The Dissertation Committee for Can Oluk Certifies that this is the approved version of the following Dissertation:

Visual Target Detection Under Multiple Dimensions of Uncertainty

Committee:

Wilson S. Geisler, Supervisor

Lawrence K. Cormack

Robbe L. Goris

Alan Bovik

Visual Target Detection Under Multiple Dimensions of Uncertainty

by

Can Oluk

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2022

To my parents, Lütfiye and Nevzat Oluk.

Acknowledgements

I would like to thank my committee - Lawrence Cormack, Robbe Goris, and Alan Bovik – for providing me support and guidance throughout my graduate studies. I am incredibly grateful to my supervisor, Wilson Geisler, for teaching me a lot about vision science and scientific practice in general. He always found time to answer my questions and listen to my ideas. Also, I am thankful to everyone in our lab for their friendship and valuable feedback.

I am lucky to have many close friends who shared my excitement and frustrations, even if we were miles apart. Their friendship and support made my graduate experience more much enjoyable. Lastly, I am grateful to my wife - Begum Ozdemir Oluk- for encouraging and supporting me in pursuing my ideals. Without her, I would not even feel the same enthusiasm for a graduate degree.

Abstract

Visual Target Detection Under Multiple Dimensions of Uncertainty

Can Oluk, Ph.D.

The University of Texas at Austin, 2022

Supervisor: Wilson S. Geisler

Detection of visual targets is integral to survival and everyday functioning. In the real world, the visual system operates under very high levels of extrinsic uncertainty (multiple simultaneous dimensions of uncertainty) about target and background properties. However, the previous literature is primarily concerned with the effect of low and modest levels of uncertainty (only a single dimension of uncertainty) on human performance. This thesis aims to measure and model human performance under multiple simultaneous dimensions of uncertainty. I primarily focus on the detection of additive targets in white noise.

First, human performance was measured under simultaneous target amplitude, and background contrast uncertainty for target prior probabilities of 0.5 and 0.2. I simulated and tested three model observers. The ideal observer, which has a dynamic decision criterion, can be approximated with a contrast normalized template-matching (NTM) observer with a single criterion. Maximum-likelihood fits revealed that the NTM and the dynamic-decision