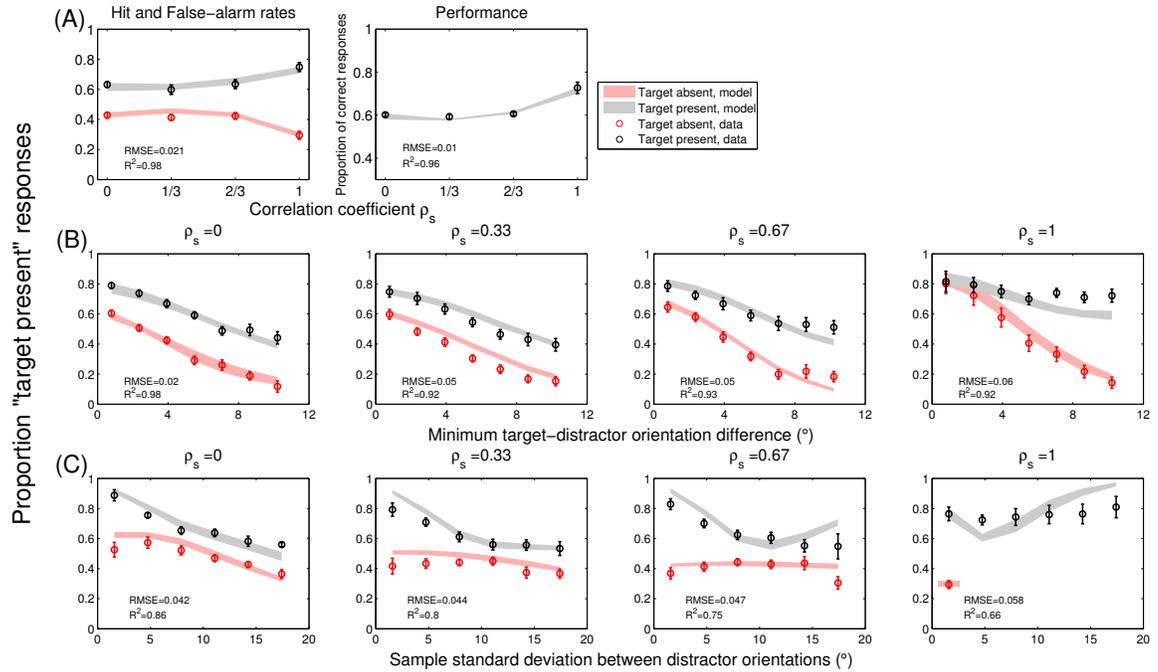
Figure 4.10: **EP model 9 (J condition-dependent, and $\rho_{s_{\mathbf{assumed}}} = (\alpha, \alpha, \alpha, \beta)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

does not explain the data based on the sample standard deviation plots.

Finally, we test our last EP model listed in Table 4.1, which is EP10 model. This is the EP model with the most flexible assumption about $\rho_{s_{\text{assumed}}}$, and precision J. We compare the fits for this model in Figure 4.11. Like many other EP models, the model does extremely well in the first three conditions, but fails to explain the behavior on the homogeneous condition. Specifically, the poor model fit is continuously seen in the sample standard deviation plot for $\rho_s = 1$ (Figure 4.11(C) here). The hit, and false-alarm rates (Figure 4.11(A)), along with the psychometric curves of minimum target-distractor orientations (Figure 4.11(B)) are well predicted by the model assumptions, and have R^2 values of more than 0.9. Further, the model fits to the data are acceptable in the first three experimental conditions in Figure 4.11(C) with small RMS errors of up to 0.044. But, a worse model fit (RMSE equal to 0.064) in the condition of $\rho_s = 1$ allows us to reject this model, and explore variable precision models.

The motivation to test VP models is to find a model that could possibly explain the subjects' behavior equally well across all experimental conditions, and in terms of all psychometric curves. We thus now analyze the fits of VP models for the experimental data.
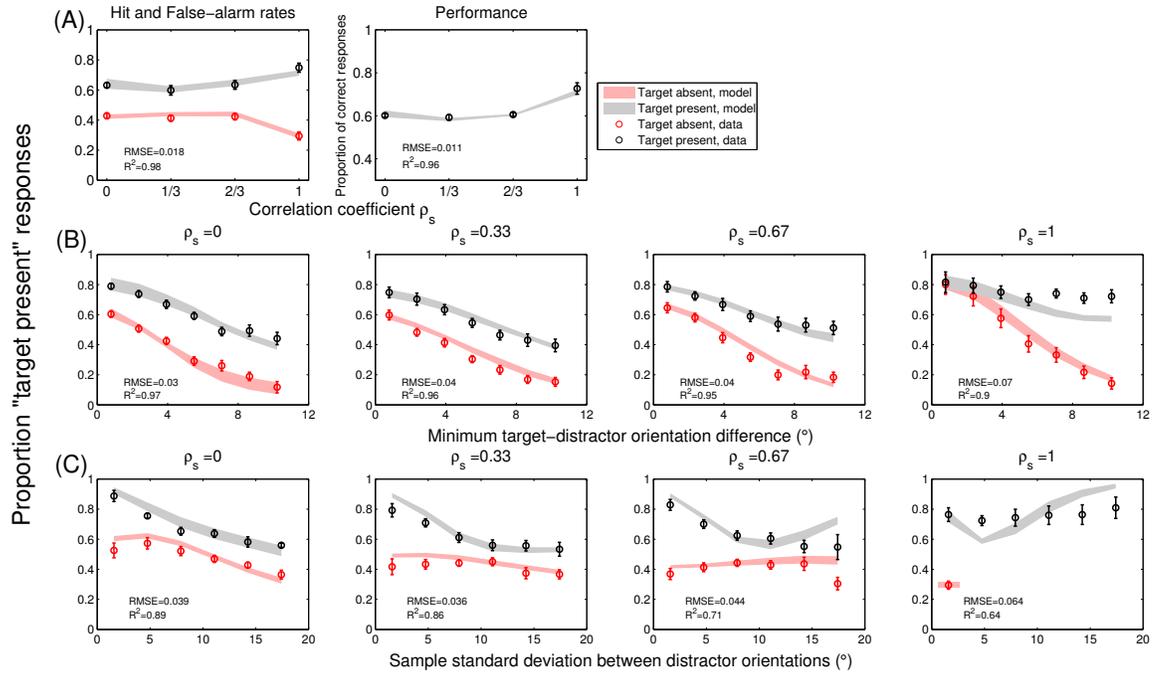
Figure 4.11: **EP model 10 (***J* **condition-dependent, and** $\rho_{s_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$**)** **fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

4.4 Variable precision models

Since equal precision models failed to explain the behavior of subjects' on the experiment, we further investigate if models with variable precision can provide an insight into how subjects inferred their responses. Variable precision models have been successfully used to explain human decisions on visual search task [94, 95] with homogeneous, and heterogeneous distractors. We thus test if varying precision can be a key factor in interpreting the decisions of subjects in the target detection experiment here. Similar to EP models, we consider both possibilities of the mean precision $\bar{J}$ of the gamma distribution to be constant, or varying across experimental conditions. We discuss the fits for both categories of models below.

4.4.1 Condition-independent mean precision $\bar{J}$

We first examine the variable precision models with mean $\bar{J}$ assumed to be constant across experimental conditions. We also assume the scale parameter τ of the gamma distribution to be constant across experimental conditions, and hence is another free parameter for the precision in the model. We check for the five possibilities for the $\rho_{s_{\text{assumed}}}$ in VP1 to VP5 models (described in Table 4.1), and analyze the resulting fits to find the best model if there is any.

Figure 4.12 shows the fitting of VP1 model to the data. We see the model explains the subjects' responses in the first three experimental conditions except for minor deviations in the target-present trials when $\rho_s = \frac{1}{3}$ in Figure 4.12(B).
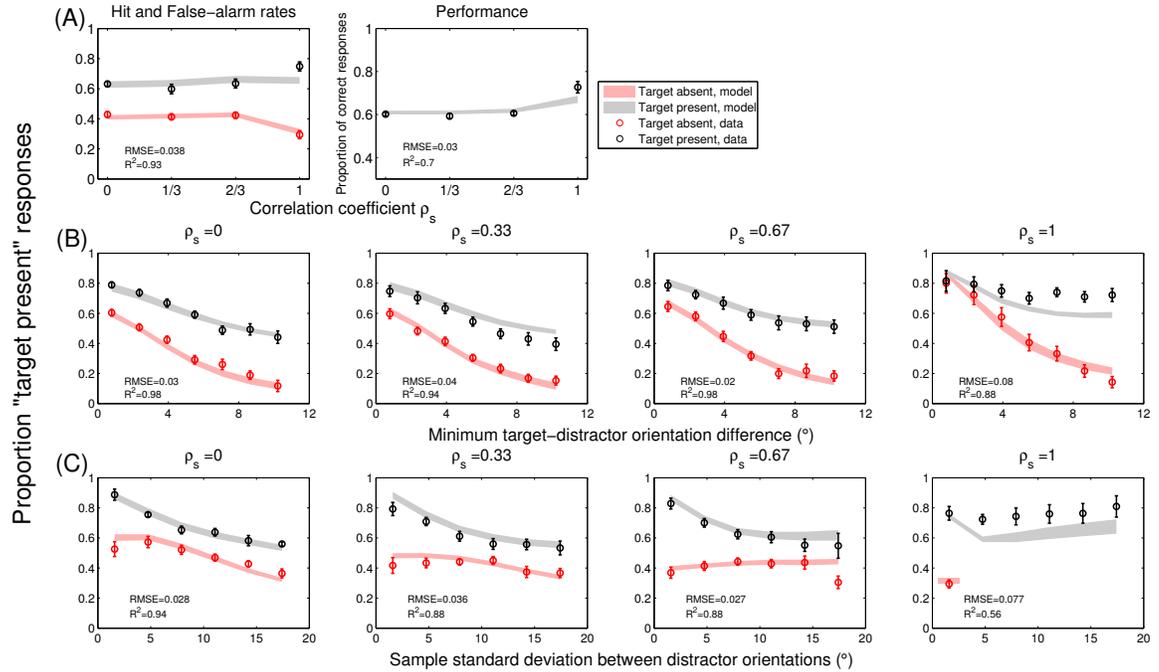
Figure 4.12: **VP model 1 ($\bar{J}$ condition-independent, and $\rho_{s_{\mathbf{assumed}}} = \rho_{s_{\mathbf{true}}}$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

However, like many EP models, the model fails to account for the behavior in the perfect correlation condition. Specifically, the model predicts a lower proportion of "target-present" responses as compared to the subjects' data in the sample standard deviation plot in Figure 4.12(C, fourth panel). The RMS error is equal to 0.077, while the goodness of fit, R^2 is measured at 0.56 in this case. Therefore, with this variable precision model having correct assumption about the generative model, the data could not be explained well in the homogeneous condition.

We further check the predictions of VP2 model for the data in Figure 4.13. Very similar to VP1 model in Figure 4.12, the predictions of VP2 model completely fail to interpret the behavior in the experimental condition with $\rho_s = 1$. Moreover, the predictions are worse in this case as compared to VP1 model, as the RMS error increases to 0.15 (in Figure 4.13(C, fourth panel)) as compared to 0.077 in VP1 model (Figure 4.12(C, fourth panel)), and 0.1 (Figure 4.13(B, fourth panel)) from 0.08 (Figure 4.12(B, fourth panel)) in the minimum target-distractor orientation plots. However, except for the condition of $\rho_s = 1$, the data points are well matched with the model predictions. For instance, the goodness of fit varies from 0.93 to 0.98 in the case of minimum target-distractor orientation plots in Figure 4.13(B, first, second, and third panels), and 0.9 to 0.97 in sample standard deviation plots in Figure 4.13(C, first, second, and third panels). This suggests that there is something really particular about the homogeneous condition that the above discussed models are missing, and subjects do treat the condition with higher statistical structure differently as compared to the partial correlation conditions.
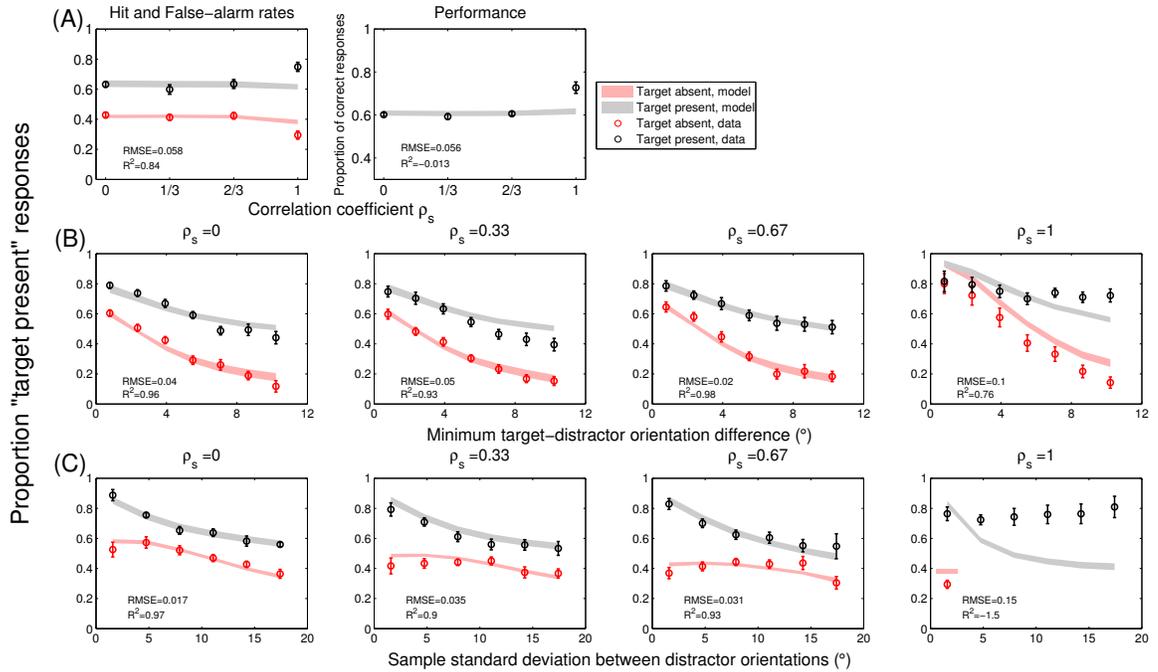
Figure 4.13: **VP model 2 ($\bar{J}$ condition-independent, and $\rho_{s_{\text{assumed}}} = (0, 0, 0, 0)$)** **fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

Figure 4.14: **VP model 3 ($\bar{J}$ condition-independent, and $\rho_{s_{\mathbf{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$ fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

Further, we analyze the fits of models VP3, VP4, and VP5 in Figures 4.14, 4.15, and 4.16, respectively. We find that all these models have a similar trend in the first three conditions, and completely fail to predict the behavior on the homogeneous condition ($\rho_s = 1$). Though these models differ in their assumptions about $\rho_{s_{\text{assumed}}}$, the predictions are quite similar in terms of model generated psychometric curves. Model VP5 has the most flexible assumption about $\rho_{s_{\text{assumed}}}$ in this category of models, and we notice that this model also shows large RMS errors of 0.091 in Figure 4.16(C, fourth panel), and a poor fit of R^2 equal to 0.43.

Thus, none of the VP models in this category could provide the explanation for subjects' responses. All models VP1 to VP5 show good agreement with the data in the first three experimental conditions (with small RMS errors), while fail to incorporate the decision strategies of subjects when the distractors are perfectly correlated. The assumption of variable precision with constant $\bar{J}$ across conditions did not improve the predictions of the models. This suggests that perhaps subjects' precision not only vary across stimuli, and trials, but may also be varying across experimental conditions. Or, in the worst case, subjects may be using some other assumptions about the generative model that our discussed models (EP1 to EP10, and VP1 to VP5) fail to incorporate.

Therefore, we finally test whether adding the assumption about varying $\bar{J}$ across conditions in variable precision models could provide a better explanation for the subjects' behavior on the experiment, specifically in the case of homogeneous condition.
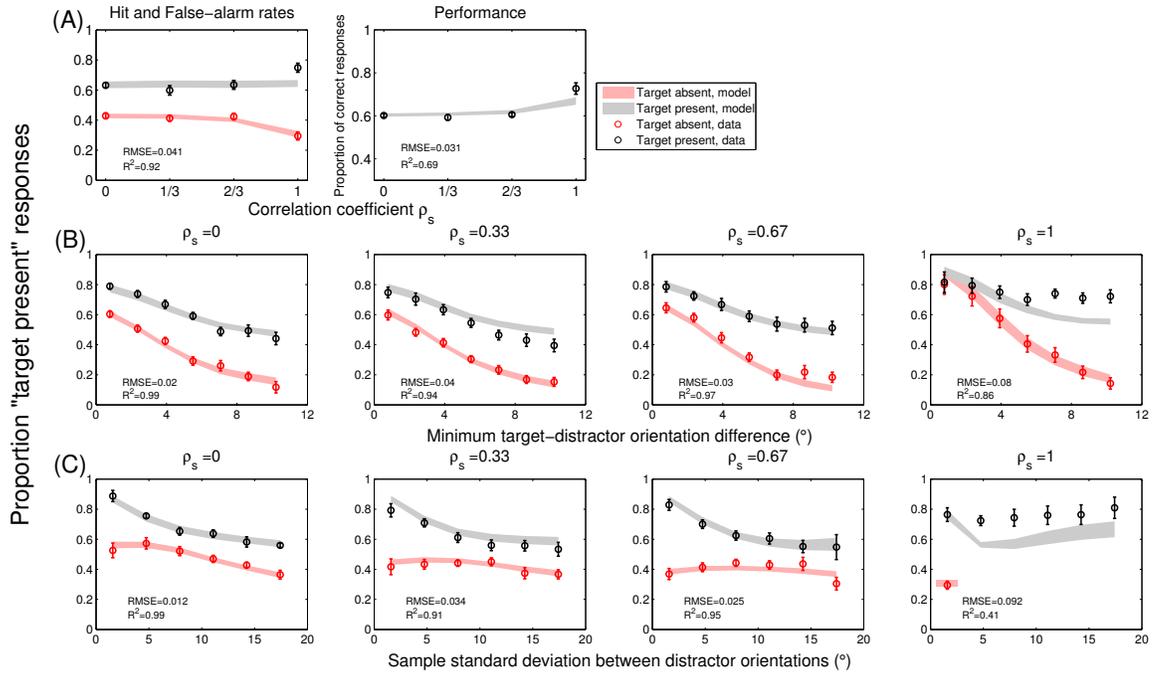
Figure 4.15: **VP model 4 ($\bar{J}$ condition-independent, and $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

Figure 4.16: **VP model 5 ($\bar{J}$ condition-independent, and $\rho_{S_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$ fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

4.4.2 Condition-dependent mean precision $\bar{J}$

In this section, we analyze the predictions of variable precision models that assume mean precision parameter $\bar{J}$ to be variable across experimental conditions, and hence we fit $\bar{J}$ per condition as a free parameter in the model for each subject. Models VP6 to VP10 described in Table 4.1 include this assumption about $\bar{J}$.

We examine the fits of VP6 model in Figure 4.17. We find that model predictions in this case are in close agreement to the data in all the conditions, and for all types of psychometric curves. The goodness of fit measured in terms of R^2 is above 0.9 for all curves, except for the minimum target-distractor plot in Figure 4.17(B, fourth panel) where it is 0.88 as the model misses few data points. But overall, we see a well match between the model predictions, and our data.

Thus, the question arises: do the subjects follow this model in their inference process? We note that VP6 model assumes that $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}}$ with varying $\bar{J}$ across experimental conditions. The good fits of this model indicate that subjects were able to infer the correct information about the generative model, and use it to make their responses. But, we must investigate other models with alternative assumptions about $\rho_{s_{\text{assumed}}}$ before making a conclusion here. We need to examine how other models perform under the assumption of varying $\bar{J}$ between experimental conditions.

To this end, we study the fitting of models VP7, VP8, VP9, and VP10 in Figures 4.18, 4.19, 4.20, and 4.21, respectively. We find that except for model VP7, all other models are equally good in their fits to the data. Model VP7 assumes
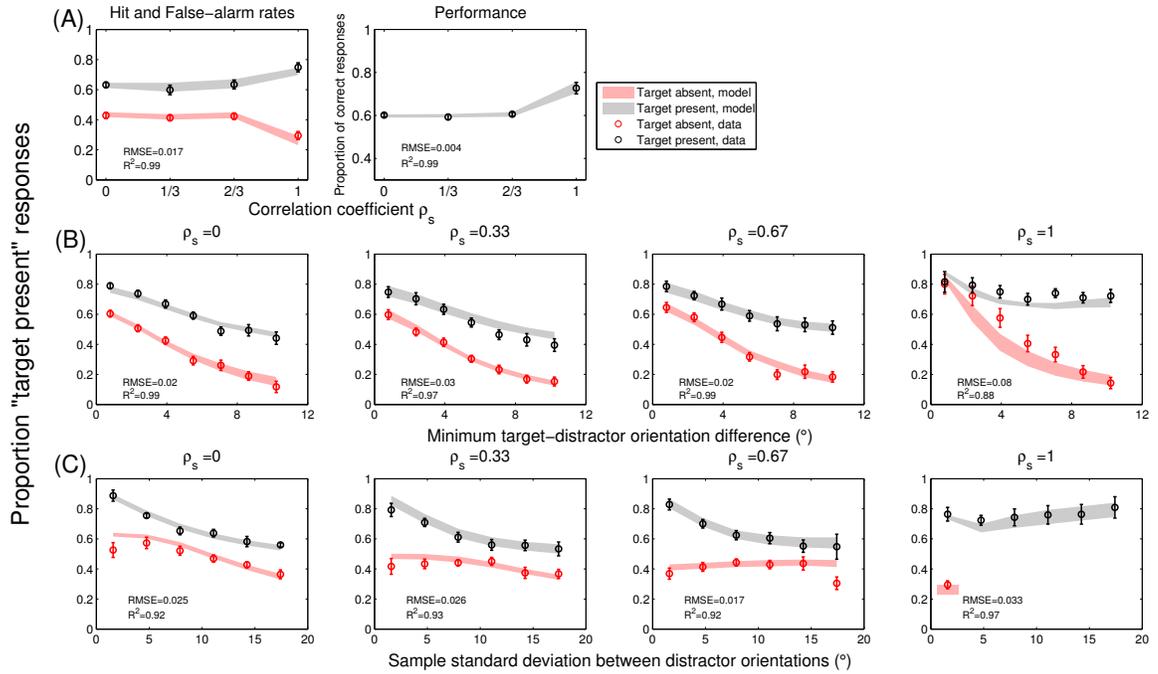
Figure 4.17: **VP model 6 ($\bar{J}$ condition-dependent, and $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}}$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

Figure 4.18: **VP model 7 ($\bar{J}$ condition-dependent, and $\rho_{s_{\textbf{assumed}}} = (0,0,0,0)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

that observers do not use any structural information in their decisions, and Figure 4.18(C, fourth panel) shows poor fits of the predictions of this model for the data having RMS error of 0.11. Based on this model, we can assert that subjects do use stimulus correlations in making their responses, and specifically the behavior of subjects is remarkably different in the homogeneous condition as compared to the partial correlation conditions.

This hypothesis is well tested with model VP9 that assumes such a structural form of $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$. Figure 4.20 illustrates a good match between the model generated curves, and the psychometric curves based on subjects' data. The goodness of fit R^2 ranges from 0.91 to 1 for different curves, and establishes that the model is successful in reproducing the responses of subjects on the experiment.

Further, the most general model VP10 perform equally well as models VP6, VP8, and VP9 in generating the predictions for the subjects' data. However, this model has more number of free parameters as we allow the possibility of $\rho_{s_{\text{assumed}}}$ fitted per condition. The magnitude of RMS errors for this model are comparable with other models in this category (except model VP7). This indicates that this is the best possible fitting we can obtain for our data given these models.

However, we still have minor deviation between model, and the data at one particular data point, the last data point in target-absent trials (red curve) in the sample standard deviation plot for $\rho_s = \frac{1}{3}$ in Figure 4.21(C, third panel). This point is nearly missed by predictions of other VP models in this category (compare third panel in row (C) of Figures 4.17, 4.19, 4.20, and 4.21). It indicates that
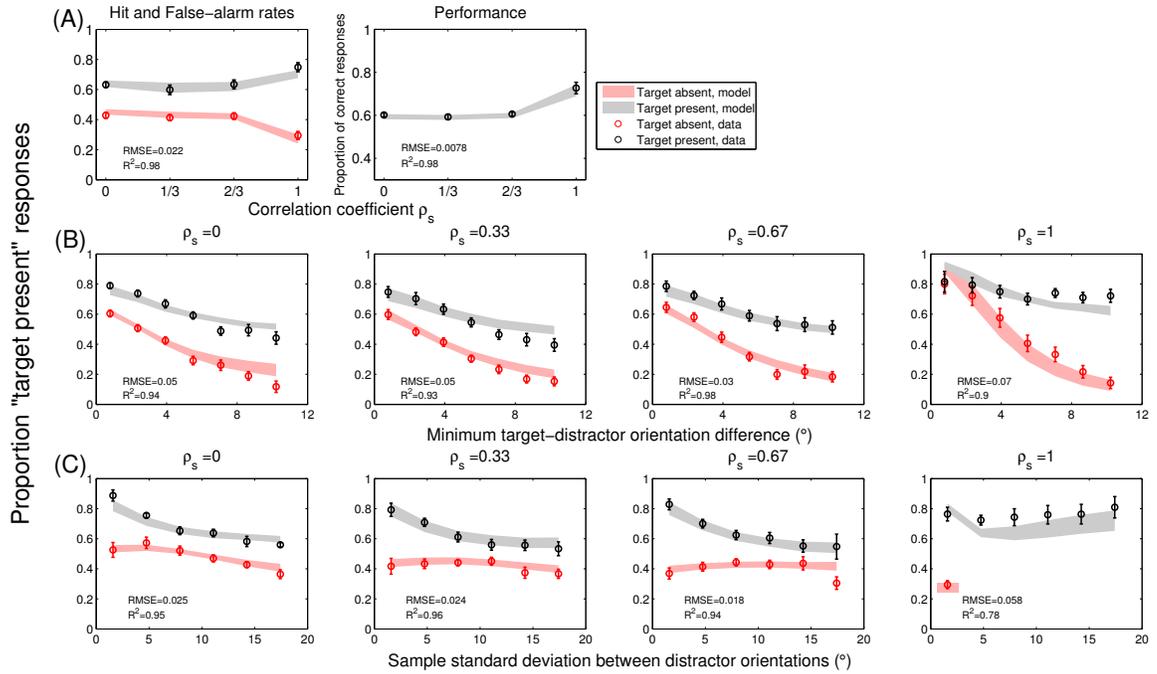
Figure 4.19: **VP model 8 ($\bar{J}$ condition-dependent, and $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$ fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

Figure 4.20: **VP model 9 ($\bar{J}$ condition-dependent, and $\rho_{s_{\mathbf{assumed}}} = (\alpha, \alpha, \alpha, \beta)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

Figure 4.21: **VP model 10 ($\bar{J}$ condition-dependent, and $\rho_{s_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$) fits for the data. (A) Hit and false-alarm rates.** (Left) Hit, and false-alarm rates as a function of correlation strength ρ_s used in the experimental conditions. (Right) Performance as a function of correlation strength ρ_s. **(B) Minimum target-distractor orientation difference.** Proportion "target present" responses as a function of minimum target-distractor orientation difference, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns). **(C) Sample standard deviation of distractor orientations.** Proportion "target present" responses as a function of sample standard deviation of distractor orientations, separately for target-present (gray), and target-absent (red) trials in each experimental condition (columns).

probably this is the best possible fitting we could obtain for the experimental data using these models. There may be something special about the characteristic of that point, and we may be missing some particular assumptions in our models that result in the deviation of models at that point. We may need to put additional assumptions in the model to incorporate the fit at that particular point in Figure 4.21(C, third panel). However, we are unaware of the mathematical complexities it will introduce in our existing high-dimensional models. Further, model VP9 at least does a better job in predicting the behavior at this last data point on target-absent trials in Figure 4.20(C, third panel). Therefore, for our purposes, we limit ourselves to this level of fitting for the data, and only focus on analyzing, and comparing these models here.

4.5 Need for model comparison

In earlier sections we analyzed the fits of equal, and variable precision models listed in Table 4.1. We find that none of the EP models could account for the behavior of subjects on the target detection experiment (described in Chapter 2). The variable precision models with the assumption of mean precision $\bar{J}$ constant across experimental conditions were also unsuccessful in reproducing the responses in the data. But VP models with variable $\bar{J}$ (except model VP7) provided good predictions for the data in terms of low RMS errors, and high R^2 values. However, we obtain more than one model that has equally well fits for the data. Specifically, models VP6, VP8, VP9, and VP10 all provide equally well predictions for the data.

But all these models have different underlying assumptions about ρ_{s_assumed}.

Therefore, we encounter the issue of discriminating multiple models that generate equally well predictions for the data. The question is how do we compare the models since each model has different number of free parameters, and different assumptions about those parameters. Further, how do we find the model that provides the best explanation for the data. We thus use the model comparison techniques discussed in Section 3.6 to compare different models, and find the best fitting model. We present the model comparison results for our models in the following chapter.

Chapter 5

Data analysis II: model comparison

The purpose of the experimental study described in Chapter 2, was to determine whether human observers take stimulus correlations into account in a target detection task. We investigate if observers are optimal in integrating the structural information in their decisions. If not, we wish to examine the alternative decision-making strategies observers use in their inference process.

To this end, we analyzed our collected experimental data in Chapter 4 using various models that could possibly explain how subjects made responses on the experiment. We found that the assumption of equal precision fails to account for the subjects' behavior. We further considered the variable precision models that test the possibility of observers using varying precision across trials, and stimuli. When we allowed the mean precision parameter of the gamma distribution used for modeling the random precision variable to be vary across experimental conditions, we discover several models that fit the data well. Particularly, models VP6,

VP8, VP9, and VP10 (Table 4.1) provided equally well matched predictions for the psychometric curves of the subjects' data. Given multiple models that generate equally well predictions of the data, we want to find the model that explains the data most parsimoniously.

This chapter is a continuation of the analysis done in Chapter 4, and the main focus here is to compare multiple models that may share a common characteristic or factor. We compare the models based on the Bayesian information, and Akaike information criteria, and find the model that best fits the data. We also use these criteria to provide an evidence why a certain category of models is preferred over others. For instance, we compare equal, and variable precision models, and show that all VP models are better than EP models. We also discriminate our models based on other factors described in the chapter. We conclude by analyzing the maximum-likelihood parameter estimates of the best fitting model, and interpret the behavior of subjects based on the best fitted model assumptions, and its fitted parameters.

5.1 Factors in the models

We note that our models listed in Table 4.1 have different number of parameters, and different assumptions about those parameters. Most of these models share an assumption about a particular parameter but, they have different assumptions about other parameters. For example, all models (EP1 to EP10, and VP1 to VP10) share an assumption about the prior over T (target presence variable), but they

differ in terms of assumptions about the encoding precision, and the correlation structure. We refer to a common characteristic (a parameter) as a *factor*, and we categorize our models based on different characteristics or factors.

We combine models sharing a common characteristic or factor in a group. There is more than one choice for each characteristic. We thus sub-divide each factor into *levels* or *family of models*. Each family has a common choice of model characteristic. We recognize the following three factors in our models: (1) precision, (2) dependence of precision parameter, and the (3) assumption about $\rho_{s_{\text{assumed}}}$. The first factor characterizes the nature of precision as either equal or variable. This factor includes two levels of EP, and VP models. The second factor describes how precision depends on experimental conditions. In particular, the parameters governing the precision can be condition-independent, or dependent. Lastly, we discriminate models based on the assumption of $\rho_{s_{\text{assumed}}}$. We have following five choices for $\rho_{s_{\text{assumed}}}$ in our models: $(0, \frac{1}{3}, \frac{2}{3}, 1)$, $(0, 0, 0, 0)$, $(\alpha, \alpha, \alpha, \alpha)$, $(\alpha, \alpha, \alpha, \beta)$, and $(\alpha, \beta, \gamma, \delta)$. Such a grouping of models into different families helps us in comparing multiple models together that share a common characteristic.

Therefore, we categorize our models based on three different factors, and each factor has different families of models given the possible choices for a factor. Table 5.1 presents a detailed classification of the models listed in Table 4.1 in different factors, and their respective levels.

Factor	Level	Models in the level	Number of models
Precision	EP	EP1 to EP10	10
	VP	VP1 to VP10	10
J or $\bar{J}$	condition-independent	EP1 to EP5, VP1 to VP5	10
	condition-dependent	EP6 to EP10, VP6 to VP10	10
$\rho_{s_{\text{assumed}}}$	$(0, \frac{1}{3}, \frac{2}{3}, 1)$	EP1, EP6, VP1, VP6	4
	$(0, 0, 0, 0)$	EP2, EP7, VP2, VP7	4
	$(\alpha, \alpha, \alpha, \alpha)$	EP3, EP8, VP3, VP8	4
	$(\alpha, \alpha, \alpha, \beta)$	EP4, EP9, VP4, VP9	4
	$(\alpha, \beta, \gamma, \delta)$	EP5, EP10, VP5, VP10	4

Table 5.1: **Factors based model list.** List of models belonging to different factors, and respective levels in each factor. Each factor has different number of levels, but all levels in the factor have equal number of models.

We compare the family of models across levels belonging to the same factor. For example, we compare the family of models belonging to equal, and variable precision models in the precision factor to determine the best level (either equal or variable precision) in the factor. We use both Bayesian information criterion (BIC), and Akaike information criterion (AIC) to compare our models. These criteria are based on the maximum log-likelihood of a model, and penalize the model for the additional free parameters. The penalizing term scales with the size of the data in BIC, while it is a multiple of 2 in the case of AIC. The model with lower

value of BIC (or AIC) wins over a model having a higher value of BIC (or AIC) in model selection based on Bayesian information criterion (or Akaike information criterion). These techniques are described in Section 3.6. We generally display the relative difference between a model, and possibly a best fitting model in our figures in this chapter. Therefore, a positive value of the relative difference for a model indicates that the model performs poorly in comparison to the hypothesized best fitting model. The magnitude of BIC (or AIC) difference between two models determines how closely the two models predict the behavior of observers on the experiment.

In the following sections, we discuss the comparison of models belonging to the same factor based on both BIC, and AIC. We present both individual based, and average subject based comparisons. We find that the two forms of comparisons (individual, and average) may lead to different winning models, thus we discuss the resulting consequences based on these comparisons.

We first examine the comparison between equal, and variable precision models corresponding to the first factor of precision.

5.2   Equal versus variable precision

We first compare models based on the precision factor. There are two model families (or levels) in this factor: equal, and variable precision. Each family has 10

models, namely EP1 to EP10 in equal precision, and VP1 to VP10 in variable precision model family.

We analyzed the model fits of both EP, and VP models in Chapter 4. We find that most of the EP models fit the data poorly compared to their variable precision counterparts. However, some of the EP models were still close to some VP models in their predictions for the psychometric curves. For instance, the predictions of model EP10 in Figure 4.11 are comparable to VP1 (Figure 4.12) model, while they are better as compared to VP2 (Figure 4.13) model. We want to determine which model is preferred for the individual subjects, and which model wins in the average (over subjects) comparison.

Figure 5.1 displays the comparison results of EP models relative to their VP versions based on both BIC (top), and AIC (bottom). Each bar in the figure represents the BIC (or AIC) difference of an EP model from its VP counterpart. For instance, we consider the difference of EP1, and VP1 models, and similarly EP2, and VP2 model. Figure 5.1(A, left) shows the BIC differences of EP models from VP for individual subjects. Similarly, left panel in (B) shows the AIC differences between the two level of models. We note that both BIC, and AIC differences are large positive values for each subject. This signifies that VP models outperform their equivalent models with equal precision. Moreover, the difference is huge (more than 500 points) for most of the models for majority of subjects indicating the poor predictions of EP models as compared to variable precision models.

We also consider the average BIC (right panel, Figure 5.1(A)), and AIC difference (right panel, Figure 5.1(B)) between EP, and VP models. Clearly, we see that

all EP models have large positive (more than 200 points) BIC, and AIC differences as compared to their VP equivalents. Therefore, the comparison in Figure 5.1 provides strong evidence for the preference of VP models over EP models.
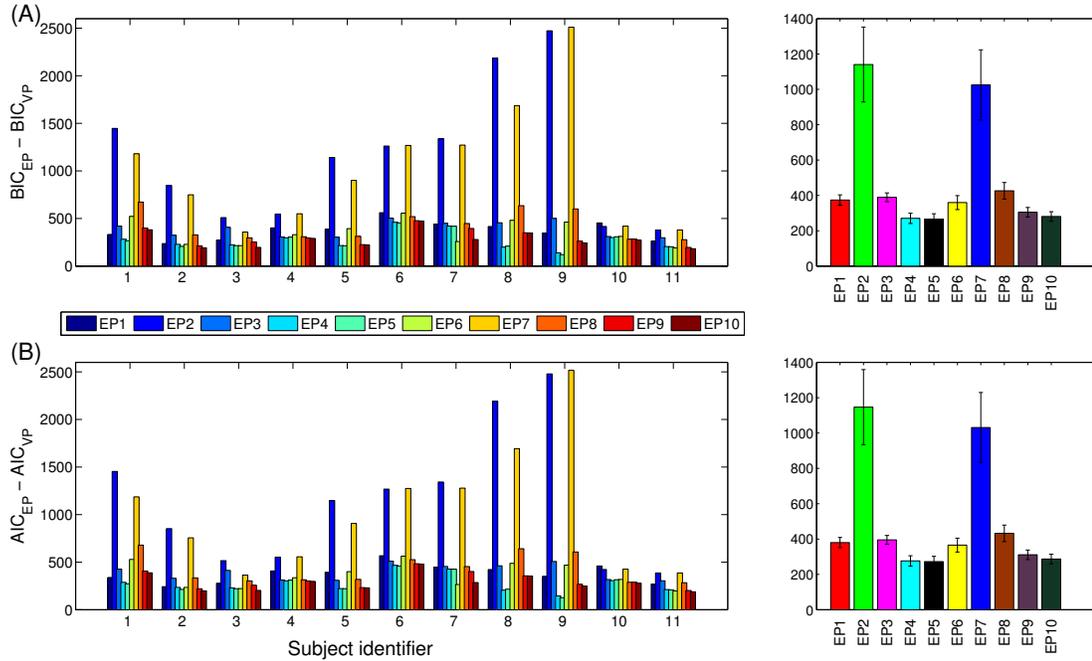
Figure 5.1: **BIC and AIC model comparisons: equal and variable precision models. (A) BIC model comparison.** (Left) The relative BIC differences of equal precision models with respect to variable precision models are shown for all 11 subjects participated in the experiment. Each bar represents the BIC difference of an EP model from its equivalent VP model for a subject. (Right) The averaged BIC differences across subjects in the experiment. **(B) AIC model comparison.** AIC relative difference of EP models with respect to VP models for each subject (left), and average over subjects (right). A positive BIC or AIC difference indicate that the EP model is worse than its equivalent VP model. Throughout the chapter, the error bars indicate the unit standard error mean (s.e.m).

5.3 Condition-independent $\bar{J}$ versus condition- dependent $\bar{J}$

We have observed in Section 4.4 that variable precision models better explained the data than equal precision models (Section 4.3). Also, Figure 5.1 demonstrates the strong preference of VP models over their EP equivalents based on Bayesian information, and Akaike information criteria. Therefore, for our further model comparisons we only present results with variable precision models since EP models have significantly higher values of BIC, and AIC than VP models.

We are now interested in determining which of the two models are preferred in model comparison: VP models with mean precision $\bar{J}$ condition-independent, or with $\bar{J}$ condition-dependent. We have seen in Section 4.4 that condition-dependent models (VP6, VP8, VP9, and VP10) fit the data well, while condition-independent models (VP1 to VP5) mis-fit in the case of homogeneous condition. Thus, in terms of model predictions, models with condition-dependent $\bar{J}$ outperform models with condition-dependent precision parameter. However, VP models with condition-independent $\bar{J}$ have many additional parameters (see Table 4.1), and that could be one of the reason for their better fits.

Therefore, we compare the two levels of variable precision models: with condition-independent $\bar{J}$ (models VP1 to VP5), and condition-dependent $\bar{J}$ (models VP6 to VP10) using BIC, and AIC in Figure 5.2. The figure displays the BIC, and AIC differences of condition-independent models relative to their condition-dependent

counterparts in the assumption of ρ_s. For example, the first bar corresponding to the first subject in Figure 5.2(A, left panel) represents the BIC difference of VP1 model with respect to VP6 model (they both have assumption about $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}}$). Hence, the different color bars represent the following differences: VP1-VP6, VP2-VP7, VP3-VP8, VP4-VP9, and VP5-VP10. Clearly, a positive value of the difference indicates that condition-dependent model has preferable BIC value.
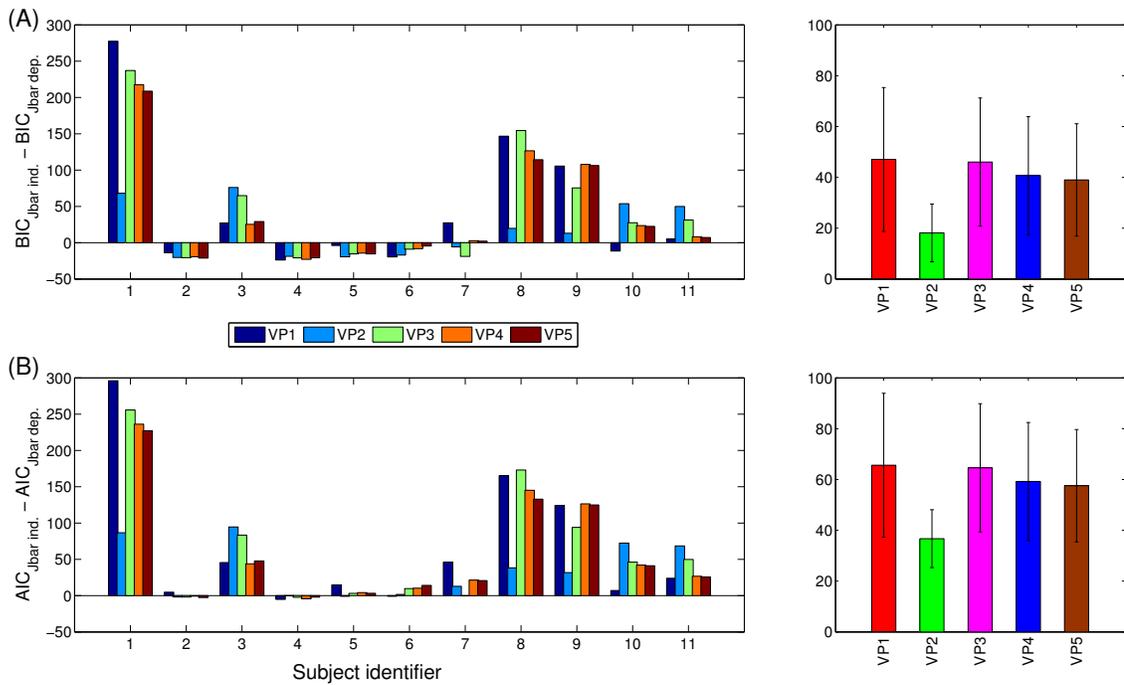
Figure 5.2: **BIC and AIC model comparisons: VP models with experimental condition-independent mean precision $\bar{J}$, and condition-dependent $\bar{J}$. (A) BIC model comparison.** The relative BIC differences of VP models with condition-independent mean precision with respect to the VP models that assume variable mean precision experimental conditions, are shown for each subject (left), and average over subjects (right). **(B) AIC model comparison.** AIC differences of VP models with $\bar{J}$ condition-independent relative to VP models with $\bar{J}$ experimental condition-dependent for each subject (left), and average over subjects (right).

We observe in Figure 5.2(A, left panel) that there is no clear winning trend of a particular level for individuals in the BIC comparison of models. Five subjects (out of 11) show preference for models with condition-dependent $\bar{J}$, while 4 subjects prefer condition-independent models. The other two subjects have mixed preference depending on the assumption of ρ_s in the different models. Similarly, the AIC comparison in the left panel of Figure 5.2(B) does not lead to conclusive results for a few subjects. However, in this case, majority of subjects' data is better fitted with condition-dependent $\bar{J}$, and even the differences are large between the two levels. The variation between BIC, and AIC comparisons is mainly because of differently scaled penalty terms in the two criteria. Bayesian information criterion more strongly penalizes a model for its parameter size, as compared to the Akaike information criterion. We note that the models with condition-dependent precision parameter have 3 extra parameters as compared to models belonging to the condition-independent level (see Table 4.1). Therefore, BIC strongly penalizes these models, and we see the results are inconclusive in Figure 5.2(A, left panel).

We also compared the models in the two levels by taking an average of the BIC, and AIC values of individual subjects. The results are presented in the right panels of Figure 5.2(A), and (B). The average comparisons show a definite preference of condition-dependent models over models with condition-independent precision. The large magnitude of difference between both levels is mainly guided by the fact that some of the subjects show a strong preference for model with condition-dependent mean precision in variable precision models. However, the standard

147

errors are considerable, and reflect the inconsistent choice of a family model between subjects.

Therefore, we find that condition-dependent $\bar{J}$ models outperform models with condition-independent $\bar{J}$ in average BIC, and AIC model selection. We also observed in Section 4.4 that predictions of these models fit well to the psychometric curves of the subjects. But, we note that the choice of a model family differ across individuals, and hence it would be difficult to conclude that VP models with condition-dependent $\bar{J}$ represent the best family of models for all subjects.

However, to find the model with the favorable assumption of ρ_s, we only consider variable precision models with condition-dependent mean precision parameter. This is because only the models in this category successfully fit the data. Moreover, these models show a winning margin in the average comparison over other models in the condition-independent level as seen in Figure 5.2(A), and (B).

5.4 Comparison based on $\rho_{S_{\textbf{assumed}}}$

Thus far, we find that variable precision models outperform equal precision models by a huge margin (Figure 5.1). Further, we observe that VP models with condition-dependent mean precision are better in generating predictions for the data as compared to the models with condition-independent $\bar{J}$ (Figure 5.2). Both these results are focused on the precision factor. We are now interested in finding the favorable model having an assumption about ρ_s that best predicts the behavior

of subjects on our target detection experiment.

Our models described in Section 4.2.1 have distinct assumptions about ρ_s, and in particular, we have considered five levels of models $\rho_{s_{assumed}} = \rho_{s_{true}}$, $\rho_{s_{assumed}} = (0,0,0,0)$, $\rho_{s_{assumed}} = (\alpha,\alpha,\alpha,\alpha)$, $\rho_{s_{assumed}} = (\alpha,\alpha,\alpha,\beta)$, and $\rho_{s_{assumed}} = (\alpha,\beta,\gamma,\delta)$. Analyzing the model predictions in Sections 4.3, and 4.4, we find that there are multiple models that fit the data equally well (see Figures 4.17, 4.19, 4.20, and 4.21). Namely, the models are VP6 ($\rho_{s_{assumed}} = \rho_{s_{true}}$), VP8 ($\rho_{s_{assumed}} = (\alpha,\alpha,\alpha,\alpha)$), VP9 ($\rho_{s_{assumed}} = (\alpha,\alpha,\alpha,\beta)$), and VP10 ($\rho_{s_{assumed}} = (\alpha,\beta,\gamma,\delta)$). Thus, there are multiple assumptions of ρ_s that equally well predict the behavior of subjects. The question thus arises: which one of these models best represents the subjects' responses? To answer this question, we compare models VP6 to VP10 using BIC, and AIC. We note that all these models have a common assumption about encoding precision, and only differ in the hypothesis about ρ_s.

Figure 5.3 displays the BIC (top), and AIC (bottom) differences for individual subjects (left panels), and the average subject comparison (right panels). The average BIC, and AIC model comparisons in Figure 5.3(A, right panel), and (B, right panel) show that each criterion chooses a different favorable model. Model VP9 has the least BIC value in the group, while VP10 is the preferable model based on AIC comparison. But the two models only differ by a few points. The relative BIC difference between VP9, and VP10 models is 6.91, while the difference between the two is about 5.46 points in the case of AIC. This difference is really small to distinguish the two models apart, and suggests that both models make similar predictions for the subject's responses on the experiment.
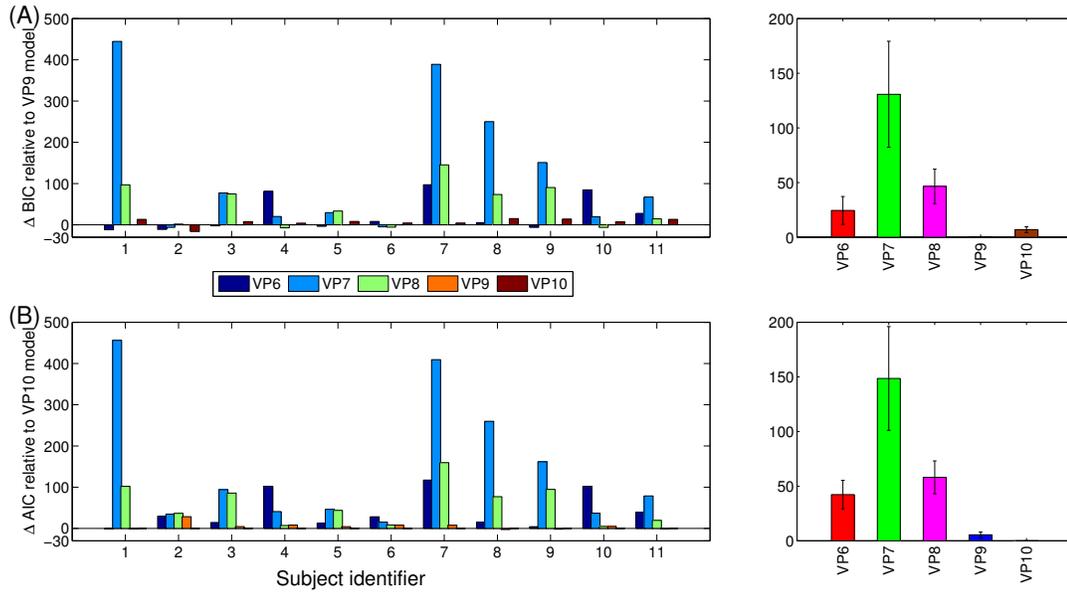
Figure 5.3: **BIC and AIC model comparisons: VP models with condition-dependent $\bar{J}$, and different assumptions about $\rho_{S_{\mathbf{assumed}}}$.** **(A) BIC model comparison.** The relative BIC differences of VP models having different assumptions about $\rho_{S_{\mathrm{assumed}}}$ with respect to VP9 model (with $\rho_{S_{\mathrm{assumed}}} = (\alpha, \alpha, \alpha, \beta)$) for each subject (left), and average over subjects (right). **(B) AIC model comparison.** AIC differences of condition-dependent mean precision VP models relative to VP10 model with $\rho_{S_{\mathrm{assumed}}} = (\alpha, \beta, \gamma, \delta)$) for each subject (left), and average over subjects (right).

Thus, BIC, and AIC are inconsistent in the choice of the favorable model. This is mainly due to the different penalty terms involved in the criteria. We note that model VP10 has 10 parameters, while VP9 has 8 free parameters (Table 4.1). We recall that each subject performed 3600 trials in the experiment (Section 2.5.3), hence, the penalty term in BIC (Eq. (3.6)) scales each additional parameter in the model by $\log(3600) \approx 8.19$. This amounts to $\log(3600) \times 10 \approx 81.9$ penalty points for a subject in the case of VP10 model. On the other hand, the penalty term is about 65.5 points for VP9 model. Therefore, the model comparison based on BIC would prefer the VP9 model in case the maximum log-likelihoods of the two models are comparable. The comparison in the right panel of Figure 5.3(A) illustrates such a preference of VP9 model as compared to the higher dimensional VP10 model.

Similarly, we can understand the average AIC comparison in Figure 5.3(B, right panel). The two models only differ by 4 points in AIC penalty term, and hence in AIC model comparison, the most general model VP10 is preferred over others.

We also find that VP7, and VP8 models with the assumptions of zero, and constant correlations perform poorly in both BIC, and AIC average model comparisons. The large relative differences of these models with respect to the winning models indicate the poor performance of these models for the data. We observe that the VP6 model having the assumption of $\rho_{S_{\text{assumed}}} = \rho_{S_{\text{true}}}$ also fails in BIC, and AIC average comparisons by 24.4, and 42.3 points, respectively. Therefore, models VP9, and VP10 provide better description of the data than the other three

models. This indicates that subjects make suboptimal inferences on the experiment, and use incorrect assumptions about the generative model (described in Figure 2.1) in detecting a target on the task.

However, we observe a large variation in model preference among individuals as seen in the left panels of Figure 5.3(A), and (B). The figure shows the relative BIC differences of other models with respect to the VP9 model, while on AIC comparison, the differences are relative to model VP10. We find that there is no unique winning model for all individuals. Models VP6, VP9, and VP10 are really close in predictions for most of the subjects, and only differ by a few points. This suggest that these models are hardly distinguishable on the BIC, and AIC measures based on individual model comparisons. Perhaps subjects use different strategies to make decisions on the experiment. However, it is difficult to characterize such a possibility given the limitations of the model comparison techniques we use here.

Therefore, based on the model selection results for models VP7, and VP8 for most of the subjects, we conclude that subjects do take into account the statistical structure present in visual scenes to make decisions. Further, they also discriminate the different experimental conditions, and use different strengths of correlations. However, it is difficult to find the best model between VP9, and VP10 models since they are relatively close in both AIC, and BIC comparisons (Figure 5.3). Hence, it is hard to say whether subjects use different strengths of correlations per experimental condition (based on VP10 model), or they treat the first three experimental conditions similarly, and the homogeneous condition differently as hypothesized by VP9 model.

In order to choose the best model, we further perform a rejection analysis on our models below.

5.5 Rejection analysis

The model comparisons in previous sections suggest that the variable precision, and assumption about ρ_s are most important ingredients of successful models. However, we are unable to determine the best model in absolute sense based on BIC, and AIC comparisons. Therefore, we now wish to determine what levels of each of the three factors best describe the subjects' behavior on the experiment. Table 5.1 describes the three different factors present among our 20 models. Further, each factor has different levels or model families that share a common assumption in the factor. Model families are subset of all models that share a particular level of a particular factor, regardless of their levels of the other factors.

We define the *rejection criterion* as the BIC (or AIC) difference of a model with respect to the winning model for each subject. For each model family in a factor, we compute the number of subjects that reject all its members as a function of the rejection criterion. That is, we first find the best model (among all models) for each subject, and then for each level in a factor, we count the number of subjects that reject all models belonging to that particular level for a given criterion. We repeat this process at all values of the (BIC, or AIC) rejection criterion, and plot the number of subjects that rejected a particular model family as a function of the criterion in Figure 5.4.

(A)

(B)

(C)
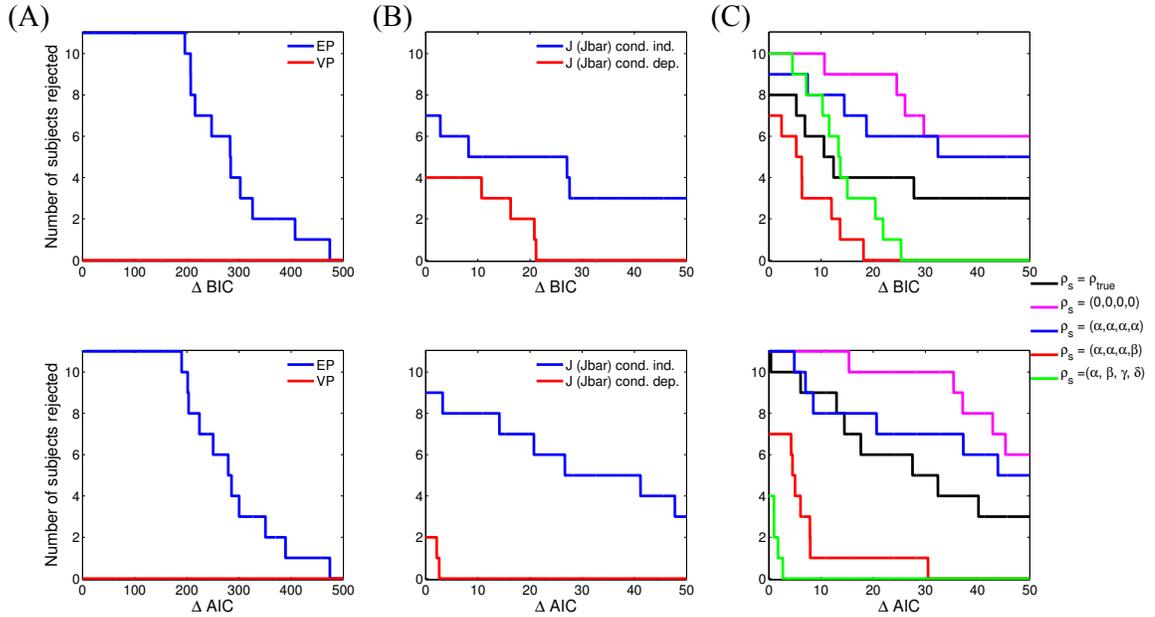
Figure 5.4: **Rejection curves: model comparison between model families for each model factor.** Each column corresponds to a factor, and a color represents a particular model family or level belonging to a factor. The model comparison is based on all 20 models for each factor. **(A) Factor 1: precision.** Number of subjects for whom *all* models belonging to a certain family or level (EP or VP) are rejected as a function of the rejection criterion based on BIC(top), and AIC (bottom) values. A model is rejected if it has a higher BIC, or AIC than that of the winning model for a subject. For example, when both BIC, and AIC rejection criteria is 100, all models of the EP family are rejected, while none of the subjects reject VP models. **(B) Factor 2: dependence of precision parameters on experimental conditions.** Similar to (A). **(C) Factor 3: assumption about correlation strength** $\rho_{s_{\text{assumed}}}$**.** Similar to (A).

Figure 5.4(A) shows the results for the first model factor, the assumption of encoding precision. Regardless of the rejection criterion, the entire family of equal precision models is rejected for majority of subjects in terms of both BIC (top), and AIC (bottom). Moreover, all subjects reject EP models for a large criterion difference of up to 200 points. While, none of the subjects reject the variable precision model family indicating that all individuals have their best models in this model family. This provides a strong evidence that variable precision models are indeed better in describing the behavior of subjects on visual search tasks as found recently in many studies [94, 141, 95, 129, 140].

Figure 5.4(B) shows the rejection analysis for the second factor, the dependence or independence of the encoding precision on experimental conditions. At any given rejection criterion for both BIC (top), and AIC (bottom), most of the subjects select the model family with condition-independent precision parameter. The number of subjects under a rejection criterion of 0 determine the number of subjects for which the winning model belongs to the model family being considered. For example, in the top panel of Figure 5.4(B), there are 4 subjects that rejected condition-dependent encoding precision models, while 7 rejected the other family of models. This implies that based on BIC model selection, majority of subjects' responses (63.6%) are best explained by models belonging to the condition-dependent family. On the other hand, 9 subjects have winning model in this family based on AIC differences (bottom panel of Figure 5.4(B)). Further, all subjects prefer model family with condition-dependent precision as the rejection criterion becomes large. Overall, there is a clear separation between the two

155

model families, and subjects more frequently reject the models that assume the encoding precision to be constant across experimental conditions.

Finally, the rejection plots based on the correlation factor are displayed in Figure 5.4(C). There are five levels in this factor corresponding to the different assumptions about ρ_s. These levels are described in Table 5.1. We observe that the distinction between models is less clear in terms of BIC differences (top). But the models with the assumption of $\rho_{s_{\text{assumed}}} = (0,0,0,0)$, and $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \alpha)$ are more frequently rejected by subjects at any given rejection criterion for both BIC (top), and AIC (bottom) differences. Next, we note that most of the subjects reject the model with $\rho_{s_{\text{assumed}}} = \rho_{s_{\text{true}}} = (0, \frac{1}{3}, \frac{2}{3}, 1)$ than with the assumptions of $\rho_{s_{\text{assumed}}} = (\alpha, \alpha, \alpha, \beta)$, and $\rho_{s_{\text{assumed}}} = (\alpha, \beta, \gamma, \delta)$. This suggest that subjects are suboptimal in inferring the true values of correlation coefficient ρ_s among stimuli in the experiment. However, it is difficult to determine which of the two strategies they follow. We see a mixed result in terms of BIC (top), and AIC (bottom) differences. The models with $\rho_{s_{\text{assumed}}}$ fitted per condition is less frequently rejected by individuals as compared to the model family that assumes constant correlations across first three experimental conditions, and a different in the case of homogeneous distractors. While, we see a reverse trend for the choice of model on AIC rejection criterion (bottom). Therefore, there is a conflict between the choice of the best model based on the two criteria. It may be the case that the mixed result here reflects individual differences: some subjects may be using different correlations across experimental conditions, while some may be treating the conditions with partial correlations identically, and the homogeneous condition differently.

These results also agree with our earlier observations about the model predictions for VP9, and VP10 models in Figures 4.20, and 4.21, respectively. The two variable precision models equally well predict the subjects' data, and also model families with these assumptions on $\rho_{s_{\text{assumed}}}$ show a comparable preference in model comparison(Figure 5.3), and on rejection plots (Figure 5.4).

Next, we compare the maximum-likelihood parameter estimates of the VP9, and VP10 models. The parameter estimates may reflect the similarity between the two models.

5.6 Parameter estimates

Parameter estimates for the models VP9, and VP10 are given in Table 5.2. Here $\bar{J}_i$ denotes the estimated mean precision for the gamma distribution over precision in the ith condition. Similarly, ρ_i represents the estimated parameter value of $\rho_{s_{\text{assumed}}}$ in the ith condition of the experiment. The parameters τ, the scale parameter of the gamma distribution over precision, and p_T, the observer's prior probability that the target is present, are estimated as a constant parameter across all conditions in the experiment.

Model	Parameter	Mean $\pm$ SEM	Median
	$\bar{J}_1$	0.1097 ± 0.0184	0.1036
	$\bar{J}_2$	0.1077 ± 0.0206	0.1036
	$\bar{J}_3$	0.1148 ± 0.0227	0.1225

Continued on next page

Table 5.2 – *Continued from previous page*

Model	Parameter	Mean $\pm$ SEM	Median
VP9	$\bar{J}_4$	0.2651 ± 0.0816	0.1447
	ρ_{123}	0.4030 ± 0.0793	0.5
	ρ_4	0.7515 ± 0.1010	0.9333
	τ	0.7433 ± 0.2877	0.3683
	p_T	0.5048 ± 0.0046	0.5
VP10	$\bar{J}_1$	0.0981 ± 0.0158	0.0877
	$\bar{J}_2$	0.0937 ± 0.0167	0.0742
	$\bar{J}_3$	0.1022 ± 0.0185	0.0877
	$\bar{J}_4$	0.2376 ± 0.0623	0.1447
	ρ_1	0.4333 ± 0.0737	0.5
	ρ_2	0.3939 ± 0.0818	0.4667
	ρ_3	0.4818 ± 0.0896	0.4667
	ρ_4	0.7939 ± 0.0856	0.9333
	τ	0.5691 ± 0.1718	0.3683
	p_T	0.5036 ± 0.0043	0.5

Table 5.2: **Maximum-likelihood parameter estimates of VP9 and VP10 models.** The estimates of the mean precision, and scale parameter of the gamma distribution over precision are given along with the estimates of correlation strength ρ_{s_assumed} in each experimental condition (where applicable), and of p_T, the observer's prior probability that the target is present.

We observe that the mean estimated values of $\bar{J}_i$ are close for $i = 1, 2, 3$, while the magnitude order of the mean value of $\bar{J}_4$ differs significantly in both models. This suggests that subjects probably use similar levels of precision in case of experimental conditions with partially correlated stimuli, while they have higher precision in detecting a target on the homogeneous condition.

The trend in the values of $\bar{J}$ matches our expectation since finding a target among identically oriented distractors would be easier compared to the randomly oriented stimuli. Therefore, an observer would be more precise in the experimental condition with $\rho_s = 1$. Also, we note that subjects use near 0.5 prior for the target present probability in both models. Further, the maximum-likelihood estimates of the two models suggest that subjects over estimate the low correlation strengths, but perform nearly optimally when distractors are perfectly correlated.

Figure 5.5 shows the behavior of individual parameter estimates for model VP9 in (A), and VP10 in (B). We note that the estimates for both VP9, and VP10 models have a similar trend. The parameters closely follow a similar behavior for most subjects, and the mean parameter estimates behave nearly identically. The prior over target presence (first column), and τ (second column) show similar estimates in the two models for majority of individuals. Also, the values of $\bar{J}$ (third column) looks nearly identical in both models for most of the subjects, except for minor deviations. The mean estimated values of $\bar{J}$ show a near constant trend in the first three experimental conditions for the model VP10 (Figure 5.5(B, third column)). Further, the estimates of $\rho_{s_{\text{assumed}}}$ are close in the partial correlation conditions.
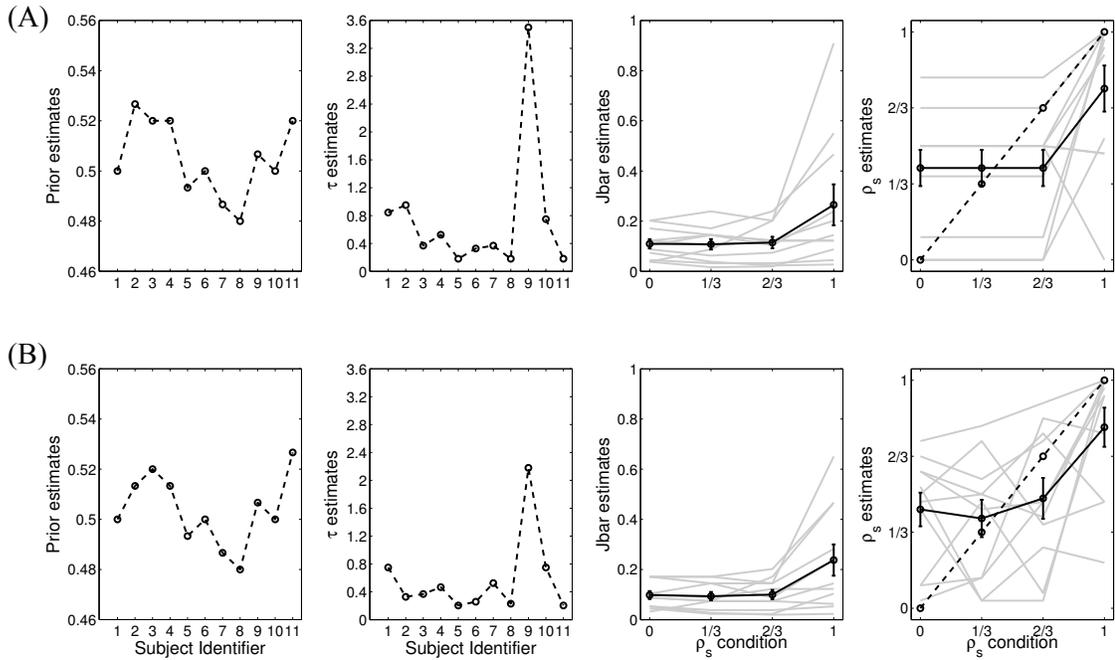
Figure 5.5: **Individual parameter estimates of VP9 and VP10 models. (A) Parameter estimates VP9 model.** Maximum-likelihood estimates of p_T, the observer's prior probability of target presence (first column), and the scale parameter τ (second column) for each subject. Individual (gray lines), and average (over subjects) (black circles) estimates of mean precision $\bar{J}$ in different experimental conditions (third column). Estimated values of correlation coefficient ρ_s for each subject (gray lines), and average (over subjects) ρ_s estimates (black circles) as a function of true correlation strength (diagonal) in the experimental conditions.

A similar trend in parameter estimates of the two models provides an explanation for the equally well model predictions of both models for the data, and the conflicting choice of models on BIC, and AIC model comparisons.

5.7 Conclusions and discussion

Based on the results of model predictions (Sections 4.3, and 4.4), and model comparisons discussed above, we conclude that subjects use the information about the statistical structure present in visual scenes to make decisions. That is, they take correlations of the distractors into account in detecting a target on the visual search experiment. However, they are suboptimal in inferring the true correlation strengths that were used to generate the stimuli in different experimental conditions. The model comparison results based on Bayesian information criterion (Figures 5.3(A, and 5.4(C, top panel)) suggest that perhaps subjects treat the partial correlation conditions identically, and find it difficult to distinguish the distinct level of weak correlations present in the scenes. However, they behave differently when the distractors are homogeneous, and show an improved performance in the average results.

While the results based on Akaike information criterion (Figures 5.3(B), and 5.4(C, bottom panel)) provide evidence that perhaps subjects infer different levels of correlation strengths in different experimental conditions. But they infer incorrect values leading to the suboptimal behavior. Further, we explore that subjects

use variable precision in making decisions ((Figures 5.1, and 5.4(A)). This precision not only varies over trials, and stimuli, but also vary across experimental conditions (Figures 5.2, and 5.4(B)). Thus, we find that distribution of encoding precision depends on the strength of correlation used in an experimental condition.

We obtain inconclusive results here about how subjects actually infer the different correlation strengths. This could be because of many underlying reasons. It may be that the individuals' differences lead to mixed results: subjects probably use different strategies to make decisions on the task. This would be difficult to examine using the measures we used, and would probably involve a detailed analysis of each subject's responses. It could also be possible that the models we considered here lack some assumptions about a parameter that could be driving subjects' decisions but we are unaware of it. For instance, we assume that subjects learn the parameter value of the standard deviation of distractors in the generative model (Figure 2.1). However, this may not necessarily be true, and in that case, it would require models that assume σ_s as a free parameter, and the model decision variable would need to computed by marginalizing over all possible values of this parameter. Further, it could be possible that subjects do not make point estimates of the correlation strengths, instead they use some unknown distribution over the true values. Even worse, another possibility would be if subjects do not at all use any information about the correlations, but instead use some suboptimal decision strategies, for example, a threshold rule (similar to the one described in Chapter 1 Example 1.4.1).

All these alternatives are possible, and may provide more explanation about subjects' behavior on the experiment. However, all these models are more likely to be more complex, and multi-dimensional. The decision variables for these models would be even more complex (compared to Eq. (2.21)) except in the case of threshold models. Therefore, it is bit uncertain that observers would use even more complex decision variables in their decisions. And if they really do, a mathematical treatment of such models would be highly intricate, and may not be possible in many cases. To explore such possibilities, an advance treatment, and model comparisons based on hierarchal modeling would probably be required.

Chapter 6

Noise correlations in a single target detection task

In many mid- to high-level visual tasks with multiple stimuli, the brain has to make categorical, global judgements. This involves extracting the relevant information from sensory input. Given the nature of a task, the brain needs to follow different mechanisms relevant to the objective of the task. For instance, in a target detection task where the goal is to detect whether a predefined object is present in a visual scene having multiple objects, the identity of any individual stimulus may not be of direct relevance in making a decision. But, in an estimation task where an observer needs to estimate the mean orientation of the stimuli in a visual display, orientation of each stimulus is equally important. Therefore, the objective of a task could play a crucial role in guiding our inference process, and how the brain extracts the meaningful information from sensory input. We explore the

importance of the objective of a task in some details in Chapter 7.

In addition to a task relevant feature, our judgements also critically depend on the accuracy of our sensory measurements. The sensory information our brain receives is usually uncertain since noise corrupts our measurements, especially when observation time is short, and multiple objects are present. The magnitude of noise in the measurements governs the accuracy of our decisions on a task, and further complicates the our inference process. Thus, measurement noise can considerably affect our decisions on a task.

Extensive work has been done both at theoretical, and experimental levels to understand the decision processes of the brain. Specifically, several models have been proposed to study the mechanisms by which the brain converts noisy sensory measurements of a set of stimuli to infer the state of the world. For example, how the brain infers a target presence or absence in a scene, or how to discriminate the mean orientation of a set of stimuli. These models generally consider decision rules that are applied to the measurements. On the other hand, the measurements themselves are usually modeled in a rather conventional fashion. They are often considered as independent (between stimuli), and normally distributed [90, 142, 94, 95]. We also use these assumptions to model an observer's measurements in Chapter 2. But, both these assumptions can be questioned.

It has been found that neural correlations can extend to distances as long as 4mm in monkey cortex [28, 36]. This indicates that the sensory measurements can be strongly correlated [118, 27]. These correlations must be accounted in the modeling of decision processes. Therefore, we are interested in examining the

effects of correlated sensory measurements on the inference process of our brain, and the accuracy of decisions.

In order to make categorical, global judgements, the brain not only needs to take into account the correlations present in the measurements, but also the statistical structure of the stimuli on a task. For instance, a target could be easily detected in the case of homogeneous distractors than when the distractors possess random orientations. In such a case, correlations between sensory measurements can further influence decisions. For instance, we consider the situation discussed by Mazyar et al. [94]. Strong measurement correlations result in more similarity between the measurements, and introduce structure. This structure could be helpful in presence of identical distractors, as similar measurements corresponding to the distractors can be grouped, and a target can be easily detected if present. By contrast, when distractors are independently drawn, there is no external stimulus structure that could be preserved, and hence, we do not necessarily expect strong measurement correlations to be beneficial. This illustrates that measurement, and stimulus correlations should not be considered in isolation.

We expect that correlations between measurements, and those between stimuli will interact to jointly influence the decisions of an observer. In this chapter, we examine how the correlations between measurements, and stimuli interplay to affect the performance of an ideal Bayesian observer in a visual search task. Specifically, we consider a target detection task similar to the one described in Section 2.1 but with the assumption of correlated measurements.

We first introduce the set up of the model for a target detection task that we use

to study the effects of measurement correlations along with correlations between distractors. We then derive the decision rule for an optimal Bayesian observer to make decisions on such a task. Our goal here is to understand how external structure together with the structure of the measurements impacts the performance of an ideal observer. Thus, we vary different parameters that determine external stimulus, and internal measurements structure, and examine the resulting behavior in the performance.

6.1 Model description

To examine how decisions of an ideal observer are determined by the statistical structure of sensory measurements, and stimuli we consider a single target detection task; where the observer is required to detect a vertical target in a set of N stimuli. The framework of the task, and mathematical notations are similar to Section 2.1. The binary variable, T, indicates target presence for $T = 1$, and target absence when $T = 0$. In half of the trials, the target is present at one of the N possible locations. Stimulus orientations, denoted by $\mathbf{s} = (s_1, s_2, \cdots, s_N)$ are the relevant characteristics of the task. The target stimulus orientation is denoted by $s_T = 0$, and the orientations of the distractors are drawn from multivariate normal distributions specified in Eqs. (2.1), and (2.6) when $T = 0$, and $T = 1$, respectively. We note that the external structure in a scene is determined by the number of stimuli, N, the variance of the distractor orientations, σ_s^2, and the pairwise correlation coefficient, ρ_s. For the purpose of this task, we consider that the number of

stimuli, and the external noise σ_s are fixed parameters. We control the amount of structure in a scene by varying the amount of correlation between distractors, that is, by varying ρ_s.

We further assume that an observer makes a noisy *measurement* of each stimulus, denoted by x_i. This measurement can be thought of as the maximum likelihood estimate obtained from the activity in a population of neurons with receptive fields including location i. We denote by $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ the vector of N measurements. It is commonly assumed that the components of $\mathbf{x}$ are unbiased, independent and normally distributed

$$\mathbf{x}_i | \mathbf{s}_i \sim \mathcal{N}(s_i, \sigma_x^2).$$

We consider here the more general situation where the measurements are unbiased, but correlated so that

$$\mathbf{x} | \mathbf{s} \sim \mathcal{N}(\mathbf{s}, \Sigma_\mathbf{x}). \tag{6.1}$$

The $N \times N$ covariance matrix $\Sigma_\mathbf{x}$ is assumed to have a similar structure as covariance matrix $\Sigma_\mathbf{s}$ (Eq. (2.3)) with constant diagonal entries, σ_x^2, and off-diagonal entries, $\rho_x \sigma_x^2$. The assumption that the measurements follow a Gaussian distribution is almost certainly an oversimplification. However, since this distribution is not well characterized, Gaussianity is a reasonable first guess that allows us to describe responses with a minimal number of parameters, and leads to analytically tractable formulations. This assumption is natural when the uncertainty of the measurement is characterized by the variance of each x_i [53, 87]. The structure in the measurements is characterized by the measurement noise, σ_x, and the covariance matrix, $\Sigma_\mathbf{x}$. The generative model of the task is shown in Figure 6.1.
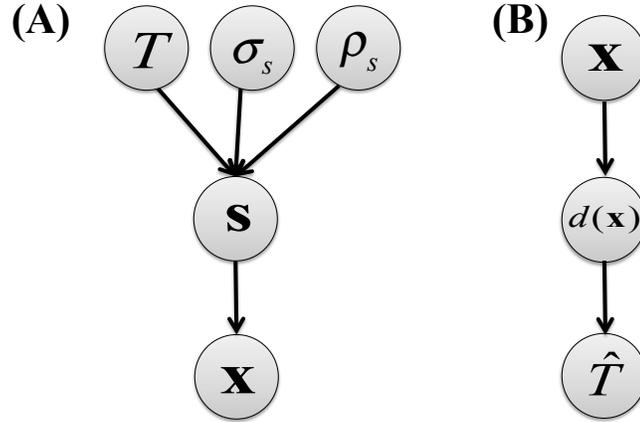
Figure 6.1: **Statistical structure of relevant task variables in the optimal-observer model for a (single) target detection task with stimulus, and measurement noise correlations. (A) Generative model.** The binary variable, T indicates the target presence for $T = 1$, and absence when $T = 0$. The stimulus orientations, $\mathbf{s} = (s_1, s_2, \cdots, s_N)$ are drawn from a multivariate normal distribution with mean vector, $\mathbf{s_D}$, and covariance matrix, $\Sigma_{\mathbf{s}}$. The standard deviation, σ_s, and the correlation coefficient, ρ_s of distractor orientations determine the statistical structure of a visual display. An observer makes measurements, $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ of the presented set of stimuli. These measurements are assumed to be unbiased, and drawn from a multivariate normal distribution having a covariance matrix, $\Sigma_{\mathbf{x}}$. The correlation coefficient, ρ_x determines the strength of correlation between these measurements. **(B) Inference process.** An optimal observer computes a decision variable, $d(\mathbf{x})$ based on the measurements $\mathbf{x}$ to make an estimate, $\hat{T}$ of the true state variable, T. The decision variable $d(\mathbf{x})$ is the log-posterior ratio between the two possibilities of making a response "target-present", or "target-absent" given the measurements, and is given by $\log \frac{p(T=1|\mathbf{x})}{p(T=0|\mathbf{x})}$. The sign of $d(\mathbf{x})$ determines the estimate $\hat{T} = 1$, or $\hat{T} = 0$.
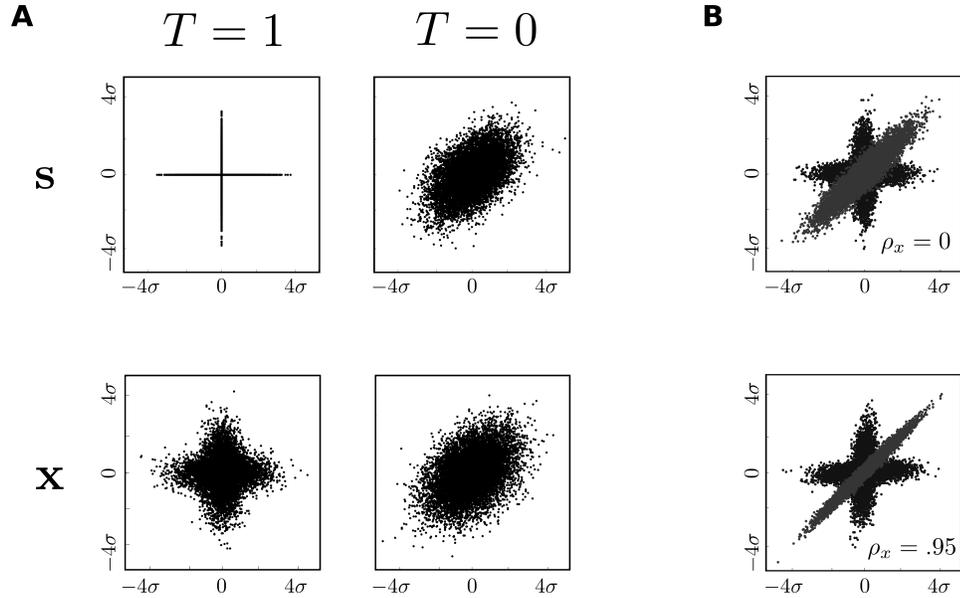
Figure 6.2: **The stimulus and measurement distributions in the single target detection with $N = 2$ stimuli when $\sigma_s^2 = 15°$, and $\sigma_x^2 = 4°$. (A) Stimulus and measurement distributions at $\rho_s = 0.5$.** The stimulus orientations (top) in the case of target present (left), and target absent (right) trials. The corresponding measurement distributions (bottom) in response to the stimulus distributions are shown for measurement correlations, $\rho_x = 0$ in both target present (left) and target absent (right) trials. **(B) Overlap of measurement distributions at $\rho_s = 0.99$.** The target present (black), and absent (dark gray) distributions are shown in the case of $\rho_x = 0$ (top), and $\rho_x = 0.95$ (bottom). The overlap between the two measurement distributions, $\mathbf{x}|T = 1$, and $\mathbf{x}|T = 0$ reduces as the strength of measurement correlation increases, and hence the two distributions are distinguishable. Throughout the chapter, the axes are measured in terms of the standard deviation, σ, which is defined by $\sigma^2 = \sigma_s^2 + \sigma_x^2$.

We now derive the mathematical model that an optimal observer follows to make decisions on the task. In particular, we compute the Bayesian decision variable as a function of different statistical parameters that govern the structure in external scenes, and the observer's measurements.

6.2 Optimal observer theory

In order to understand how noise correlations in the measurements impact the decisions of an ideal Bayesian observer, we derive an analytical expression for the decision variable based on a similar mechanism discussed in Section 2.2.

An optimal Bayesian observer makes a decision based on the log posterior ratio given in Eq. (2.9). The observer infers target presence when the decision variable, denoted by $d_{\text{NST}}(\mathbf{x})$, is positive, and target absence otherwise. That is,

$$d_{\text{NST}}(\mathbf{x}) = \log \frac{P(T=1|\mathbf{x})}{P(T=0|\mathbf{x})} = \underbrace{\log \frac{P(\mathbf{x}|T=1)}{P(\mathbf{x}|T=0)}}_{L_{\text{NST}}(\mathbf{x})} + \log \frac{P(T=1)}{P(T=0)} > 0. \qquad (6.2)$$

6.2.1 The log-likelihood ratio

We compute the log-likelihood ratio, $L_{\text{NST}}(\mathbf{x})$ by marginalizing the information over intermediate variables; the spatial location, and the stimulus orientations, $\mathbf{s}$. We readily observe that it is identical to obtaining Eq. (2.11) in Section 2.2.

Therefore,

$$L_{\text{NST}}(\mathbf{x}) = \log \frac{P(\mathbf{x}|T=1)}{P(\mathbf{x}|T=0)} = \log \frac{1}{N} \sum_{j=1}^{N} \frac{\int P(\mathbf{x}|\mathbf{s})P(\mathbf{s}|T=1)d\mathbf{s}}{\int P(\mathbf{x}|\mathbf{s})P(\mathbf{s}|T=0)d\mathbf{s}}. \tag{6.3}$$

However, we note that the computation of $P(\mathbf{x}|T=1)$ is not simple, and straight in this case as compared to the one in Section 2.2. It is because $P(\mathbf{x}|\mathbf{s})$ can no longer be written as a product of one-dimensional normal density functions. Therefore, to evaluate the integral in case of $T=1$, we construct a new $N \times N$ matrix $\mathbf{\Sigma}_{\mathbf{s},j}^{\eta}$ for $\eta > 0$ as:

$$(\mathbf{\Sigma}_{\mathbf{s},j}^{\eta})_{k,l} = \begin{cases} \sigma_s^2, & \text{if } k = l \neq j, \\ \rho_s \sigma_s^2, & \text{if } k, l \neq j, \\ \eta, & \text{if } k = l = j, \\ 0, & \text{if } k = j, l \neq j \text{ or } k \neq j, l = j. \end{cases} \tag{6.4}$$

Here $j \in \{1, 2, \cdots, N\}$ represents the spatial location of a target. By denoting $\tilde{\mathbf{s}_{\mathbf{D}}}$ as an N-dimensional vector with all components s_D, but the jth as s_T, we can write

$$\mathbf{s}|(T=1, j) \sim \lim_{\eta \to 0^+} \mathcal{N}(\tilde{\mathbf{s}_{\mathbf{D}}}, \mathbf{\Sigma}_{\mathbf{s},j}^{\eta}). \tag{6.5}$$

We introduced the auxiliary covariance $\mathbf{\Sigma}_{\mathbf{s},j}^{\eta}$ because $\mathbf{\Sigma}_{\mathbf{s},j}^{0}$ is not invertible, since the density is singular in the variable with index j.

In the limit of $\eta \to 0^+$, we compute the log-likelihood ratio

$$L_{\text{NST}}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \frac{\displaystyle\int P(\mathbf{x}|\mathbf{s})P(\mathbf{s}|T=1)d\mathbf{s}}{\displaystyle\int P(\mathbf{x}|\mathbf{s})P(\mathbf{s}|T=0)d\mathbf{s}}$$

$$= \log \frac{1}{N} \sum_{j=1}^{N} \frac{\displaystyle\lim_{\eta \to 0^+} \int f(\mathbf{x}; \mathbf{s}, \Sigma_\mathbf{x})f(\mathbf{s}; \tilde{\mathbf{s}}_\mathbf{D}, \mathbf{\Sigma}_{\mathbf{s},j}^{\eta})d\mathbf{s}}{\displaystyle\int f(\mathbf{x}; \mathbf{s}, \Sigma_\mathbf{x})f(\mathbf{s}; \mathbf{s}_\mathbf{D}, \Sigma_\mathbf{s})d\mathbf{s}}.$$

We use Eqs. (B.3), and (B.4) about product, and integral properties of multivariate normal distributions, and denote

$$\mathbf{C} = \Sigma_\mathbf{s} + \Sigma_\mathbf{x}, \text{ and } \mathbf{C_j} = \mathbf{\Sigma}_{\mathbf{s},j}^{0} + \Sigma_\mathbf{x}, \tag{6.6}$$

to obtain

$$L_{\text{NST}}(\mathbf{x}) = \log \frac{1}{N} \sum_{j=1}^{N} \frac{f(\mathbf{x}; \tilde{\mathbf{s}}_\mathbf{D}, \mathbf{C_j})}{f(\mathbf{x}; \mathbf{s}_\mathbf{D}, \mathbf{C})}$$

$$= \log \frac{1}{N} \sqrt{\frac{|\mathbf{C}|}{|\mathbf{C_j}|}} \sum_{j=1}^{N} \exp\left(-\frac{1}{2}(\mathbf{x} - \tilde{\mathbf{s}}_\mathbf{D})^{\mathsf{T}} \mathbf{C_j}^{-1}(\mathbf{x} - \tilde{\mathbf{s}}_\mathbf{D}) + \frac{1}{2}(\mathbf{x} - \mathbf{s}_\mathbf{D})^{\mathsf{T}} \mathbf{C}^{-1}(\mathbf{x} - \mathbf{s}_\mathbf{D})\right).$$

$$\tag{6.7}$$

The determinant of the matrix $\mathbf{C_j}$ does not depend on the spatial location parameter j since all matrices of type $\mathbf{C_j}$ can be obtained from each other by permuting appropriate rows, and columns. Moreover, to describe the distributions of the Gaussian variables $\mathbf{x}|(T=0)$, and $\mathbf{x}|(T=1)$, we note that their respective covariances are determined by the matrices $\mathbf{C}$, and $\mathbf{C_j}$.

6.2.1.1 Determinants and inverses of matrices $\mathbf{C}$ and $\mathbf{C_j}$

We further simplify Eq. (6.7) by computing the determinants, and inverses of the covariance matrices $\mathbf{C}$, and $\mathbf{C_j}$ when $\rho_s, \rho_x \neq 1$, otherwise these matrices are singular. We first decompose these matrices as:

$$\mathbf{C} = \mathbf{D} + (\rho_s \sigma_s^2 + \rho_x \sigma_x^2)\mathbf{J}, \text{ and } \mathbf{C_j} = \mathbf{A_j} + \mathbf{U_j} \mathbf{E} \mathbf{V_j}$$

where $\mathbf{D}$ is a diagonal matrix with constant diagonal entries, $\sigma_s^2(1 - \rho_s) + \sigma_x^2(1 - \rho_x)$, $\mathbf{J}$ is a matrix of ones, $\mathbf{E}$ is a 2×2 identity matrix,

$$\text{and } (\mathbf{A_j})_{k,l} = \begin{cases} \eta + \sigma_x^2(1 - \rho_x), & \text{if } k = l = j, \\ \sigma_s^2(1 - \rho_s) + \sigma_x^2(1 - \rho_x), & \text{if } k = l \neq j, \\ 0, & \text{otherwise .} \end{cases}$$

Furthermore, the columns of $N \times 2$ matrix $\mathbf{U_j}$, and rows of $2 \times N$ matrix $\mathbf{V_j}$ are given by

$$\mathbf{U_j}_{(k,1)} = \begin{cases} \sigma_s^2(1 - \rho_s) + \sigma_x^2(1 - \rho_x), & \text{if } k \neq j, \\ \sigma_x^2(1 - \rho_x), & \text{if } k = j, \end{cases} \text{ and } \mathbf{U_j}_{(k,2)} = \sigma_x^2(1 - \rho_x) \; \forall \, k,$$

$$\mathbf{V_j}_{(1,k)} = \begin{cases} 1, & \text{if } k \neq j, \\ 0, & \text{if } k = j, \end{cases} \text{ and } \mathbf{V_j}_{(2,k)} = \begin{cases} 0, & \text{if } k \neq j, \\ 1, & \text{if } k = j. \end{cases}$$

Additionally, we define the following variables

$$v = \frac{1}{\sigma_s^2(1 - \rho_s) + \sigma_x^2(1 - \rho_x)}, \quad \tilde{v} = \frac{1}{\sigma_x^2(1 - \rho_x)}, \quad a = \rho_s \sigma_s^2 + \rho_x \sigma_x^2, \tag{6.8}$$

$$V = Nv, \quad V_{\backslash j} = (N - 1)v, \quad \beta = \frac{a}{1 + aV}, \quad q = a + \rho_s \sigma_s^2 \rho_x \sigma_x^2 \tilde{v}, \tag{6.9}$$

$$r = 1 + \rho_s \sigma_s^2 V_{\backslash j}, \text{ and } \gamma = 1 + aV_{\backslash j} + \rho_x \sigma_x^2 \tilde{v} r. \tag{6.10}$$

We apply the generalized Matrix Determinant Lemma (see Lemma 2 in Appendix B.3) to obtain the following determinants:

$$|\mathbf{C}| = \frac{1 + aV}{v^N}, \text{ and } |\mathbf{C_j}| = \frac{\gamma}{v^{N-1}\,\tilde{v}}.$$

We also apply the Woodbury matrix identity described in Theorem 2 in Appendix B.3 to compute the following inverses of both covariance matrices

$$(\mathbf{C}^{-1})_{k,l} = \begin{cases} v - \beta v^2 & \text{if } k = l, \\ -\beta v^2 & \text{if } k \neq l, \end{cases}$$

$$\text{and } (\mathbf{C_j}^{-1})_{k,l} = \begin{cases} \tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 r}{\gamma} & \text{if } k = l = j, \\ v - \frac{v^2 q}{\gamma} & \text{if } k = l \neq j, \\ -\frac{v^2 q}{\gamma} & \text{if } k \neq l, \text{ and } k, l \neq j, \\ -\frac{\rho_x \sigma_x^2 v \tilde{v}}{\gamma} & \text{if } k = j, \, l \neq j, \text{ or } l = j, \, k \neq j. \end{cases}$$

We now substitute the above obtained determinants, and inverses of the covariance matrices to compute the following expressions:

$$(\mathbf{x} - \mathbf{s_D})^{\mathsf{T}} \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s_D}) = (v - \beta v^2) \sum_{k=1}^{N} (x_k - s_D)^2 - \beta v^2 \sum_{k \neq l}^{N} (x_k - s_D)(x_l - s_D)$$

$$(\mathbf{x} - \mathbf{s_{\tilde{D}}})^{\mathsf{T}} \mathbf{C_j}^{-1} (\mathbf{x} - \mathbf{s_{\tilde{D}}}) = \left(\tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 r}{\gamma} \right) (x_j - s_T)^2 + v \sum_{k \neq j}^{N} (x_k - s_D)^2$$

$$- \frac{2\rho_x \sigma_x^2 v \tilde{v}}{\gamma} (x_j - s_T) \sum_{k \neq j}^{N} (x_k - s_D) - \frac{v^2 q}{\gamma} \sum_{k, l \neq j}^{N} (x_k - s_D)(x_l - s_D).$$

Substituting the above obtained expressions in Eq. (6.7) gives us the following simplified equation for the log-likelihood ratio of a single target detection task

with correlated measurements:

$$
\begin{aligned}
L_{\text{NST}} = \log \frac{1}{N} & \sqrt{\frac{\tilde{v}(1+aV)}{v\,\gamma}} \sum_{j=1}^{N} \exp\left(-\frac{1}{2} \left(\tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 r}{\gamma} \right) (x_j - s_T)^2 \right. \\
& - \frac{1}{2} \left(\beta v^2 - v \right) (x_j - s_D)^2 - \left(\beta v^2 - \frac{\rho_x \sigma_x^2 v \tilde{v}}{\gamma} \right) (x_j - s_D) \sum_{k \neq j}^{N} (x_k - s_D) \\
& \left. - \frac{1}{2} \left(\beta v^2 - \frac{v^2 q}{\gamma} \right) \sum_{k,l \neq j}^{N} (x_k - s_D)(x_l - s_D) \right).
\end{aligned}
\tag{6.11}
$$

6.2.2 Bayesian decision variable

The decision variable $d_{\text{NST}}(\mathbf{x})$ defined in Eq. (6.2) completely characterizes the decision-making strategy of an ideal Bayesian observer on a single trial. It also quantifies the impact of different (stimulus, and measurement structure) parameters on the decisions of the optimal observer.

Since a target is present with a $1/2$ probability on each trial of the task, the ideal observer uses a uniform prior on T. Thus, the decision variable to report "target present" on the task is given by the log-likelihood ratio obtained in Eq. (6.11)

$$d_{\text{NST}} = \log \frac{1}{N} \sqrt{\frac{\tilde{v}(1 + aV)}{v\,\gamma}} \sum_{j=1}^{N} \exp\left(\underbrace{-\frac{1}{2}\left(\tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 r}{\gamma}\right)(x_j - s_T)^2}_{I} \right.$$

$$\underbrace{-\frac{1}{2}\left(\beta v^2 - v\right)(x_j - s_D)^2}_{II} - \underbrace{\left(\beta v^2 - \frac{\rho_x \sigma_x^2 v\tilde{v}}{\gamma}\right)(x_j - s_D)\sum_{k \neq j}^{N}(x_k - s_D)}_{III}$$

$$\left. \underbrace{-\frac{1}{2}\left(\beta v^2 - \frac{v^2 q}{\gamma}\right)\sum_{k,l \neq j}^{N}(x_k - s_D)(x_l - s_D)}_{IV} \right). \tag{6.12}$$

The above equation gives the decision variable in terms of the model parameters, and the measurements, $\mathbf{x}$. While the total number of stimuli N, the variability, and co-variability between the distractors orientation determined respectively by σ_s^2, and ρ_s, regulate the structure of visual stimuli. The variability σ_x^2, and correlation ρ_x describe the structure of the observer's measurements.

6.2.3 Interpretation of the decision variable

Eq. (6.12) defines a nonlinear decision boundary in the space of measurements $\mathbf{x}$. Although the expression is explicit, the decision variable depends in a complicated way on the different parameters that describe the structure of the stimulus, and the response. The variables v, and $\tilde{v}$ represent scaled inverse variances corresponding to distractor, and target stimulus, while V, and $V_{\backslash j}$ denote the population sums of v in target present, and absent cases, respectively. The parameters

β, r, q, and γ (Eq. (6.8)) are defined in terms of $\sigma_s^2, \rho_s, \sigma_x^2, \rho_x$, and it is difficult to quantify their explicit impact on decisions, and performance of an observer.

However, we note that the formulation of Eq. (6.12) is similar to the decision variable derived in Eq. (2.21) except for different scaling parameters. Thus, Eq. (6.12) can be interpreted in a similar way as in Section 2.2.2. Specifically, we interpret each term in the exponent of Eq. (6.12) as an evidence towards the jth stimulus being a target: (I) if the jth measurement is close to the target orientation, this term increases the likelihood of the jth stimulus being a target; (II) on the other hand, if the jth measurement is similar to the mean distractor orientation, this term decreases such a likelihood; (III) the third term compares the sample mean of distractor measurements, with the difference of jth measurement from the mean distractor orientation; if the term is large, it is less likely that the jth stimulus is the target; (IV) the fourth term can be rewritten as the sample covariance of distractor measurements; if this term is large it is more likely that the jth stimulus is the target. Thus, each term is relevant in making a decision whether the jth measurement corresponds to a target stimulus or not. The coefficient of each exponent term carries the information about parameters of stimulus, and measurement structure. The influence of these parameters such as $\sigma_s^2, \rho_s, \sigma_x^2, \rho_x$ on different terms of the decision variable $d_{\text{NST}}(\mathbf{x})$ is difficult to understand since their dependence is much more intricate.

We next consider a simple, and special case of the decision variable in Eq. (6.12). Specifically, we evaluate the equation in the absence of correlations between measurements.

6.2.4 No noise correlations, $\rho_x = 0$

We note that in the case of no noise correlations in the measurements, *i.e.,* when $\rho_x = 0$, the mathematical model of the task reduces to the one described in Section 2.1. Thus, on substituting $\rho_x = 0$ in Eq. (6.12), the variables defined in Eq. (6.8) reduces to the ones earlier defined in Chapter 2 (Eqs. (2.15) and (2.19))

$$a = \rho_s \sigma_s^2, \ v = \tilde{w}, \ \tilde{v} = w, \ V = \tilde{W}, \beta = \alpha, \ \frac{q}{\gamma} = \alpha_{\setminus j}.$$

Subsequently, the decision variable $d_{\mathrm{NST}}(\mathbf{x})$ given by Eq. (6.12) reduces to the variable $d_{\mathrm{ST}}(\mathbf{x})$ derived in Eq. (2.21). Thus, the target detection task discussed in Chapter 2 is a special case of the one here with a generalized assumption on the measurement noise.

We now consider how variations in the external structure parameters together with the parameters governing the structure in the measurements impact the performance of an ideal Bayesian observer. External structure is characterized by the number of stimuli, and the pairwise correlations between them, while the structure of the measurements is characterized by the covariance matrix $\Sigma_{\mathbf{x}}$. In general, we cannot assume that these parameters can be varied independently. For example, neural mechanisms that impact σ_x^2 almost certainly impact ρ_x [28, 33, 117]. However, the dependence between these parameters is not fully characterized, and we therefore explore a range of possible parameter values below.

6.3 Analysis and Results

Our goal is to describe how stimulus structure, along with the structure of the corresponding measurements affect the decisions of an optimal observer in a target detection task. We examine how performance changes as both correlations between distractors, and between measurements are varied.

We first note that the stimulus variable **s**, follows two different distributions depending on whether $T = 0$, or $T = 1$. In general, measurement noise increases the overlap between these distributions. The higher the overlap between the two distributions, the more difficult the decision. However, structured noise in the measurements can reduce such overlap. Therefore, performance of an ideal observer depends not only on the level, but also on the structure of measurement noise [6, 4].

We examine Eq. (6.12) numerically in different regimes of stimulus, and measurement statistical parameters. Specifically, we consider the following two regimes determined by the dominance of either external noise, σ_s^2, or measurement noise, σ_x^2:

(a) weak measurement noise, $\sigma_x^2 \ll \sigma_s^2$,

(b) strong measurement noise, $\sigma_x^2 = \sigma_s^2$.

In the following sections, we elaborate how noise correlations in the measurements affect the decision variable $d_{\mathrm{NST}}(\mathbf{x})$ (Eq. (6.12)), and subsequently the performance of an ideal Bayesian observer. We obtain analytical expressions for the

decision variable in these cases, and provide supporting numerical results along with the best possible intuitive interpretation.

6.3.1   Weak measurement noise, $\sigma_x^2 \ll \sigma_s^2$

The external structure in a scene is determined by the noise level σ_s, and also by the pairwise correlation coefficient ρ_s. For a fixed noise level σ_s, the structure is introduced in the visual stimuli by varying the amount of correlations among pairs of distractor orientations. A strong external structure is introduced when $\rho_s = 1$ as all distractors are identical, and the target is an odd-ball stimulus. While in the cases of $\rho_s < 1$, the external structure is weaker, and hence detecting a target may not necessarily be easier. Therefore, we individually treat the cases of weak, and strong external structure in the regime of weak noise in the observer's measurement.

We denote $\epsilon = \frac{\sigma_x^2}{\sigma_s^2}$, and expand different terms of Eq. (6.12) about $\epsilon \ll 1$. We only consider terms with larger contribution, *i.e.,* terms of the orders of $\mathcal{O}(\frac{1}{\epsilon})$, to obtain approximations for the decision variable $d_{\text{NST}}(\mathbf{x})$ in different regimes of parameters.

6.3.1.1 Strong external structure, $\rho_s = 1$

In the case of $\rho_s = 1$, and $\epsilon \ll 1$, the coefficients of the different exponent terms in Eq. (6.12) reduces to the following simplified expressions:

$$I : \tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 r}{\gamma} \approx \frac{1}{\sigma_x^2}$$

$$II : \beta v^2 - v \approx \frac{-(N-1)}{N\sigma_x^2(1-\rho_x)}$$

$$III : \beta v^2 - \frac{\rho_x \sigma_x^2 v \tilde{v}}{\gamma} \approx \frac{1}{N\sigma_x^2(1-\rho_x)} + \mathcal{O}(1)$$

$$IV : \beta v^2 - \frac{v^2 q}{\gamma} \approx -\frac{1}{N(N-1)\sigma_x^2(1-\rho_x)},$$

and the leading determinant coefficient becomes

$$\sqrt{\frac{\tilde{v}(1+aV)}{v\,\gamma}} \approx \sqrt{\frac{N(1-\rho_x)}{N-1}}.$$

Therefore, we obtain the following approximation of the Bayesian decision variable $d_{\mathrm{NST}}(\mathbf{x})$ in this case

$$d_{\mathrm{NST}}(\mathbf{x}) \approx \log \frac{1}{N} \sqrt{\frac{N(1-\rho_x)}{N-1}} \sum_{j=1}^{N} \exp\left(-\frac{1}{2N\sigma_x^2(1-\rho_x)} \left((1-N\rho_x)(x_j - s_T)^2 \right.\right.$$

$$\left.\left. -(N-1)(x_j - s_D)^2 + 2(x_j - s_D)\sum_{k \neq j}^{N}(x_k - s_D) - \frac{1}{N-1}\sum_{k,l \neq j}^{N}(x_k - s_D)(x_l - s_D) \right) \right).$$

$$(6.13)$$

Furthermore, in the particular case of $s_T = s_D = 0$, the above equation reduces to a much simpler expression that is easier to interpret:

$$d_{\mathrm{NST}}(\mathbf{x}) \approx \log \frac{1}{N} \sqrt{\frac{N(1 - \rho_x)}{N - 1}}$$

$$\sum_{j=1}^{N} \exp \underbrace{\left(-\frac{1}{2N\sigma_x^2(1 - \rho_x)} \left((1 - N\rho_x)x_j^2 + 2x_j \sum_{k \neq j}^{N} x_k - \frac{1}{N - 1} \left(\sum_{k \neq j}^{N} x_k \right)^2 \right) \right)}_{E_j}.$$

$$(6.14)$$

Clearly, the above expression indicates that an ideal observer uses the strength of measurement correlations ρ_x in a decision. Specifically, in the limit of perfect noise correlations, *i.e.*, $\rho_x \approx 1$, the exponential term in Eq. (6.14) is approximately

$$E_j \approx \left(\frac{N - 1}{2N\sigma_x^2(1 - \rho_x)} \left(x_j - \frac{1}{N - 1} \sum_{k \neq j} x_k \right)^2 \right). \qquad (6.15)$$

Therefore, the optimal observer simply subtracts the mean of the $N - 1$ measurements of putative distractors from that of the putative target. In the case of perfect noise correlations, the measurements of the distractor stimuli are identical. Thus, on "target absent" trials, the term E_j's in Eq. (6.15) are zero, and therefore, $d_{\mathrm{NST}}(\mathbf{x}) \to -\infty$. While on a target present trial, E_j is positive, and the exponential diverges to infinity as $\rho_x \to 1$. The prefactor in Eq. (6.14) approaches zero, however, the divergence of exponential is stronger. Therefore, $d_{\mathrm{NST}}(\mathbf{x}) \to \infty$, and an ideal observer performs perfectly in this case.

We confirm the above analytical observation through numerical simulations in the case of $N = 4$ stimuli. We vary the strength of stimulus, and measurement

noise correlations (ρ_s, and ρ_x, respectively), and examine the trend in the performance of an optimal Bayesian observer shown in Figure 6.3(A). For simulation purposes, we used $\sigma_s = 15°$, and $\sigma_x = 4°$ to obtain the weak measurement noise ($\sigma_x^2 \ll \sigma_s^2$) regime.

We observe in Figure. 6.3(A) that the performance of an ideal observer in weak measurement noise case is nearly independent of ρ_s, and ρ_x for weak external structure ($\rho_s < 1$). However, it depends strongly on ρ_x when the external structure is relatively stronger, *i.e.*, $\rho_s \approx 1$. In particular, any amount of noise correlations strongly drives the performance in such a case (seen in Figure. 6.3(A), and (C)). Moreover, we note that perfect performance is observed at $\rho_s = \rho_x = 1$ (Figure 6.3(C)). Hence, in the presence of strong external structure (all distractors having identical orientations), strong noise correlations enhance the performance of an optimal Bayesian observer.

The increased performance with increasing external correlations ρ_s, accords with intuition that strong external structure makes detecting a target easier. However, *measurement* structure can play an equally important role. In the presence of strong external structure (homogeneous distractors), noise correlations can significantly improve performance as seen in Figure 6.3(B).

Perfect performance is achieved at $\rho_s = \rho_x = 1$, and we can understand it intuitively. In this case measurements x_i, of the stimuli are obtained by adding the *same* realization of a random variable (noise) to each stimulus characteristic, s_i. In target absent trials, all measurements are hence identical. If the target is present, the measurements contain a single outlier, and an ideal observer can thus
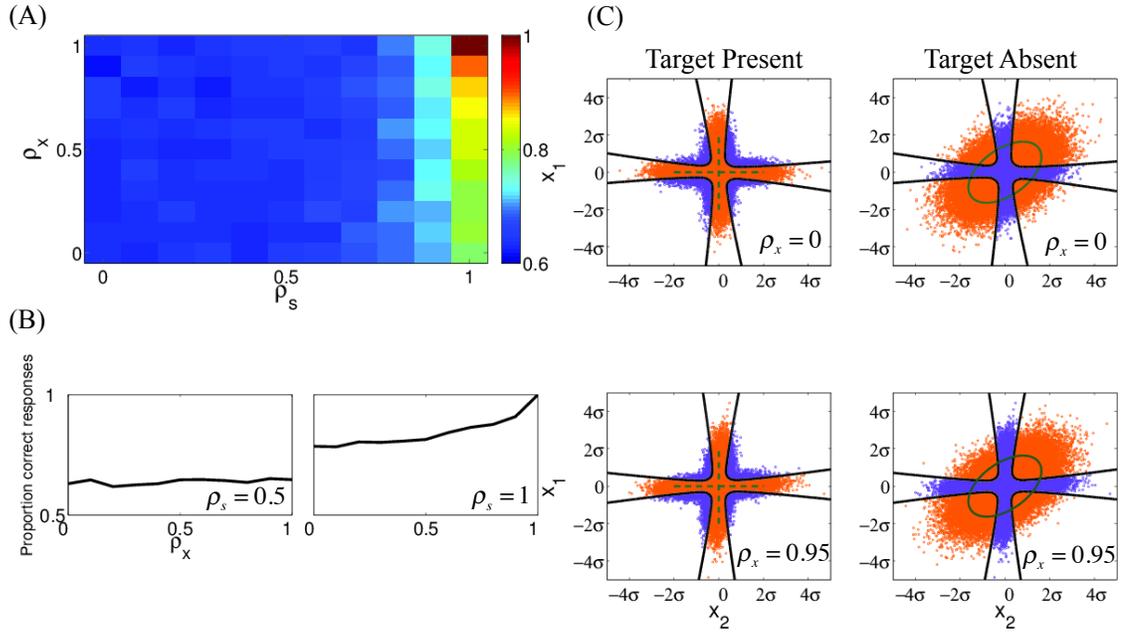
Figure 6.3: **Impact of measurement noise correlations on the performance of an ideal Bayesian observer on the single target detection task when the measurement noise is weak as compared to the external noise,** $\sigma_x^2 \ll \sigma_s^2$ ($\sigma_s = 15°, \sigma_x = 4°$). **(A)** Performance of an ideal observer as a function of external correlation strength, ρ_s, and measurement noise correlations, ρ_x on the task with $N = 4$ stimuli. **(B)** The proportion of correct responses as a function of ρ_x in the case of weak external structure, $\rho_s = 0.5$ (left), and strong external structure, $\rho_s = 1$ (right) when $N = 4$. **(C)** The decision boundary, $d_{\mathrm{NST}}(\mathbf{x}) = 0$, with target present (left), and target absent (right) distributions in the case of $\rho_s = 0.5$, and $N = 2$ stimuli. Here purple dots correspond to incorrect inferences while orange dots represents correct responses. The green lines (target present) or ellipses (target absent cases) show 2 units of standard deviation for the stimulus distribution.

distinguish the two cases perfectly.

6.3.1.2 Weak external structure, $\rho_s < 1$

We also observe that noise correlations have little effect on performance when external structure is weaker, *i.e.* when $\rho_s < 1$ (Figure. 6.3(A)). To intuitively understand such a behavior, we again consider the case of $\rho_x \approx 1$ so that measurements are obtained by equal rotation of each stimulus orientation. An ideal observer can use the structure of measurement noise in making a decision. However, the observer cannot use knowledge about a *particular realization* of noise. In other words, the observer can use the fact that the measurements are obtained by rotating the stimulus by approximately the same angle, but not the exact angle of the rotation. If there is only weak external structure, it is now difficult to tell whether one of the stimuli is an outlier. An ideal observer must therefore infer whether a target is present from the individual measurements of stimuli. Hence measurement noise correlations provide little help in the absence of external structure.

These observations are also reflected in the structure of the decision boundaries ($d_{\text{NST}}(\mathbf{x}) = 0$), and distributions of the measurements illustrated in Figure 6.3(C). In the target absent (left column, Figure 6.3(C)), and present trials (right column) the distribution of measurements is determined predominantly by variability in the stimulus. Noise correlations present in the observer's measurements have little effect on these distributions. As a result, the decision boundary also changes little with an increase in ρ_x. This is in contrast to the case when $\rho_s \approx 1$, where

internal variability can be important in increasing the overlap between the distributions of $\mathbf{x}|T = 1$ and $\mathbf{x}|T = 0$ (see Figure 6.4).

We confirm this intuition about the role of noise correlations by approximating the decision variable in this case, and expanding the coefficients of exponent terms in Eq. (6.12) as follows

$$I : \tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 r}{\gamma} \approx \frac{1}{\sigma_x^2}$$

$$II : v - \beta v^2 \approx \frac{1 + (N-2)\rho_s}{\sigma_s^2(1 - \rho_s)(1 + (N-1)\rho_s)} = \mathcal{O}(1)$$

$$III : \beta v^2 - \frac{\rho_x \sigma_x^2 v \tilde{v}}{\gamma} \approx \mathcal{O}(1)$$

$$IV : \beta v^2 - \frac{v^2 q}{\gamma} \approx \mathcal{O}(1).$$

We further approximate the leading determinant coefficient in Eq. (6.12) as

$$\sqrt{\frac{\tilde{v}(1 + aV)}{v \gamma}} \approx \sqrt{\frac{\sigma_s^2(1 - \rho_s)(1 + (N-1)\rho_s)}{\sigma_x^2(1 + (N-2)\rho_s)}}.$$

Combining the above terms gives us the following approximation of the decision variable when $\sigma_x^2 \ll \sigma_s^2$, and $\rho_s < 1$,

$$d_{\text{NST}}(\mathbf{x}) \approx \log \frac{1}{N} \sqrt{\frac{\sigma_s^2(1 - \rho_s)(1 + (N-1)\rho_s)}{\sigma_x^2(1 + (N-2)\rho_s)}} \sum_{j=1}^{N} \exp\left(-\frac{(x_j - s_T)^2}{2\sigma_x^2}\right). \quad (6.16)$$

This approximation is relatively simpler, and easier to understand. We easily see that the strength of noise correlations ρ_x does not affect the decisions of an ideal observer at highest order in (σ_x^2/σ_s^2). Therefore noise correlations only weakly impact the decision, and hence performance. Additionally, the terms in the exponent are $x_j^2/2\sigma_x^2$. Hence, the orientation at each location is considered separately, and weighted by the precision of measurements, $1/\sigma_x^2$. The ideal observer hence
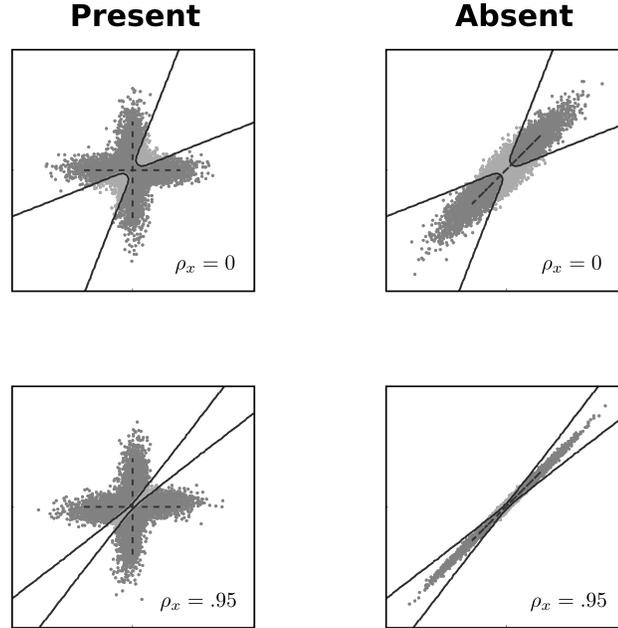
Figure 6.4: **The stimulus and response distributions along with the decision boundary on the single target detection task in presence of weak measurement noise ($\sigma_x^2 \ll \sigma_s^2$) and strong external structure, $\rho_s = 1$ as a function of noise correlations in the measurements.** The distributions of the stimuli and corresponding responses on the single target detection task with $N = 4$ stimuli, weak measurement noise ($\sigma_s = 15°, \sigma_x = 4°$), and strong external structure ($\rho_s = 1$) for weak (top panels) and strong noise (bottom panels) correlation. Black solid lines represent the decision boundary, $d_{\text{NST}}(\mathbf{x}) = 0$. The dashed black lines correspond to the 2 units standard deviation about the stimulus distribution, $\mathbf{s}|T = 1$ (left), and $\mathbf{s}|T = 0$ (right). The distribution of the responses is represented by the dots. Light grey dots correspond to the incorrect inferences while dark grey dots to the correct responses in both target present (left) and absent (right) cases. We note that noise correlations decrease the overlap between the measurement distributions, $\mathbf{x}|T = 1$, and $\mathbf{x}|T = 0$, and makes them distinguishable.

makes a decision primarily based on the evidence from each stimulus separately in this case.

In sum, external structure of a scene implies that stimulus distributions $\mathbf{s}|T = 1$, and $\mathbf{s}|T = 0$ are concentrated on low dimensional subspaces. Even small noise can increase the overlap between the distributions of measurements $\mathbf{x}|T = 1$, and $\mathbf{x}|T = 0$. However, noise in the measurement, and external structure can conspire to decrease the overlap between the distribution of measurements. This happened in the present case when ρ_s, and ρ_x are both close to 1 (Figure 6.3(B)).

In the absence of external structure, the signal distributions $\mathbf{s}|T = 1$, and $\mathbf{s}|T = 0$ are not concentrated along low dimensional manifolds. Here external variability σ_s^2 always dominates, and low intensity measurement variability σ_x^2 has little effect on the performance of an ideal observer.

6.3.2 Strong measurement noise, $\sigma_x^2 = \sigma_s^2$

With a single target, noise correlations have little impact on the performance of an ideal observer, unless external structure is strong. Although, external variability is typically expected to be much stronger than measurement variability in most situations [14], we next consider the case when the two sources of variability are comparable, $\sigma_x^2 = \sigma_s^2$. The case of $\sigma_x^2 \gg \sigma_s^2$ is insignificant to consider since the increased measurement noise leads to poor performance, and the observer essentially makes a guess on each trial.

Increasing measurement noise σ_x^2 trivially leads to a decrease in performance.

However, in the limit of perfect external, and measurement noise correlations *i.e.*, $\rho_s, \rho_x \approx 1$, an ideal observer still performs perfectly. It is not clear, however, how the performance is affected at intermediate values of correlations, and noise. We thus explore this case numerically, and provide an intuitive reasoning for the observed behavior. However, an analytical approximation of the Eq. (6.12) is difficult to obtain in this particular case. For simulation purposes, we consider the case of $N = 4$ stimuli, and $\sigma_s = \sigma_x = 15°$. Figure 6.5(A) shows the behavior in the performance of an optimal Bayesian observer with varying amounts of external (ρ_s), and measurement noise correlations (ρ_x).

On comparing Figure 6.5(A),(C) with Figure 6.3(A),(C), we observe that noise correlations have a completely different effect on performance than in the case of weak measurement noise. Even with no external structure, $\rho_s = 0$, performance increases slightly (approximately 5-6%) with an increase in ρ_x. Surprisingly, for intermediate values of external correlations, $\rho_s = 0.5$, noise correlations have a negative impact on performance. The reason for this unexpected behavior is not clear, as Eq. (6.12) is difficult to analyze in this case. However it is obvious that structure present in the observer's measurements play a much greater role in decision making when noise is very large, unlike the weak noise case where we see little change in structure when ρ_x is varied (Figure 6.3(A)).

We see this fact reflected in the distributions of the responses, and the decision boundary in Figure 6.5(B), which are both strongly affected by noise correlations ρ_x. Moreover, we see a similarity between this decision boundary, and the one shown in Figure 6.3(B) in that the mid-section is elongated along the diagonal to
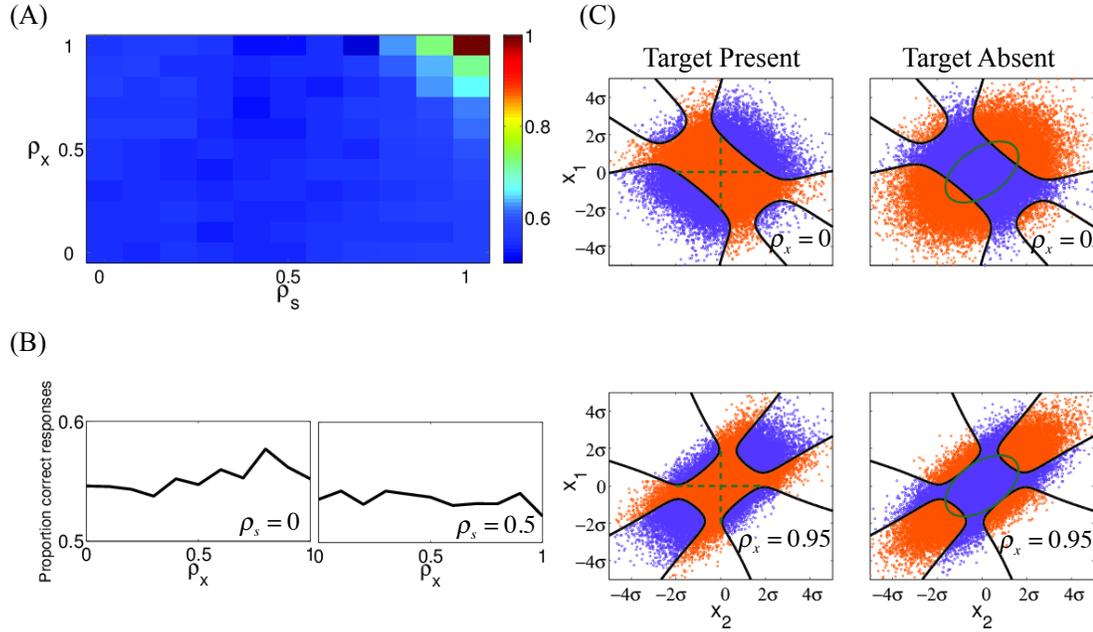
Figure 6.5: **Impact of measurement noise correlations on the performance of an ideal Bayesian observer on the single target detection task when the measurement noise is comparable to the external noise, $\sigma_x^2 = \sigma_s^2$ ($\sigma_s = \sigma_x = 15°$). (A)** Performance of an ideal observer as a function of external correlation strength, ρ_s, and measurement noise correlations, ρ_x on the task with $N = 4$ stimuli. **(B)** The proportion of correct responses as a function of ρ_x in the case of $\rho_s = 0$ (left), and $\rho_s = 0.5$ (right) when $N = 4$. **(C)** The decision boundary, $d_{\mathrm{NST}}(\mathbf{x}) = 0$, with target present (left), and target absent (right) distributions in the case of $\rho_s = 0.5$, and $N = 2$ stimuli. Again, purple dots correspond to the incorrect inferences while orange dots represents the correct responses. The green lines or ellipses show 2 units of standard deviation for the stimulus distribution.

capture more of the response distribution with an increase in noise correlations. This transformation makes intuitive sense considering the fact that with increased noise correlations, noise simply moves the stimulus elements roughly along the diagonal, and therefore the decision boundary changes accordingly.

6.4 Summary

The performance of an ideal observer on a detection task with a single target is greatly influenced by the interplay between the statistical structure of the stimuli, and the correlations between the observer's measurements. The decisions are largely affected by how different parameters that control the stimulus, and measurements structure are varied. Specifically, we find that the performance of the observer is mostly unaffected at all levels of measurements correlations when the stimulus structure is not strong, and the measurement noise is weaker than the external noise. However, the observer always make correct decisions when distractors are homogeneous ($\rho_s = 1$), and measurements are perfectly correlated. This is because identical distractors induces a strong statistical structure in a scene, and strongly correlated measurements preserves this structure, and help in making correct decisions. In the case of strong measurement noise which is comparable to the external noise ($\sigma_x^2 = \sigma_s^2$), the change in performance for weak external structure is little unpredictable, and largely depends on the interaction between the stimulus, and measurement noise correlations. While, the observer performs perfectly when both distractors, and measurements are maximally correlated.

Chapter 7

Noise correlations in a multiple target detection task

In a detection task with a single target, the impact of measurement noise correlations on the performance of an ideal Bayesian observer strongly depends on the relation between the statistics governing the external structure, and the parameters associated with the observer's measurements. Noise correlations influence the decisions of the ideal observer to varying extents given different parameter regimes (see Chapter 6).

Specifically, the interaction between these correlations are beneficial when the amount of structure in visual scenes is sufficiently strong, and pronounced (Figure 6.3(A)). Any strength of measurement noise correlations results in improved performance (Figure 6.3(C)). The measurement noise is assumed to be relatively

weak than the external stimuli noise in such a case. However, in the case of measurement noise being comparable to the external noise, the decisions of an ideal observer are also affected when external structure is weak (Figure 6.5(A)). Therefore, the role of measurement noise correlations in the decisions of an optimal observer heavily depends on the amount of structure present in an external visual scene.

The external statistical structure in a target detection task is determined by various factors: the number of stimuli, the number of targets, the mean orientation of the target, and distractors, the external noise level in the distractors, and the pairwise correlation coefficient between them. By fixing all other factors at a constant level, the amount of structure in a scene can be introduced by varying the amount of correlations between distractor orientations; namely from heterogeneous to homogeneous distractors (see Chapter 2). However, we can also vary any other parameter that drives the framework of a visual scene, and then study how noise correlations in the observer's measurements could possibly interact with the structured input to effect the performance of an ideal observer.

One possibility is to increase the number of stimuli in the task. The increased number of stimuli, and hence the distractors could possibly help an ideal observer in the case of homogeneous distractors; however, in general we would expect a lower performance when the distractors are not sufficiently aligned in a scene, for instance in the case of heterogeneous distractors. This is because the observer would be required to integrate more information from additional stimuli, and their respective locations here.

Rather, we consider an alternative to introduce more structure in the external scene. We allow the possibility of more than one target in a visual display. Since multiple targets are possible here, we refer to such a task as *multiple target detection task*. For a fixed number of stimuli in the task, increasing the number of targets certainly enhances the structure in the visual scene, and can potentially benefits the decisions of an ideal observer. However, the impact of measurement noise correlations is unknown in the case of such structured inputs. We thus explore how an optimal observer would behave in presence of external, and the measurement noise correlations on a multiple target detection task. In particular, we analyze the responses of the ideal observer in different regimes of parameters determining the structure in observer's measurements, and the visual scene.

Furthermore, we inspect whether noise correlations always reinforce the decisions of an observer or could they possibly hurt them? Alternatively, could a different objective in a same set of stimuli lead to different performance behavior? This is an important question, and we have briefly discussed this in the introduction of Chapter 6. We examine this question in a different task, namely, a *discrimination task*. We design the discrimination task with similar structural characteristics as that of the multiple target detection task, but consider a different objective for the observer. The aim is to study how noise correlations effect the performance of an optimal Bayesian observer in such a case, and how their influence varies with the objective of a task.

We begin this chapter with the description of the model for the multiple target detection task, followed by the mathematical derivation of the optimal-observer

model. We further analyze the influence of varying amounts of noise correlations on the distributions of the observer's responses, and the resulting performance. We conclude by presenting an example of a discrimination task which has similar structural properties as that of the multiple target detection task; but the noise correlations show an adverse effect on the performance.

7.1 Model description

The model set up for a multiple target detection task is nearly similar to the single target detection task described in Sections 2.1, and 6.1. We present a set of N stimuli to the observer, and the observer needs to infer whether a target stimulus is present in the visual display. A target is a stimulus with a particular orientation, denoted by s_T. For simplicity, we assume $s_T = 0°$, and measure stimulus orientations relative to that of a target. On half of the trials one or multiple targets are present among a set of *distractors*. All n targets have identical orientation. A distractor stimulus has a non-target orientation. Again, we denote target presence by $T = 1$, and absence by $T = 0$.

When $T = 0$, there are no targets, and stimulus orientations are drawn from a multivariate normal distribution described in Eq. (2.1) with mean $\mathbf{s_D} = \mathbf{0}_N = (0, 0, \cdots, 0)$, and covariance matrix, $\Sigma_\mathbf{s}$, so that

$$\mathbf{s}|T = 0 \sim \mathcal{N}(\mathbf{0}_N, \Sigma_\mathbf{s}).$$

We again assume $\Sigma_\mathbf{s}$ with constant diagonal terms, σ_s^2, and off-diagonal terms,

$\rho_s \sigma_s^2$, where the pairwise correlation coefficient, ρ_s determines the relation between the components of the distractors.

If $T = 1$, then $n \geq 1$ targets are present. To place the targets, we choose, with equal probability, n out of the N possible locations. The set of the n target locations is denoted by L. There are $\binom{N}{n}$ possible choices for this set, and we denote the collection of all possible choices of sets of target locations by $\mathcal{L}$. We also write

$$M = \binom{N}{n}$$

which is the cardinality of the set $\mathcal{L}$.

The orientations of the remaining $N - n$ distractors are chosen from a multivariate normal distribution with mean $\mathbf{0}_{N-n}$, and covariance matrix $\Sigma_{\mathbf{s}\backslash L}$ of dimension $(N - n) \times (N - n)$. Once the locations of the targets are chosen, let $\mathbf{s}_L = (s_{i_1}, s_{i_2}, \cdots, s_{i_n})$, $i_l \in L$ denote the orientations of the target stimuli, and $\mathbf{s}_{\backslash L} = (s_{j_1}, s_{j_2}, \cdots, s_{j_{N-n}})$, $j_l \notin L$ those of the distractors. We can therefore write

$$\mathbf{s}_L | T = 1 \sim \sum_{i \in L} \delta(s_i), \qquad \text{and} \qquad \mathbf{s}_{\backslash L} | T = 1 \sim \mathcal{N}(\mathbf{0}_{N-n}, \Sigma_{\mathbf{s}\backslash L}). \tag{7.1}$$

For $\eta > 0$, we introduce the following auxiliary covariance,

$$(\Sigma_{\mathbf{s},L}^{\eta})_{i,j} = \begin{cases} (\Sigma_{\mathbf{s}\backslash L})_{i,j}, & \text{if } i, j \notin L, \\ \eta, & \text{if } i = j \in L, \\ 0, & \text{if } i \in L, \text{ or } j \in L, \text{ and } i \neq j, \end{cases} \tag{7.2}$$

and write,

$$\mathbf{s} | (T = 1, L) \sim \lim_{\eta \to 0^+} \mathcal{N}(\mathbf{0}_N, \Sigma_{\mathbf{s},L}^{\eta}).$$

We note that the matrix $\Sigma_{\mathbf{s},L}^{\eta}$ reduces to $\Sigma_{\mathbf{s},j}^{\eta}$ defined in Eq. (6.4) in the case of a detection task with a single target.

We further assume that the observer's measurements, $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ are correlated so that it follows the multivariate normal distribution described in Eq. (6.1) as

$$\mathbf{x}|\mathbf{s} \sim \mathcal{N}(\mathbf{s}, \Sigma_\mathbf{x}).$$

Additionally, the covariance matrix $\Sigma_\mathbf{x}$ is assumed to have similar structure as $\Sigma_\mathbf{s}$ with constant diagonal terms, σ_x^2, and off-diagonal terms, $\rho_x \sigma_x^2$. Since the task variables are similar to the target detection task specified in Section 6.1, the generative model is same as shown in Figure 6.1.

7.2 Optimal observer theory

We now derive the mathematical model to understand the behavior of an optimal Bayesian observer on the task. In Sections 2.2, and 6.2, we described that an ideal observer makes a decision based on the sign of the log-posterior ratio. Here we denote the Bayesian decision variable by $d_{\mathrm{NMT}}(\mathbf{x})$,

$$d_{\mathrm{NMT}}(\mathbf{x}) = \log \frac{p(T = 1|\mathbf{x})}{p(T = 0|\mathbf{x})}. \tag{7.3}$$

Since the optimal observer uses a uniform prior distribution over T, we essentially compute the log-likelihood ratio

$$d_{\mathrm{NMT}}(\mathbf{x}) = \log \frac{p(\mathbf{x}|T = 1)}{p(\mathbf{x}|T = 0)}.$$

The optimal observer needs to marginalize over the intermediate variables, the spatial location vector, and the stimulus, $\mathbf{s}$ to compute the distribution of the measurements given the target presence variable, *i.e.*, $p(\mathbf{x}|T)$. Thus, we have

$$d_{\text{NMT}}(\mathbf{x}) = \log \frac{\int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|T=1)d\mathbf{s}}{\int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|T=0)d\mathbf{s}}.$$

We note that

$$p(\mathbf{s}|T=1) = \sum_{L\in\mathcal{L}} p(\mathbf{s}|T=1,L)p(L) = \frac{1}{M}\sum_{L\in\mathcal{L}} p(\mathbf{s}|T=1,L).$$

Therefore,

$$
\begin{aligned}
p(\mathbf{x}|T=1) &= \frac{1}{M}\int p(\mathbf{x}|\mathbf{s})\sum_{L\in\mathcal{L}} p(\mathbf{s}|T=1,L)d\mathbf{s} \\
&= \frac{1}{M}\sum_{L\in\mathcal{L}}\int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|T=1,L)d\mathbf{s} \\
&= \frac{1}{M}\lim_{\eta\to 0^+}\sum_{L\in\mathcal{L}}\int f(\mathbf{x};\mathbf{s},\mathbf{\Sigma_x})f(\mathbf{s};\mathbf{0}_N,\mathbf{\Sigma}^{\eta}_{\mathbf{s},L})d\mathbf{s} \\
&= \frac{1}{M}\sum_{L\in\mathcal{L}} f(\mathbf{x};\mathbf{0}_N,\mathbf{C}_L).
\end{aligned}
$$

Here we have defined $\mathbf{C}_L = \mathbf{\Sigma_x} + \mathbf{\Sigma}^0_{\mathbf{s},K}$. The above equation is obtained by applying the general rule of product, and integral for multivariate normal distributions described in Appendix B.1.2. Similarly, we define $\mathbf{C} = \mathbf{\Sigma_x} + \mathbf{\Sigma_s}$ to obtain

$$
\begin{aligned}
p(\mathbf{x}|T=0) &= \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|T=0)d\mathbf{s} \\
&= \int f(\mathbf{x};\mathbf{s},\mathbf{\Sigma_x})f(\mathbf{s};\mathbf{0}_N,\mathbf{\Sigma_s})d\mathbf{s} \\
&= f(\mathbf{x};\mathbf{0}_N,\mathbf{C}).
\end{aligned}
$$

Taking the required ratio of $P(\mathbf{x}|T = 1)$, and $P(\mathbf{x}|T = 0)$ gives us

$$d_{\text{NMT}}(\mathbf{x}) = \log \frac{1}{M} \sum_{L \in \mathcal{L}} \frac{f(\mathbf{x}; \mathbf{0}_N, \mathbf{C}_L)}{f(\mathbf{x}; \mathbf{0}_N, \mathbf{C})}$$

$$= \log \frac{1}{M} \sqrt{\frac{|\mathbf{C}|}{|\mathbf{C}_L|}} \sum_{L \in \mathcal{L}} \exp\left(-\frac{1}{2}\mathbf{x}^T \left(\mathbf{C}_L^{-1} - \mathbf{C}^{-1} \right) \mathbf{x} \right). \qquad (7.4)$$

We note that the determinant of the matrix $\mathbf{C}_L$ does not depend on the set L since all matrices of this form can be obtained from each other by permuting appropriate rows, and columns.

We further simplify the above equation by computing the determinant, and inverse of matrix, $\mathbf{C}_L$. The determinant, and inverse of $\mathbf{C}$ has already been computed in Section 6.2.1.1. We note that the covariance matrix $\mathbf{C}_L$ has the following form,

$$(\mathbf{C}_L)_{i,j} = \begin{cases} \eta + \sigma_x^2, & \text{if } i = j \in L, \\ \sigma_s^2 + \sigma_x^2, & \text{if } i = j \notin L, \\ \rho_x \sigma_x^2, & \text{if either } i \text{ or } j \in L, \text{ and } i \neq j, \\ \rho_s \sigma_s^2 + \rho_x \sigma_x^2, & \text{if } i, j \notin L, \text{ and } i \neq j. \end{cases}$$

We closely follow the process described in Section 6.2.1.1 for the inverse, and determinant computations. In particular, we decompose $\mathbf{C}_L = \mathbf{A_L} + \mathbf{U_L}\mathbf{E}\mathbf{V_L}$. The different matrix components in this decomposition are defined in a similar manner as seen in Section 6.2.1.1. Specifically, $\mathbf{E}$ is a 2×2 identity matrix,

$$(\mathbf{A_L})_{k,l} = \begin{cases} \eta + \sigma_x^2(1 - \rho_x), & \text{if } k = l \in L, \\ \sigma_s^2(1 - \rho_s) + \sigma_x^2(1 - \rho_x), & \text{if } k = l \notin L, \\ 0, & \text{otherwise}. \end{cases}$$

200

and, the columns of $N \times 2$ matrix $\mathbf{U_L}$ and rows of $2 \times N$ matrix $\mathbf{V_L}$ are given by

$$
\mathbf{U}_{\mathbf{L}(k,1)} = \begin{cases} \sigma_s^2(1 - \rho_s) + \sigma_x^2(1 - \rho_x), & \text{if } k \notin L, \\ \sigma_x^2(1 - \rho_x), & \text{if } k \in L, \end{cases} \quad \text{and } \mathbf{U}_{\mathbf{L}(k,2)} = \sigma_x^2(1 - \rho_x) \ \forall\, k,
$$

$$
\mathbf{V}_{\mathbf{L}(1,k)} = \begin{cases} 1, & \text{if } k \notin L, \\ 0, & \text{if } k \in L, \end{cases} \quad \text{and } \mathbf{V}_{\mathbf{L}(2,k)} = \begin{cases} 0, & \text{if } k \notin L, \\ 1, & \text{if } k \in L. \end{cases}
$$

We also recall the definition of the following quantities from Eq. (6.8)

$$
v = \frac{1}{\sigma_s^2(1 - \rho_s) + \sigma_x^2(1 - \rho_x)}, \quad \tilde{v} = \frac{1}{\sigma_x^2(1 - \rho_x)},
$$

$$
V = Nv, \quad a = \rho_s \sigma_s^2 + \rho_x \sigma_x^2, \quad \beta = \frac{a}{1 + aV},
$$

and further define

$$
V_{\backslash L} = (N - n)v, \quad \tilde{q} = a + n\rho_s \sigma_s^2 \rho_x \sigma_x^2 \tilde{v}, \tag{7.5}
$$

$$
\tilde{r} = 1 + \rho_s \sigma_s^2 V_{\backslash L}, \quad \text{and } \tilde{\gamma} = 1 + aV_{\backslash L} + n\rho_x \sigma_x^2 \tilde{v}\tilde{r}. \tag{7.6}
$$

Applying the Generalized Matrix Determinant Lemma, and the Woodbury formula (Appendix B.3), we obtain the following determinant, and inverse of the matrix $\mathbf{C}_L$:

$$
|\mathbf{C}_L| = \frac{\tilde{\gamma}}{\tilde{v}^n \, v^{N-n}},
$$

and

$$
(\mathbf{C}_L^{-1})_{i,j} = \begin{cases}
\tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 \tilde{r}}{\tilde{\gamma}}, & \text{if } i = j \in L, \\[2mm]
v - \frac{v^2 \tilde{q}}{\tilde{\gamma}}, & \text{if } i = j \notin L, \\[2mm]
-\frac{\rho_x \sigma_x^2 \tilde{v}^2 \tilde{r}}{\tilde{\gamma}}, & \text{if } i, j \in L, \text{ and } i \neq j, \\[2mm]
-\frac{v^2 \tilde{q}}{\tilde{\gamma}}, & \text{if } i, j \notin L, \text{ and } i \neq j, \\[2mm]
-\frac{\rho_x \sigma_x^2 v \tilde{v}}{\tilde{\gamma}}, & \text{if } i \neq j, i \in L, j \notin L \text{ or } i \neq j, i \notin L, j \in L.
\end{cases}
$$

As required in Eq. (7.4), we simplify

$$
\sqrt{\frac{|\mathbf{C}|}{|\mathbf{C}_L|}} = \sqrt{\frac{1 + aV}{\tilde{\gamma}}} \left(\frac{\tilde{v}}{v} \right)^n,
$$

and compute

$$
\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} = (v - \beta v^2) \sum_{i=1}^{N} x_i^2 - \beta v^2 \sum_{i \neq j}^{N} x_i x_j,
$$

$$
\mathbf{x}^T \mathbf{C}_L^{-1} \mathbf{x} = \left(\tilde{v} - \frac{\rho_x \sigma_x^2 \tilde{v}^2 \tilde{r}}{\tilde{\gamma}} \right) \sum_{i \in L} x_i^2 + \left(v - \frac{v^2 \tilde{q}}{\tilde{\gamma}} \right) \sum_{i \notin L} x_i^2
$$

$$
- \frac{\rho_x \sigma_x^2 \tilde{v}^2 \tilde{r}}{\tilde{\gamma}} \sum_{\substack{i,j \in L \\ i \neq j}} x_i x_j - \frac{2 \rho_x \sigma_x^2 v \tilde{v}}{\tilde{\gamma}} \sum_{\substack{i \in L \\ j \notin L}} x_i x_j - \frac{v^2 \tilde{q}}{\tilde{\gamma}} \sum_{\substack{i,j \notin L \\ i \neq j}} x_i x_j.
$$

We substitute the above expressions in Eq. (7.4) to obtain the following expression

for the optimal Bayesian decision variable on the task:

$$
\begin{aligned}
d_{\text{NMT}}(\mathbf{x}) = \log \frac{1}{M} &\sqrt{\frac{1 + aV}{\tilde{\gamma}} \left(\frac{\tilde{v}}{v}\right)^n} \sum_{L \in \mathcal{L}} \exp\left(-\frac{1}{2} \underbrace{\sigma_s^2 (1 - \rho_s) v \tilde{v} \sum_{i \in L} x_i^2}_{I} \right. \\
&- \frac{1}{2}\left(\beta v^2 - \frac{\rho_x \sigma_x^2 \tilde{v}^2 \tilde{r}}{\tilde{\gamma}}\right) \underbrace{\sum_{i,j \in L} x_i x_j}_{II} - \left(\beta v^2 - \frac{\rho_x \sigma_x^2 v \tilde{v}}{\tilde{\gamma}}\right) \underbrace{\sum_{i \in L, j \notin L} x_i x_j}_{III} \\
&\left. - \frac{1}{2}\left(\beta v^2 - \frac{v^2 \tilde{q}}{\tilde{\gamma}}\right) \underbrace{\sum_{i,j \notin L} x_i x_j}_{IV} \right).
\end{aligned}
\tag{7.7}
$$

The above equation describes the decision strategy of an optimal observer on the multiple target detection task. If $d_{\text{NMT}}(\mathbf{x}) > 0$, the observer responds the target is present, and target absent otherwise. Eq. (7.7) shows the intricate dependence of the decision variable, $d_{\text{NMT}}(\mathbf{x})$ on several parameters governing the structure of the task, such as the total number of stimuli N, number of targets n, variance σ_s^2 and correlation ρ_s of the distractors' orientations, and the parameters determining the structure of the measurements, σ_x^2, and ρ_x. We assume that the observer knows these parameters.

7.2.1 Interpretation of the decision variable

The decision variable derived in Eq. (7.7) is a generalized form of the decision variable computed in Eq. (6.12). Eq. (7.7) is much more complex, and depends in a complicated way on different parameters that describe the structure of the

stimulus, and the response. However, we can interpret it in a similar manner as in Section 6.2.3. The outer sum is over all putative sets of targets $L \in \mathcal{L}$. The summands correspond to the evidence that the set L contains targets. Each term in the exponent has an intuitive interpretation: (I) for the putative set of targets, L, the first term represents the sample second moment of the potential target measurements; if this term is large, the measurements are more likely to be away from zero, and hence it is less likely that the set L consists of targets; (II) the second term can be written in terms of sample covariance between putative targets, and a smaller value of this term corresponds to an increased chance of set L being a set of targets; (III) this term compares the sample means of measurements in the putative target set to those outside of it, a small value of this term provides evidence that the set L contains targets; (IV) the last term can be rewritten in terms of sample covariance of the measurements corresponding to the putative distractor set, *i.e.,* stimuli outside of set L, if these measurements are correlated, a large covariance would signify that this is a set of distractors, and hence it is more likely that the set L consists of targets. Again, it is more difficult to provide a precise interpretation of the different prefactors involved in Eq. (7.7) since they have more complicated dependence on various parameters in the generative model of the task.

We now wish to analyze the impact of different statistical parameters that govern the structure of the scene, and the measurements, on the performance of an ideal Bayesian observer. This goal is similar to our analysis performed in Chapter 6, but here in the case of the multiple target detection task. Specifically, we aim

to understand how the relationship between the parameters that determine the external structure of the visual display, namely, σ_x^2, and ρ_x; and the measurement parameters, σ_x^2, and ρ_x affect the decisions of an optimal observer. Since no explicit dependence is known among these parameters, and we therefore examine a range of possibilities here.

7.3 Analysis and results

Here we analyze the decision variable $d_{\mathrm{NMT}}(\mathbf{x})$ (Eq. (7.7)) in the regime of weak measurement noise, that is, when $\sigma_x^2 \ll \sigma_s^2$. In the case of strong measurement noise, $\sigma_x^2 \gg \sigma_s^2$, the noise dominates over other statistical parameters, and the observer only makes a guess about target presence. Therefore, we limit our analysis in the regime of weak measurement noise, and understand the impact of noise correlations on the performance of an Bayes-optimal observer.

As seen in Section 6.3.1, we again let $\epsilon = \frac{\sigma_x^2}{\sigma_s^2}$, and approximate Eq. (7.7) in the limit of $\epsilon \to 0$. We further split our analysis in the case of weak, and strong external structure, that is when $\rho_s < 1$, and $\rho_s \approx 1$, respectively.

7.3.1 Weak external structure, $\rho_s < 1$

In the absence of structured visual scenes, the coefficients of the exponent terms in Eq. (7.7) reduces to the following expressions in the limit of $\epsilon \to 0$

$$I : \sigma_s^2(1 - \rho_s)v\tilde{v} = \frac{1}{\sigma_x^2(1 - \rho_x)} + \mathcal{O}(1)$$

$$II : \beta v^2 - \frac{\rho_x \sigma_x^2 \tilde{v}^2 \tilde{r}}{\tilde{\gamma}} = -\frac{\rho_x}{\sigma_x^2(1 - \rho_x)(1 + (n-1)\rho_x)} + \mathcal{O}(1)$$

$$III : \beta v^2 - \frac{\rho_x \sigma_x^2 v\tilde{v}}{\tilde{\gamma}} \approx \mathcal{O}(1)$$

$$IV : \beta v^2 - \frac{v^2 \tilde{q}}{\tilde{\gamma}} \approx \mathcal{O}(1).$$

Also, the leading determinant term approximates to

$$\sqrt{\frac{1 + aV}{\tilde{\gamma}}} \left(\frac{\tilde{v}}{v}\right)^n \approx \sqrt{\frac{(1 - \rho_x)(1 + (N-1)\rho_s)}{(1 + (N - n - 1)\rho_s)(1 + (n-1)\rho_x)}} \left(\frac{\sigma_s^2(1 - \rho_s)}{\sigma_x^2(1 - \rho_x)}\right)^n.$$

Therefore, we obtain the following approximation for the decision variable in the parameter regime of $\sigma_x^2 \ll \sigma_s^2$, and $\rho_s < 1$:

$$d_{\text{NMT}}(\mathbf{x}) \approx \log \frac{1}{M} \sqrt{\frac{(1 - \rho_x)(1 + (N-1)\rho_s)}{(1 + (N - n - 1)\rho_s)(1 + (n-1)\rho_x)}} \left(\frac{\sigma_s^2(1 - \rho_s)}{\sigma_x^2(1 - \rho_x)}\right)^n$$

$$\sum_{L \in \mathcal{L}} \exp\left(-\frac{1}{2\sigma_x^2(1 - \rho_x)}\left(\sum_{i \in L} x_i^2 - \frac{\rho_x}{1 + (n-1)\rho_x}\sum_{i,j \in L} x_i x_j\right)\right). \quad (7.8)$$

Special case: $n = 1$ We note that in the case of a single target, set L has only one element, and $\mathcal{L} = \{1, 2, \cdots, N\}$. Thus, when $s_T = s_D = 0°$, Eq. (7.8) reduces to the Eq. (6.16) that has been examined in details in Chapter 6.

We observe that Eq. (7.8) explicitly depends on ρ_x, unlike the case of a single target in the same regime (Section 6.3.1.2). Also, these noise correlations clearly

influence the performance of an ideal observer as seen in Figure 7.1(A). The simulations were performed with $N = 4$ stimuli, and $n = 3$ targets in the task. The external noise, and measurement noise were fixed at $\sigma_s^2 = 15°$, and $\sigma_x^2 = 4°$. Since the n target stimuli have identical orientation, they are perfectly correlated. This introduces a relation between a subset of the stimuli on half of the trials, even when $\rho_s = 0$.

In general, for a fixed ρ_s, we observe in Figure 7.1(A) that the performance increases with ρ_x. This is because the stimuli have a strong structure as all the targets are identically oriented, and thus the distribution $\mathbf{s}|T = 1$ lies on a low dimensional subspace. Therefore, we expect that measurement noise correlations can have a significant impact in the presence of such structure. This is in addition to the trivial increase in performance expected simply from having more targets.

As seen in Figure 7.1(A), and (B), an ideal observer takes into account noise correlations for all values of ρ_s even when measurement noise is low. This lends to an increase in performance as ρ_x is increased beyond some critical value for any ρ_s. This is different from the single target case (Eq. (6.16)) where the decision variable was independent of measurement noise correlations when $\rho_s < 1$.

When $\rho_s < 1$, performance gradually increases with ρ_x. The target stimuli are a perfectly correlated subset of the stimulus set. The measurements of these stimuli are identical when noise correlations are perfect. In such a case, an ideal observer performs perfectly by detecting whether any n of the N measurements x_i are equal.
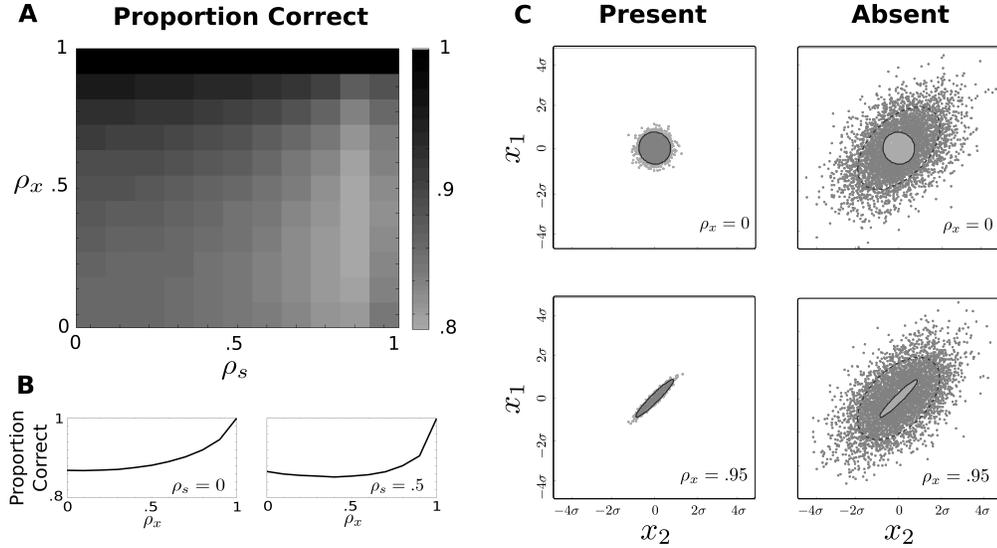
Figure 7.1: **Impact of measurement noise correlations on the performance of an ideal Bayesian observer on the multiple target detection task when the measurement noise is weak as compared to the external noise,** $\sigma_x^2 \ll \sigma_s^2$ $(\sigma_s = 15°, \sigma_x = 4°)$. **(A)** Performance of an ideal Bayesian observer as a function of external correlation strength, ρ_s, and measurement noise correlations, ρ_x on the task with $N = 4$ stimuli, and $n = 3$ targets. **(B)** The proportion of correct responses as a function of ρ_x in the case of $\rho_s = 0$ (left), and $\rho_s = 0.5$ (right) when $N = 4$, and $n = 3$. **(C)** The decision boundary, $d_{\text{NMT}}(\mathbf{x}) = 0$, with target present (left), and target absent (right) distributions in the case of $\rho_s = 0.5$, and $N = n = 2$ stimuli. Here light gray dots correspond to incorrect inferences while dark gray dots represents correct responses. The black dashed lines show 2 units of standard deviation for the stimulus distribution. The axes are measured in terms of the standard deviation σ which is defined by $\sigma^2 = \sigma_s^2 + \sigma_x^2$.

We now analytically understand how different amounts of noise correlations impact the performance in presence of weak external structure, $\rho_s < 1$ as seen in Figure 7.1(A).

7.3.1.1   No noise correlations, $\rho_x = 0$

In the absence of noise correlations *i.e.*, $\rho_x = 0$, a decision is based solely on the sample second moment *about the target orientation* of the measurements of the n stimuli in the putative target set.

$$d_{\mathrm{NMT}}(\mathbf{x}) \approx \log \frac{1}{M} \sqrt{\frac{(1 + (N-1)\rho_s)}{(1 + (N-n-1)\rho_s)(1 + (n-1)\rho_s)}} \left(\frac{\sigma_s^2(1 - \rho_s)}{\sigma_x^2} \right)^n$$

$$\sum_{L \in \mathcal{L}} \exp \left(-\frac{1}{2\sigma_x^2} \sum_{i \in L} x_i^2 \right). \tag{7.9}$$

Since we have assumed the target orientation to be $0°$, this value does not appear explicitly in the centered second moment in Eq. (7.9). If i is a target, the variance of the measurement x_i is σ_x^2, and if i is a distractor the variance is $\sigma_s^2 + \sigma_x^2$. Therefore, if a subset, L, of stimuli contains targets, the centered sample second moment will be smaller, than if it does not. Hence, in the absence of noise correlations, the ideal observer does not take relations between the measurements into account, and instead compares the stimulus orientations in the putative set of targets to the known target orientation.

7.3.1.2 Perfect noise correlations, $\rho_x \to 1$

On the other hand, when the measurements are structured, *i.e.*, in the case of $\rho_x > 0$, the ideal observer computes the second moment about a point between the target orientation, and the sample mean, $\frac{1}{n}\sum_{i \in L} x_i$. This can be seen by rewriting Eq. (7.8) as

$$
d_{\text{NMT}}(\mathbf{x}) \approx \log \frac{1}{M} \sqrt{\frac{(1-\rho_x)(1+(N-1)\rho_s)}{(1+(N-n-1)\rho_s)(1+(n-1)\rho_s)} \left(\frac{\sigma_s^2(1-\rho_s)}{\sigma_x^2(1-\rho_x)}\right)^n}
$$
$$
\sum_{L \in \mathcal{L}} \exp\left(-\frac{n}{2\sigma_x^2(1-\rho_x)}\underline{\left(\frac{1}{n}\sum_{i \in L} x_i^2 - \frac{n\rho_x}{1+(n-1)\rho_x}\left(\frac{1}{n}\sum_{i \in L} x_i\right)^2\right)}\right).
$$

(7.10)

Furthermore, in the limit of $\rho_x \to 1$, the underlined term in Eq. (7.10) approaches the sample variance – the sample second moment centered at the sample mean,

$$
\frac{1}{n}\sum_{i \in L} x_i^2 - \frac{n\rho_x}{1+(n-1)\rho_x}\left(\frac{1}{n}\sum_{i \in L} x_i\right)^2 \to E_L = \frac{1}{n}\sum_{i \in L} x_i^2 - \left(\frac{1}{n}\sum_{i \in L} x_i\right)^2. \quad (7.11)
$$

In the case of strong noise correlations, the target stimuli in set L have approximately equal measurements. Therefore, the optimal observer computes the sample second moment centered around the sample mean (the sample variance).

We first consider the case of target present trials, $T = 1$. We note that E_L in Eq. (7.11) approaches zero for a set of targets, L_T, and therefore, the exponential in Eq. (7.10) becomes unity. In the event that L is a set not consisting of all targets, $E_L > 0$, and

$$
\sum_{L \in \mathcal{L}} \exp\left(-\frac{nE_L}{2\sigma_x^2(1-\rho_x)}\right) = 1 + \sum_{L \in \mathcal{L} \setminus L_T} \exp\left(-\frac{nE_L}{2\sigma_x^2(1-\rho_x)}\right) \to 1.
$$

The prefactor in Eq. (7.10) diverges. However, since $\lim_{x\to\infty} \exp(-x)x^n = 0$ for any finite n, the product of the prefactor, and the exponential terms (not equal to 1) still converge to zero. Therefore, on target present trials, we obtain $d_{\mathrm{NMT}}(\mathbf{x}) \to \infty$ as $\rho_x \to 1$.

On target absent trials, $T = 0$, the expression E_L will be greater than zero for all sets, L. Thus, $d_{\mathrm{NMT}}(\mathbf{x}) \to -\infty$ in the limit of $\rho_x \to 1$. Hence, regardless of external correlations ρ_s, the ideal observer performs perfectly when $\rho_x = 1$.

7.3.1.3 Intermediate ρ_x

When $0 < \rho_x < 1$, the ideal observer makes a decision by adopting an intermediate strategy, and computes a second moment about a point between the target orientation, and the sample mean (Eq. (7.8)). Moreover, the weight on the sample mean increases with the number of targets in set L, since the prefactor $n\rho_x/(1 + (n - 1)\rho_x)$ grows with n for fixed ρ_x.

We also observe in Figure 7.1(C) that noise correlations significantly impact the distribution of measurements. With $N = n = 2$, the distribution approaches the diagonal ($x_1 = x_2$) as $\rho_x \to 1$. This decreases the overlap between in the target present, and absent distributions.

Therefore, with multiple targets, there is always structure in the stimulus set, and $P(\mathbf{s}|T = 1)$ is always concentrated on a low dimensional subspaces. In this case, the structure present in the observer's measurements can again decrease the

overlap between the distribution of measurements, and significantly impact deci-sions, and hence, the performance.

7.3.2   Strong external structure, $\rho_s = 1$

In Figure 7.1(A), we also observe that the performance steadily increases with ρ_x when $\rho_s = 1$, and a perfect performance is achieved at $\rho_s = \rho_x = 1$. This behavior is quite similar to the single target case seen in Figure 6.3(A). We thus expect a similar approximation of the decision variable, $d_{\mathrm{NMT}}(\mathbf{x})$ in this case.

In the case of homogeneous distractors, $\rho_s = 1$, we obtain the following re-duced expressions for the different exponent terms in Eq. (7.7):

$$I : \sigma_s^2(1 - \rho_s)v\tilde{v} = 0$$

$$II : \beta v^2 - \frac{\rho_x \sigma_x^2 \tilde{v}^2 \tilde{r}}{\tilde{\gamma}} = \frac{1 - (N - n + 1)\rho_x}{N(1 + (n-1)\rho_x)\sigma_x^2(1 - \rho_x)} + \mathcal{O}(1)$$

$$III : \beta v^2 - \frac{\rho_x \sigma_x^2 v\tilde{v}}{\tilde{\gamma}} = \frac{1}{N\sigma_x^2(1 - \rho_x)} + \mathcal{O}(1)$$

$$IV : \beta v^2 - \frac{v^2 \tilde{q}}{\tilde{\gamma}} = \frac{-n - N(n-1)\rho_x}{N(N-n)\sigma_x^2(1 - \rho_x)} + \mathcal{O}(1).$$

The determinant prefactor becomes

$$\sqrt{\frac{1 + aV}{\tilde{\gamma}}} \left(\frac{\tilde{v}}{v}\right)^n \approx \sqrt{\frac{N(1 - \rho_x)}{(N - n)(1 + (n-1)\rho_x)}}.$$

Combining the above expressions gives us the following approximated decision

variable:

$$d_{\text{NMT}}(\mathbf{x}) \approx \log \frac{1}{M} \sqrt{\frac{N(1-\rho_x)}{(N-n)(1+(n-1)\rho_x)}} \sum_{L \in \mathcal{L}} \exp\left(-\frac{1}{2}\frac{1}{N\sigma_x^2(1-\rho_x)}\right.$$

$$\left.\underbrace{\left(\frac{1-(N-n+1)\rho_x}{(1+(n-1)\rho_x)}\sum_{i,j \in L} x_i x_j + 2\sum_{i \in L, j \notin L} x_i x_j - \frac{n+N(n-1)\rho_x}{(N-n)}\sum_{i,j \notin L} x_i x_j\right)}_{F_L}\right).$$

(7.12)

Again, the above equation reduces to Eq. (6.14) studied in Chapter 6 in the case of a single target, $n = 1$.

It is evident in Eq. (7.12) that ρ_x impacts the decisions of an optimal observer when external structure is predominant. When $\rho_x = 1$, the expression F_L in Eq. (7.12) becomes

$$F_L = \frac{n-N}{n}\left(\sum_{i \in L} x_i\right)^2 + 2\sum_{i \in L, j \notin L} x_i x_j - \frac{n+N(n-1)}{(N-n)}\left(\sum_{i,j \notin L} x_i\right)^2.$$

Therefore, the ideal observer computes the decision in this case by comparing the sample second moment of the elements in the putative target set, L to those not in the set, and also considering the product of the sample means of the two sets. The measurements of the target stimuli are equal, and also the distractors measurements are similar when noise correlations are perfect. Thus, on both target absent and present trials, F_L converges to a finite number, but the exponential prefactor in Eq. (7.12) diverges in the limit of $\rho_x \to 1$. Therefore, $d_{\text{NMT}}(\mathbf{x}) \to +\infty$ on target present trials, and it goes to $-\infty$ on target absent trials. Hence, the observer performs perfectly in such a case. Further, the trend in performance for any ρ_x

when $\rho_s = 1$ can be explained in similar terms as seen in the case of a single target (Section 6.3.1.1).

In sum, we found that the impact of noise correlations on the performance of an optimal Bayesian observer is significant with multiple targets as compared to the case of a single target (Chapter 6). This is attributed to the presence of more structured input, and the interplay between the noise correlations present in the stimuli, and the observer's measurements. Thus, presence of structured displays enhance the joint effects of stimulus, and measurement noise correlations.

However, the role of measurement correlations in response to the fixed input could also depend on the actual objective of the task rather than the true structure present in a visual scene. For instance, two different tasks can be performed with the same visual input but the impact of noise correlations can vary extensively given the aim of the task.

In order to explore the possible dependence of noise correlations on the objective of a task, we consider an example of a discrimination task in the following section.

7.4 Mean stimulus orientation discrimination task

As before, we consider the task where an observer is presented with N stimuli in a visual display. The stimulus orientations denoted by $\mathbf{s} = (s_1, s_2, \cdots, s_N)$ are relevant features of the task. We measure orientations relative to the vertical,

which we denote by 0. The task for the observer is to decide whether the mean orientation of the stimulus set is to the left (this is denoted by $C = -1$), or right ($C = 1$) of the vertical. The binary variable, C represents the two classes that need to be discriminated. Hence, this type of a task is known as a *discrimination task*.

Stimulus orientations are drawn from a multivariate normal distribution with mean vector, $\mathbf{0}_N$, and covariance matrix, $\Sigma_{\mathbf{s}}$, so that

$$\mathbf{s} \sim \mathcal{N}(\mathbf{0}_N, \Sigma_{\mathbf{s}}). \tag{7.13}$$

We have defined the matrix $\Sigma_{\mathbf{s}}$ in Eq. (2.3). The observer makes a decision based on the measurements of the presented stimuli orientations, denoted by $\mathbf{x} = (x_1, x_2, \cdots, x_N)$. Similar to the target detection tasks discussed in Sections 6.1, and 7.1, we assume the measurements to be unbiased, and follow multivariate normal distribution with mean, $\mathbf{s}$ and covariance matrix, $\Sigma_{\mathbf{x}}$ as

$$\mathbf{x}|\mathbf{s} \sim \mathcal{N}(\mathbf{s}, \Sigma_{\mathbf{x}}).$$

We observe that the framework of the task, and thus the structure of visual inputs is similar to the target detection tasks analyzed in Chapter 6, and here in Section 7.1. Except that, we do not have any characterization of the target in this case. The observer is only interested in determining whether $C = 1$, or $C = -1$, instead of detecting any particular orientation of the stimuli.

7.4.1 Inference process

An optimal Bayesian observer needs to determine whether the mean orientation of the set of stimuli is oriented to the left, or right of the vertical. We denote the mean orientation of the set of stimuli on a trial by $\bar{s}$, so that $\bar{s} = \sum_{i=1}^{N} s_i$. The optimal observer performs the computation based on the log-posterior ratio of the mean stimulus orientation being left, or right given the measurements. We denote the Bayesian decision variable on the task by $d_{\mathrm{MD}}(\mathbf{x})$, is given by

$$
\begin{aligned}
d_{\mathrm{MD}}(\mathbf{x}) &= \log \frac{p(C = 1|\mathbf{x})}{p(C = -1|\mathbf{x})} = \log \frac{p(\bar{s} > 0|\mathbf{x})}{p(\bar{s} < 0|\mathbf{x})} \\
&= \log \frac{p(\mathbf{x}|\bar{s} > 0)}{p(\mathbf{x}|\bar{s} < 0)} + \log \frac{p(\bar{s} > 0)}{p(\bar{s} < 0)}.
\end{aligned} \tag{7.14}
$$

We denote the observer's MAP estimate of C by $\hat{C}$. When the decision variable, $d_{\mathrm{MD}}(\mathbf{x}) > 0$, the observer infers $\hat{C} = 1$, and responds the mean stimulus orientation to the right of the vertical, that is, $\bar{s} > 0$. If $d_{\mathrm{MD}}(\mathbf{x}) < 0$, the observer reports $\hat{C} = -1$, and therefore $\bar{s} < 0$.

To compute the density function $p(\mathbf{x}|C)$ in Eq. (7.14), we marginalize the observer's information over the variable $\mathbf{s}$, and further apply Bayes' rule to obtain

$$
\begin{aligned}
p(\mathbf{x}|\bar{s} > 0) &= \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|\bar{s} > 0)d\mathbf{s} \\
&= \int p(\mathbf{x}|\mathbf{s})p(\bar{s} > 0|\mathbf{s})\frac{p(\mathbf{s})}{p(\bar{s} > 0)}d\mathbf{s} \\
&= \frac{1}{p(\bar{s} > 0)} \int_{\bar{s}>0} p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s} \\
&= \frac{1}{p(\bar{s} > 0)} \int_{\bar{s}>0} f(\mathbf{x}; \mathbf{s}, \Sigma_{\mathbf{x}})f(\mathbf{s}; \mathbf{0}_N, \Sigma_{\mathbf{s}})d\mathbf{s} \\
&= \frac{k_{\mathbf{x}}}{p(\bar{s} > 0)} \int_{\bar{s}>0} f\left(\mathbf{s}; \left(\mathbf{I} + \Sigma_{\mathbf{x}}\Sigma_{\mathbf{s}}^{-1}\right)^{-1}\mathbf{x}, \left(\Sigma_{\mathbf{s}}^{-1} + \Sigma_{\mathbf{x}}^{-1}\right)^{-1}\right)d\mathbf{s}.
\end{aligned}
$$

Appendix Eq. (B.3) is used above to simplify the product of two multivariate normal distributions, and here $k_{\mathbf{x}} = \frac{1}{\sqrt{2\pi|\Sigma_{\mathbf{s}}+\Sigma_{\mathbf{x}}|}} \exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}(\Sigma_{\mathbf{s}} + \Sigma_{\mathbf{x}})^{-1}\mathbf{x}\right)$ is a normalization constant. Similarly, we compute $p(\mathbf{x}|\bar{s} < 0)$, and obtain the following expression for the Bayesian decision variable, $d_{\mathrm{MD}}(\mathbf{x})$ that an optimal observer uses to make a decision:

$$
d_{\mathrm{MD}}(\mathbf{x}) = \log \left(\frac{\int_{\bar{s}>0} f\left(\mathbf{s}; \left(\mathbf{I} + \Sigma_{\mathbf{x}}\Sigma_{\mathbf{s}}^{-1}\right)^{-1}\mathbf{x}, \left(\Sigma_{\mathbf{s}}^{-1} + \Sigma_{\mathbf{x}}^{-1}\right)^{-1}\right) d\mathbf{s}}{\int_{\bar{s}<0} f\left(\mathbf{s}; \left(\mathbf{I} + \Sigma_{\mathbf{x}}\Sigma_{\mathbf{s}}^{-1}\right)^{-1}\mathbf{x}, \left(\Sigma_{\mathbf{s}}^{-1} + \Sigma_{\mathbf{x}}^{-1}\right)^{-1}\right) d\mathbf{s}} \right). \tag{7.15}
$$

The above equation characterizes the decisions of the ideal observer on the task where the observer needs to discriminate whether the mean orientation of the presented stimuli on a trial is to the left, or right of the vertical. Though the expression is not explicit, it depends on the various parameters that govern the external structure of a visual scene, and those that determine the structure in the observer's measurements. For instance, the total number of stimuli N, the variance and correlation of stimuli, σ_s^2, and ρ_s shape the structure of the input stimuli, while σ_x^2, and ρ_x drive the configuration of the measurements.

We find that Eq. (7.15) is difficult to further simplify analytically, and therefore, we analyze the decisions of the ideal observer by performing numerical simulations. We present the obtained simulation results below.

7.4.2 Results

We observe that the structure of stimulus input in the task is similar to the multiple target detection task discussed in Section 7.1. We are interested in determining

whether the effect of noise correlations changes with the objective of the task here. We thus compare the performance of an optimal observer on the two tasks, and analyze the impact of noise correlations in both cases.

For comparison purposes, we simulate the mean orientation discrimination task with $N = 4$ stimuli, $\sigma_s^2 = 15°$, and $\sigma_x^2 = 4°$. Figure 7.2(A) shows the trend in the performance of an optimal Bayesian observer on the task as a function of ρ_x when $\rho_s = 0.5$.
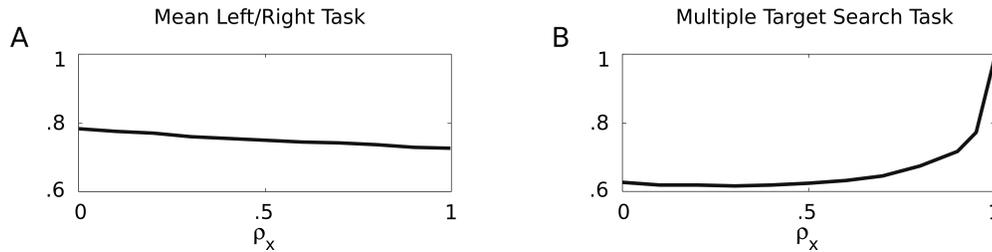
Figure 7.2: Performance of an optimal observer as a function of strength of noise correlations **(A)** on the mean left/right discrimination task and, **(B)** target detection task with two targets.

We note that the performance gradually decreases as ρ_x increases. This is in contrast to the performance behavior on a multiple target detection task in Figure 7.2(B). The increasing noise correlations improve the performance of the ideal observer on the multiple target detection task, and indeed perfect performance is possible when responses are perfectly correlated. However, noise correlations have a negative impact on the performance when the task is to discriminate the mean stimulus orientation, instead of finding a target in the same visual structured scene.

Though the input is structured in a similar way in both cases, the observer needs to use different strategies to make decisions on the two tasks. In the discrimination task, the observer is asked to integrate information from different sources. The detection task requires extracting information about targets buried in a sea of distractors. Intuitively, a choice between two possibilities needs to be made based on the measurement of a set of stimuli. Each choice corresponds to a distribution of measurements. The difficulty of the task depends on how much the degree of overlap between these two distributions. The higher the overlap, the more difficult it is to tell which distribution an measurement belongs to, and the more difficult the decision. External structure, as well as structured noise in the measurements impacts the overlap between these distributions. Therefore, performance of an ideal observer depends not only on the level, but also on the structure of the measurement noise [6, 4]. Thus the particulars of the task, the structure of the stimulus, as well as the level, and structure of the measurements jointly determine performance.

Hence, the role of noise correlations can be subtle, and depend on the nature of the task. We thus have examined their impact on detecting a single (Chapter 6), and multiple targets in a group of distractors whose orientations are chosen with varying degrees of dependence.

Chapter 8

Discussion

Understanding how correlations between stimuli, and measurements affect our decisions is important to understand how our visual perceptual system responds to structured input along with the structured measurements. Here we presented a thorough analysis of an experimental study designed to investigate how humans make decisions in response to stimuli having varying degrees of structure on a visual search task. Further, we theoretically analyzed the joint effects of stimulus, and measurement correlations on the performance of an ideal observer in a family of visual search tasks. Below, we discuss our findings, and their limitations along with potential generalizations.

Stimulus correlations in visual search

Many recent studies reported that humans performed near optimally on visual search tasks [90, 142, 94, 95]. That is, humans were able to make best possible decisions while searching for a predefined target among non-relevant distractors. These studies only used two types of distractors - homogeneous (identical), and heterogeneous (independent). Moreover, the orientations of the homogeneous distractors were completely predictable (same across trials) except in [95]. Mazyar et al. [95] were the first to distinguish homogeneity from predictability. They manipulated the distractors statistics by using trial-to-trial variability in homogeneous displays. The distractors were still identical to each other, but the orientation of the distractor was randomly chosen across trials. Regardless of introducing variability across trials in homogeneous displays, and changing the degree of heterogeneity, their experiments focused on two extreme conditions of distractors - identical, and independent random orientations.

In our work, we thus explored the intermediate regime of correlations between distractors. Visual stimuli in natural scenes can be correlated to different extents with each other, therefore, it is important to understand how our visual decisions are affected by inputs that have intermediate correlations. Using a target detection experiment, we examined whether humans take into account stimulus (distractors) correlations in visual search. We varied the amount of correlations between stimuli across different experimental sessions. Different correlations introduced

varying amount of structure in visual displays - from none (heterogeneous distractors) to mid-level (partial correlated distractors) to high-level (identical distractors). Due to individual differences, we obtained mixed results in our model comparisons, and do not have a clear presentation of how subjects inferred different correlation strengths. But, based on the assumptions of our best fitted models on different criteria, we found that subjects accounted for stimulus correlations in their responses, however, they were suboptimal in inferring the true correlation strength of distractors in an experimental session.

Specifically, the favorable model based on Bayesian information criterion suggests that perhaps subjects were unable to distinguish the partially correlated conditions, and they used a constant correlation strength to make decisions on those experimental sessions. While, subjects inferred a near-to-optimal correlation strength when distractors were perfectly correlated. This indicates that probably subjects use different inference process when making decisions on homogeneous displays than when they are not. Perhaps humans perceive completely structured input differently while they may be unable to make a clear distinction between inputs having partial correlations, and those having no structure. Further, we found that distribution of encoding precision was dependent on the correlation strength used across experimental sessions. This dependence is difficult to explain, but suggests that possibly subjects encode stimuli with different level of precision when they are highly structured, and differently when the statistical structure in a scene is relatively weak. A similar observation was also made by Mazyar et al. [95] about different dependence of precision in homogeneous, and

heterogeneous conditions, however, they analyzed the set-size effects.

Another winning model based on Akaike information criterion suggests that subjects treated each experimental condition differently, but suboptimally. They incorrectly inferred the true correlation strength used to generate stimuli in an experimental session. Based on the parameter estimates of this model, and experimentally obtained psychometric curves, we concluded that possibly subjects behave similarly in the conditions when distractors have intermediate strengths of correlations, and they perform distinctly in the case of identical distractors.

We note that these findings depend on the choice of models, and the model comparison techniques. It is always possible that there are better models, and better explanations. For instance, we assumed that subjects correctly inferred the standard deviation of distractor orientations in the experiment. It might be possible that subjects used some other possible values of this standard deviation to make decisions. In that case, we need to use plausible assumptions on this parameter in our models, and test whether subjects were able to infer this value correctly. It is also possible that subjects do not use any information about stimulus correlations, and other parameters in the generative model of the task, and instead they use alternative suboptimal strategies such as threshold criteria [90], to make their decisions. Apart from these model assumptions, one of the limitations in obtaining clear results could be the small number of subjects in our experimental study. We only have data collected from small finite number of participants in the experiment. Thus, the individual differences are dominant in average model comparisons, and lead to mixed results.

There are many more alternatives that exist, and could possibly provide a better explanation for our experimental data. However, due to computational, and time constraints, not every possibility can be tested. The models we tested encompassed a variety of possible assumptions about subject's behavior, and are general enough to include large number of alternatives. Also, the question we examined in our study is relevant for any psychophysical study in general, not only to visual search. Structured, or correlated information is available in a variety of stimuli (not just visual scenes) such as audio signals, odors. It is important to understand how brain integrates correlated inputs to extract relevant information about the state of the world. Our work is a little step in this direction, and we hope that our findings may have more general implications.

Interplay between stimulus and measurement correlations

Several theoretical, and experimental studies have modeled decision processes by which brain converts sensory measurements of a set of stimuli into a judgement about the world. But, many of them relied on stereotypical assumptions about the measurements being independent (across stimuli), and normally distributed [90, 94, 95, 142]. We extended our work here by focusing on the effects of violation of the assumption of independent measurements on performance in categorical, global perceptual judgements. It has been found that neural correlations can extend to long distances in visual cortex [36, 28] which suggests that

the sensory measurements can be strongly correlated [118, 27]. But the effects of measurement correlations can not be studied in isolation, they may be paired with statistical structure of stimuli [95]. Therefore, in this work, we examined the joint influence of stimulus, and measurement correlations on the performance of an ideal observer in a family of visual search tasks, namely target detection tasks.

We found that the relation between correlations of stimuli, and sensory measurements play a significant role in the decision-making strategy of the ideal observer. The measurement correlations help in preserving the statistical structure of the stimuli on a multiple target detection task, and hence enhances the performance. While, they have no effect on decisions in the case of a single target. This demonstrates that probably the role of measurement correlations only come into play when sufficiently strong structure is present in external scenes. In the case of weak external structure, there is perhaps very little structure present in visual displays, and correlations between sensory measurements are inefficient in preserving those weak structures. We observed that an ideal observer always perform perfectly when the distractors are homogeneous, and measurements are perfectly correlated in both single, and multiple target detection tasks. Such an observation clearly reflects the interplay between both correlations, and how preserved structure of stimuli by measurement correlations benefits the decisions of the observer.

We also examined that the interaction between stimulus, and measurement correlations also depend on the relation between the external noise, and the standard deviation of the measurements distribution. In the regime of strong external

noise than measurement noise, the trend in performance is easier to interpret in a single target detection task. While in the case of dominating measurement noise, a clear trend was not seen, and it was relatively difficult to understand how the two correlations interact to increase the performance at a particular level, but not at others.

Apart from the dependence on different parameters that control the statistical structure of visual displays, we found that influence of correlations between sensory measurements can also depend on the nature, or objective of a task. The measurement correlations can have different effects in response to same structured stimuli in tasks with different objectives. We observed that while these correlations enhance the performance of the ideal observer on a multiple target detection task, they negatively impact the performance on a discrimination task where the observer needs to decide whether the mean stimulus orientation is to the left or right of the vertical. The statistical structure of visual scenes in the two tasks were kept the same, but the observer was asked to make two different decisions - finding a predefined target among distractors, and judging whether mean stimulus orientation of a set of stimuli orients to the left or right of vertical. This finding suggests that the role of measurement correlations can be subtle, and the nature of the task, along with the structure of stimulus jointly determine the influence of these correlations on performance of the observer in visual perceptual tasks.

We understand that our findings are limited to perceptual decisions in visual search tasks, and the interaction between stimulus, and measurement correlations can have a variety of other possible effects on our perceptual system. Therefore,

it is important to understand how these correlations jointly interact, and govern our decisions. Further work needs to be done to identify the regimes where the effects of both these correlations are dominant, and can significantly impact our decision-making strategies.

A

Appendix A

Notation table

Symbol	Description
$\mathcal{N}(\mu, \sigma^2)$	a one-dimensional normal distribution with mean μ, and variance σ^2
$f(z; \mu, \sigma^2)$	a one-dimensional normal density function of variable, z with mean μ, and variance, σ^2
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	an N-dimensional multivariate normal distribution with N-dimensional mean vector $\boldsymbol{\mu}$, and $N \times N$ covariance matrix $\boldsymbol{\Sigma}$
$f(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	an N-dimensional normal density function of variable, $\mathbf{z}$ with mean vector, $\boldsymbol{\mu}$, and covariance, $\boldsymbol{\Sigma}$; $\frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})\right)$
$\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	a random variable $\mathbf{z}$ having a normal distribution with probability density function, $f(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

Continued on next page

Symbol	Description
$\mathbf{z}\|Y \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	a random variable $\mathbf{z}$ conditioned on Y having a normal distribution with probability density function, $f(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$
$\mathbf{0}_N$	an N-dimensional zero vector, $(0, 0, \cdots, 0)$
$\mathbf{1}_j$	an N-dimensional vector with j^{th} entry as 1, and rest zeros, $(0, 0, \cdots, 0, 1, 0, \cdots, 0)$
$\mathbf{z}_{\backslash j}$	$(z_1, z_2, \cdots, z_{j-1}, z_{j+1}, \cdots, z_N)$
$\boldsymbol{\Sigma}_{\backslash j}$	an $N \times N$ matrix obtained by removing j^{th} row, and column of matrix $\boldsymbol{\Sigma}$
$\|\mathbf{A}\|$	determinant of the matrix $\mathbf{A}$

Table A.1: **Mathematical Notation.** Description of the mathematical notation used in the text.

Some mathematical results

We present some known mathematical results, and theorems here that are used in the main text to derive several results.

B.1 Product and integral of normal distributions

The product of m ($m \geq 2$) normal distributions over a single variable is computable, and is itself a normal distribution. Also, the integral of obtained product distribution is analytically tractable. We list below few relevant results about the product, and integral of normal distributions.

B.1.1 Univariate normal distributions

Consider two Gaussian probability distributions over a single random variable, z. Assume that the means of the two distributions are denoted by μ_1, and μ_2; and variances as σ_1^2, and σ_2^2. The product of these two probability distributions is a normal distribution, and the (normalized) probability density function of the resulting distribution can be computed as

$$f(z; \mu_1, \sigma_1^2) \cdot f(z; \mu_2, \sigma_2^2) = f(\mu_1; \mu_2, \sigma_1^2 + \sigma_2^2) f\left(z; \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}\right) \quad \text{(B.1)}$$

$$= k_c f(z; \mu_c, \sigma_c^2).$$

In general, the product of m such one-dimensional normal distributions over the same variable z is also a normal distribution. If the mean of the ith distribution is denoted by μ_i, and the variance by σ_i^2, then the probability density function of the product of m such normal distributions is provided in [89], and is given by

$$\prod_{i=1}^{m} f(z; \mu_i, \sigma_i^2) = c_p f\left(z; \frac{\sum_{i=1}^{m} \mu_i}{\sum_{i=1}^{m} \frac{1}{\sigma_i^2}}, \frac{1}{\sum_{i=1}^{m} \frac{1}{\sigma_i^2}}\right), \quad \text{(B.2)}$$

where $c_p = \dfrac{1}{\left(\prod_{i=1}^{m} \sigma_i\right) \sqrt{\sum_{i=1}^{m} \frac{1}{\sigma_i^2}}} \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{m} \frac{\mu_i^2}{\sigma_i^2} - \frac{\left(\sum_{i=1}^{m} \frac{\mu_i}{\sigma_i^2}\right)^2}{\sum_{i=1}^{m} \frac{1}{\sigma_i^2}}\right)\right)$ is a normalization constant.

We now describe the results for the product, and integral of the resulting product in the case of multivariate normal distributions.

B.1.2 Multivariate normal distributions

Consider two N-dimensional multivariate normal distributions in a single random variable, $\mathbf{z}$ with mean vectors $\boldsymbol{\mu}_1$, and $\boldsymbol{\mu}_2$; and covariance matrices, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$. Given that the covariance matrices are nonsingular, the product of these two N-dimensional multivariate normal distributions in the random variable $\mathbf{z}$ is another (unnormalized) N-dimensional multivariate normal distribution [2, 111] given by

$$f(\mathbf{z}; \boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}) \cdot f(\mathbf{z}; \boldsymbol{\mu_2}, \boldsymbol{\Sigma_2}) = k_p f(\mathbf{z}; \boldsymbol{\mu_p}, \boldsymbol{\Sigma}_p), \tag{B.3}$$

with

$$\boldsymbol{\Sigma}_p = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}, \quad \boldsymbol{\mu_p} = \boldsymbol{\Sigma}_p(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu_1} + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu_2}), \text{ and}$$

$$k_p = |2\pi\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_p^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_p\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})\right)$$

$$= |2\pi(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})^{\mathrm{T}}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})\right).$$

In the case of one-dimensional normal distributions, Eq. (B.3) reduces to Eq. (B.1) with k_p reducing to $k_c = f(\mu_1; \mu_2, \sigma_1^2 + \sigma_2^2)$.

Further, since $\int f(\mathbf{z}; \boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})d\mathbf{z} = 1$, the above results for the product of normal distributions implies that

N-dimensional: $\displaystyle\int_{\mathbb{R}^N} f(\mathbf{z}; \boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}) \cdot f(\mathbf{z}; \boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})\, d\mathbf{z} = \int_{\mathbb{R}^N} k_p f(\mathbf{z}; \boldsymbol{\mu_p}, \boldsymbol{\Sigma}_p)d\mathbf{z} = k_p$

one-dimensional: $\displaystyle\int_{\mathbb{R}} f(z; \mu_1, \sigma_1^2) \cdot f(z; \mu_2, \sigma_2^2)\, dz = \int_{\mathbb{R}} k_c f(z; \mu_c, \sigma_c^2)dz = k_c.$

$$\tag{B.4}$$

B.2 Determinant and inverse of a rank-1 matrix

The determinant of a rank-1 matrix can be computed using the following matrix determinant lemma [60, 24].

Lemma 1 (Matrix Determinant Lemma). *Suppose* $\mathbf{A}$ *is an invertible square matrix, and* $\mathbf{u}, \mathbf{v}$ *are column vectors. Then*

$$|\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}}| = (1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u})\,|\mathbf{A}|.$$

Here $\mathbf{u}\mathbf{v}^{\mathrm{T}}$ *is the outer product of two vectors u, and v.*

Further, the inverse of a rank-1 matrix can be obtained using the Sherman-Morrison formula [35, 54, 130, 9, 55, 98, 113]. This is a special case of the generalized Woodbury formula given in Theorem 2.

Theorem 1 (Sherman-Morrison formula). *Suppose* $\mathbf{A}$ *is an invertible square matrix, and* $\mathbf{u}, \mathbf{v}$ *are vectors. Assume that* $1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u} \neq 0$*. Then*

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}})^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}}.$$

Here $\mathbf{u}\mathbf{v}^{\mathrm{T}}$ *is the outer product of two vectors u, and v.*

B.3 Determinant and inverse of a rank-k matrix

We now present the relevant general theorems for computing the determinant, and inverse of a rank-k matrix.

Lemma 2 (Generalized Matrix Determinant Lemma)**.** *Suppose* $\mathbf{A}$ *is an invertible* $n \times n$ *matrix,* $\mathbf{U}, \mathbf{V}$ *are* $n \times m$ *matrices, and* $\mathbf{W}$ *is an invertible* $m \times m$ *matrix. Then*

$$|\mathbf{A} + \mathbf{U}\mathbf{W}\mathbf{V}^{\mathsf{T}}| = |\mathbf{W}^{-1} + \mathbf{V}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{U}||\mathbf{W}||\mathbf{A}|.$$

The inverse of a rank-k matrix can be computed using the following generalized Woodbury matrix identity [152, 55, 62].

Theorem 2 (Sherman-Morrison Woodbury formula or Woodbury matrix identity)**.** *Suppose* $\mathbf{A}$ *is an invertible* $n \times n$ *matrix,* $\mathbf{U}, \mathbf{W}$*, and* $\mathbf{V}$ *are* $n \times m$*,* $m \times m$*, and* $m \times n$ *matrices. Then*

$$(\mathbf{A} + \mathbf{U}\mathbf{W}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{W}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1}.$$

C

Appendix

Gabor Filter

A Gabor filter is a linear filter having frequency, and orientation similar to that of human visual system, and is obtained by multiplication of a sinusoidal wave with Gaussian kernel function [42, 100]. It can be used as a band-pass filter for unidimensional signals (e.g. speech). The filter has a real, and imaginary component given by

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right),$$

and

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{x'}{\lambda} + \psi\right).$$

Here $x' = x \cos\theta + y \sin\theta$, and $y' = -x \sin\theta + y \cos\theta$. In the above equations, λ represents the wavelength of the sinusoidal wave, θ is the angle of the normal to the parallel stripes of a Gabor function, ψ is the phase offset, σ is the standard deviation of the Gaussian envelope, and γ denotes the spatial aspect ratio, and describes the ellipticity of the support of the Gabor function.

Experimental studies have shown that simple cells in human visual system can be modeled by Gabor functions [30, 66]. Thus, Gabor patches are extensively used as stimuli in psychophysical studies. An example of a Gabor patch is shown in Figure 2.2(A).

Bibliography

[1] L.F. Abbott and P. Dayan. The effect of correlated variability on the accuracy of a population code. *Neural Computation*, 11(1):91–101, 1999.

[2] P. Ahrendt. The multivariate Gaussian probability distribution. Lecture notes, 2005.

[3] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on automatic control*, 19(6):716–723, 1974.

[4] B.B. Averbeck. *Noise correlations and information encoding and decoding*. Springer, 2009.

[5] B.B. Averbeck and D. Lee. Effects of noise correlations on information encoding and decoding. *Journal of Neurophysiology*, 95(6):3633–3644, 2006.

[6] B.B. Averbeck, P.E.Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Neuroscience Reviews*, 7(5):358–366, 2006.

[7] S. Baldassi and D.C. Burr. Feature-based integration of orientation signals in visual search. *Vision research*, 40(10):1293–1300, 2000.

[8] S. Baldassi and P. Verghese. Comparing integration rules in visual search. *Journal of Vision*, 2(8), 2002.

[9] M.S. Bartlett. An inverse matrix adjustment arising in discriminant analysis. *The Annals of Mathematical Statistics*, 22(1):107–111, 1951.

[10] P.W. Battaglia, R.A. Jacobs, and R.N. Aslin. Bayesian integration of visual and auditory signals for spatial localization. *JOSA A*, 20(7):1391–1397, 2003.

[11] W.F. Bauer. The Monte Carlo method. *Journal of the Society for Industrial & Applied Mathematics*, 6(4):438–451, 1958.

[12] T. Bayes and R. Price. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions (1683-1775)*, pages 370–418, 1763.

[13] J. Beck, J. Kanitscheider, X. Pitkow, P. Latham, and A. Pouget. The perils of inferring information from correlations. In *Cosyne Abstracts 2013, Salt Lake City USA*, 2013.

[14] J.M. Beck, W.J. Ma, X. Pitkow, P.E. Latham, and A. Pouget. Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, 74(1):30–39, 2012.

[15] B.A. Berg. *Markov chain Monte Carlo simulations and their statistical analysis: with web-based Fortran code*. World Scientific Publishers, Singapore, 2004.

[16] M. Bhardwaj, R. van den Berg, W.J. Ma, and Krešimir Josić. Do humans account for stimulus correlations in visual perception? In *Cosyne Abstracts 2013, Salt Lake City, USA*, 2013.

[17] H.S. Bhat and N. Kumar. On the derivation of the Bayesian information criterion. *School report, University of California*, 2010.

[18] C.M. Bishop and N.M. Nasrabadi. *Pattern recognition and machine learning*, volume 1. Springer, New York, 2006.

[19] G.E.P. Box. Evolutionary operation: A method for increasing industrial productivity. *Applied Statistics*, pages 81–101, 1957.

[20] H. Bozdogan. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

[21] H. Bozdogan. Akaike's information criterion and recent developments in information complexity. *Journal of mathematical psychology*, 44(1):62–91, 2000.

[22] T.F. Brady and J.B. Tenenbaum. Encoding higher-order structure in visual working memory: A probabilistic model. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pages 411–416, 2010.

[23] D.H. Brainard. The psychophysics toolbox. *Spatial vision*, 10(4):433–436, 1997.

[24] M. Brookes. The matrix reference manual (online), 2005.

[25] K.P. Burnham and D.R. Anderson. Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2):261–304, 2004.

[26] S. Carroll, M. Bhardwaj, W.J. Ma, and K. Josić. Visual decisions in presence of external and internal correlations. *in preparation*, 2013.

[27] Y. Chen, W.S. Geisler, and E. Seidemann. Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature neuroscience*, 9(11):1412–1420, 2006.

[28] M.R. Cohen and A. Kohn. Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819, 2011.

[29] C.A. Colin and F.A.G. Windmeijer. An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2):329–342, 1997.

[30] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Optical Society of America, Journal, A: Optics and Image Science*, 2(7):1160–1169, 1985.

[31] P. Dayan and L. Abbott. Theoretical neuroscience. *MIT Press*, 15:460, 2001.

[32] P. Dayan, G.E. Hinton, R.M. Neal, and R.S. Zemel. The Helmholtz machine. *Neural computation*, 7(5):889–904, 1995.

[33] J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A. Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806, 2007.

[34] J. Duncan and G.W. Humphreys. Visual search and stimulus similarity. *Psychological review*, 96(3):433, 1989.

[35] W.J. Duncan. Some devices for the solution of large sets of simultaneous linear equations (with an appendix on the reciprocation of partitioned matrices). *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Seventh Series*, 35:660–670, 1944.

[36] A.S. Ecker, P. Berens, G.A. Keliris, M. Bethge, N.K. Logothetis, and A.S. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–587, 2010.

[37] A.S. Ecker, P. Berens, A.S. Tolias, and M. Bethge. The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience*, 31(40):14272–14283, 2011.

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

[39] M.P. Eckstein, J.P. Thomas, J. Palmer, and S.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, 2000.

[40] M.O. Ernst and M.S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.

[41] W.K. Estes and H.A. Taylor. A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences of the United States of America*, 52(2):446, 1964.

[42] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biological cybernetics*, 61(2):103–113, 1989.

[43] D. Fougnie, J.W. Suchow, and G.A. Alvarez. Variability in the quality of visual working memory. *Nature communications*, 3:1229, 2012.

[44] F. Gabbiani and S.J. Cox. *Mathematics for neuroscientists*. Academic Press, 2010.

[45] E. Ganmor, R. Segev, and E. Schneidman. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proceedings of the National Academy of Scienes*, 108(23):9679–9684, 2011.

[46] T.J. Gawne and B.J. Richmond. How independent are the messages carried by adjacent inferior temporal cortical neurons? *The Journal of Neuroscience*, 13(7):2758–2771, 1993.

[47] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian data analysis*. Chapman & Hall, CRC press, 2003.

[48] G.A. Gescheider. *Psychophysics: the fundamentals*. Psychology Press, 1997.

[49] D.E. Goldberg and J.H. Holland. Genetic algorithms and machine learning. *Machine learning*, 3(2):95–99, 1988.

[50] D. Goldreich. A Bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions. *PLoS One*, 2(3):e333, 2007.

[51] D. Goldreich and M.A. Peterson. A Bayesian observer replicates convexity context effects in figure-ground perception. *Seeing and Perceiving*, 25(3-4):3–4, 2012.

[52] N. Graham, P. Kramer, and D. Yager. Signal-detection models for multidimensional stimuli: Probability distributions and combination rules. *Journal of Mathematical Psychology*, 31(4):366–409, 1987.

[53] D.M. Green and J.A. Swets. *Signal detection theory and psychophysics*, volume 1. John Wiley & sons, New York, 1966.

[54] L. Guttman. Enlargement methods for computing the inverse matrix. *The annals of mathematical statistics*, 17(3):336–343, 1946.

[55] W.H. Hager. Updating the inverse of a matrix. *SIAM review*, 31(2):221–239, 1989.

[56] J.H. Halton. A retrospective and prospective survey of the Monte Carlo method. *SIAM review*, 12(1):1–63, 1970.

[57] W. Hamilton. *Lectures on metaphysics and logic*, volume 1. Gould and Lincoln, 1859.

[58] J.M. Hammersley, D.C. Handscomb, and G. Weiss. Monte Carlo methods. *Physics today*, 18:55, 1965.

[59] B.J. Hansen, M.I. Chelaru, and V. Dragoi. Correlated variability in laminar cortical circuits. *Neuron*, 76:590–602, 2012.

[60] D.A. Harville. Matrix algebra from a statistician's perspective. *Technometrics*, 40(2):164–164, 1998.

[61] G. Hatfield. Perception as unconscious inference. *Perception and the physical world: Psychological and philosophical issues in perception*, pages 113–143, 2002.

[62] N.J. Higham. *Accuracy and Stability of Numberical Algorithms*, volume 48. SIAM, 1996.

[63] R. Hooke and T.A. Jeeves. Direct search solution of numerical and statistical problems. *Journal of the ACM (JACM)*, 8(2):212–229, 1961.

[64] R. A. Jacobs. Optimal integration of texture and motion cues to depth. *Vision research*, 39(21):3621–3629, 1999.

[65] E.T. Jaynes. *How does the brain do plausible reasoning?* Springer, 1988.

[66] J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

[67] K. Josić, E. Shea-Brown, B. Doiron, and J. de la Rocha. Stimulus-dependent correlations and population codes. *Neural computation*, 21(10):2774–2804, 2009.

[68] M.H. Kalos and P.A. Whitlock. *Monte Carlo methods*. John Wiley & Sons, 2008.

[69] S.M. Kay. *Fundamentals of statistical signal processing, Detection Theory, volume 2*. Prentice Hall PTR, New Jersey, USA, 1998.

[70] D. Kersten, P. Mamassian, and A. Yuille. Object perception as Bayesian inference. *Annual Reiew of Psychology*, 55:271–304, 2004.

[71] S. Keshvari, R. van den Berg, and W.J. Ma. Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE*, 7(6):e40216, 2012.

[72] S. Keshvari, R. van den Berg, and W.J. Ma. No evidence for an item limit in change detection. *PLoS computational biology*, 9(2):e1002927, 2013.

[73] S.A. Klein. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & psychophysics*, 63(8):1421–1455, 2001.

[74] D.C. Knill. Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception. *Journal of Vision*, 5(2), 2005.

[75] D.C. Knill and A. Pouget. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719, 2004.

[76] D.C. Knill and W. Richards. *Perception as Bayesian inference*. Cambridge University Press, 1996.

[77] A. Kohn and M.A. Smith. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *The Journal of Neuroscience*, 25(14):3661–3673, 2005.

[78] K. Körding. Decision theory: what should the nervous system do? *Science*, 318(5850):606–610, 2007.

[79] K.P. Körding, S. Ku, and D.M. Wolpert. Bayesian integration in force estimation. *Journal of Neurophysiology*, 92(5):3161–3165, 2004.

[80] K.P. Körding and D.M. Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247, 2004.

[81] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[82] P.E. Latham and S. Nirenberg. Synergy, redundancy, and independence in population codes, revisited. *The Journal of neuroscience*, 25(21):5195–5206, 2005.

[83] P.E. Latham and Y. Roudi. Role of correlations in population coding. *arXiv preprint:1109.6524*, 2011.

[84] P.M. Lee. *Bayesian statistics: an introduction*. John Wiley & Sons, 2012.

[85] R. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9(4):1082–1099, 1999.

[86] W.J. Ma. Bayesian approach to perception. *SAGE Encyclopedia of Perception*, pages 201–205, 2009.

[87] W.J. Ma. Signal detection theory, uncertainty, and Poisson-like population codes. *Vision research*, 50(22):2308–2319, 2010.

[88] W.J. Ma. Theoretical neuroscience-learning, perception and cognition. Course Lecture notes, Spring 2011.

[89] W.J. Ma, K. Körding, and D. Goldreich. Bayesian modeling of perception. unpublished, 2013.

[90] W.J. Ma, V. Navalpakkam, J. M. Beck, R. van den Berg, and A. Pouget. Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, 14(6):783–790, 2011.

[91] D.J.C. MacKay. Bayesian interpolation. *Neural computation*, 4(3):415–447, 1992.

[92] D.J.C. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[93] N.A. Macmillan and D.C. Creelman. *Detection theory: A user's guide*. Psychology press, 2004.

[94] H. Mazyar, R. van den Berg, and W.J. Ma. Does precision decrease with set size? *Journal of Vision*, 12(6), 2012.

[95] H. Mazyar, R. van den Berg, R.L. Seilheimer, and W.J. Ma. Independence is elusive:set size effects on encoding precision in visual search. *Journal of Vision*, 13(5):1–14, 2013.

[96] J.L. McClelland. Connectionist models and bayesian inference. *Rational models of cognition*, pages 21–53, 1998.

[97] D. McNicol. *A primer of signal detection theory*. Psychology Press, 2005.

[98] C.D. Meyer. *Matrix analysis and applied linear algebra book and solutions manual*, volume 2. SIAM, 2000.

[99] M. Mitchell and S. Forrest. Genetic algorithms and artificial life. *Artificial Life*, 1(3):267–289, 1994.

[100] J.R. Movellan. Tutorial on gabor filters, 2008.

[101] A.L. Nagy and G. Thomas. Distractor heterogeneity, attention, and color in visual search. *Vision research*, 43(14):1541–1552, 2003.

[102] V. Navalpakkam and L. Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2049–2056. IEEE, 2006.

[103] R.M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. *Technical report*, 1993.

[104] L.W. Nolte and D. Jaarsma. More on the detection of one of m orthogonal signals. *The Journal of the Acoustical Society of America*, 41:497, 1967.

[105] I.F. Ohiorhenuan, F. Mechler, K.P. Purpura, A.M. Schmid, Q. Hu, and J.D. Victor. Sparse coding and high-order correlations in fine-scale cortical network. *Nature*, 466:617–621, 2010.

[106] J. Palmer. Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2):332, 1990.

[107] J. Palmer. Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision research*, 34(13):1703–1721, 1994.

[108] J. Palmer, P. Verghese, and M. Pavel. The psychophysics of visual search. *Vision research*, 40(10):1227–1268, 2000.

[109] D.G. Pelli. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4):437–442, 1997.

[110] D.H. Perkel, G.L. Gerstein, and G.P. Moore. Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains. *Biophysical journal*, 7(4):419–440, 1967.

[111] K.B. Petersen and M.S. Pedersen. The matrix cookbook. Lecture notes, 2008.

[112] W.W. Peterson, T.G. Birdsall, and W. Fox. The theory of signal detectability. *Information Theory, IRE Professional Group on*, 4(4):171–212, 1954.

[113] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press, 2007.

[114] R.P.N. Rao, B.A. Olshausen, and M.S. Lewicki. *Probabilistic models of the brain: Perception and neural function*. MIT Press, 2002.

[115] C.P. Robert and G. Casella. *Monte Carlo statistical methods*, volume 319. Springer-Verlag, 2004.

[116] R. Romo, A. Hernandez, A. Zainos, and E. Salinas. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron*, 38(4):649–657, 2003.

[117] R. Rosenbaum and K. Josić. Mechanisms that modulate the transfer of spiking correlations. *Neural computation*, 23(5):1261–1305, 2011.

[118] R. Rosenbaum, J. Trousdale, and Krešimir Josić. Pooling and correlated neural activity. *Frontiers in Computational Neuroscience*, 4(9), 2010.

[119] R. Rosenholtz. Visual search for orientation among heterogeneous distractors: experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4):985, 2001.

[120] R.Y. Rubinstein and D.P. Kroese. *Simulation and the Monte Carlo method*. Wiley Series in Probability and Mathematical Statistics, New York, 1981.

[121] L.L. Scharf. *Statistical signal processing*, volume 98. Addison-Wesley Reading, 1991.

[122] E. Schneidman, W. Bialek, and M.J. Berry. Synergy, redundancy, and independence in population codes. *The Journal of Neuroscience*, 23(37):11539–11553, 2003.

[123] G.E. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[124] P. Seriès, P.E. Latham, and A. Pouget. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nature Neuroscience*, 7(10):1129–1135, 2004.

[125] M. Shamir and H. Sompolinsky. Correlation codes in neuronal populations. *Advances in neural information processing systems*, 1:277–284, 2002.

[126] L. Shams, W.J. Ma, and U. Beierholm. Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16(17):1923–1927, 2005.

[127] M.L. Shaw. Identifying attentional and decision-making components in information processing. *Attention and performance VIII*, 8:277–295, 1980.

[128] M.L. Shaw. Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. *Attention and performance X*, pages 109–121, 1984.

[129] S. Shen, R. van den Berg, and W.J. Ma. When is sensory precision variable? In *Cosyne Abstracts 2013, Salt Lake City, USA*, 2013.

[130] J. Sherman and W.J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.

[131] H. Shin, R. van den Berg, and W.J. Ma. Independent pools of visual short-term memory resource for different features. In *Cosyne Abstracts 2013, Salt Lake City, USA*, 2013.

[132] M.A. Smith. *Alhacen's Theory of Visual Perception: A Critical Edition, with English Translation and Commentary, of the First Three Books of Alhacen's De Aspectibus, the Medieval Latin Version of Ibn Al-Haytham's Kitāb Al-Manazir*, volume 1. American Philosophical Society, 2001.

[133] H. Sompolinsky, H. Yoon, K. Kang, and M. Shamir. Population coding in neuronal systems with correlated noise. *Physical Review E*, 64(5):051904, 2001.

[134] D. L. Swets and B. Punch. Genetic algorithms for object localization in a complex scene. In *IEEE International conference on image processing*, pages 595–598. IEEE, 1995.

[135] H. Tassinari, T.E. Hudson, and M.S. Landy. Combining priors and noisy visual cues in a rapid pointing task. *The Journal of neuroscience*, 26(40):10154–10163, 2006.

[136] G. Tkačik, J.S. Prentice, V. Balasubramanian, and E. Schneidman. Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences*, 107(32):14419–14424, 2010.

[137] V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on optimization*, 7(1):1–25, 1997.

[138] A. Treisman. Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2):194, 1982.

[139] A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[140] R. van den Berg and W.J. Ma. Factorial comparison of working memory models. unpublished, 2013.

[141] R. van den Berg, H. Shin, W.C. Chou, R. George, and W.J. Ma. Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109(22):8780–8785, 2012.

[142] R. van den Berg, M. Vogel, K. Josić, and W.J. Ma. Optimal inference of sameness. *Proceedings of the National Academy of Sciences*, 109(8):3178–83, 2012.

[143] P. Verghese. Visual search and attention: A signal detection theory approach. *Neuron*, 31(4):523–535, 2001.

[144] B. T. Vincent, R. J. Baddeley, T. Troscianko, and I.D. Gilchrist. Optimal feature integration in visual search. *Journal of Vision*, 9(5), 2009.

[145] H. von Helmholtz. Handbuch der Physiologischen Optik, vol. II. *In G. Karsten (Ed.) Allgemeine Encyklopddie der Physik*, 9:37–51, 1867.

[146] L. Wasserman. Bayesian model selection and model averaging. *Journal of mathematical psychology*, 44(1):92–107, 2000.

[147] F.A. Wichmann and N.J. Hill. The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & psychophysics*, 63(8):1293–1313, 2001.

[148] T.D. Wickens. Elementary signal detection theory. *New York, Oxford university press*, 2002.

[149] P. Wilken and W.J. Ma. A detection theory account of change detection. *Journal of Vision*, 4(12), 2004.

[150] R.L. Winkler. *An introduction to Bayesian inference and decision*. Holt, Rinehart and Winston New York, 1972.

[151] J.M. Wolfe, S.R. Friedman-Hill, M.I. Stewart, and K.M. O'Connell. The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1):34, 1992.

[152] M.A. Woodbury. Inverting modified matrices. *Memorandum report*, 42:106, 1950.

[153] W. Zhang and S.J. Luck. Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192):233–235, 2008.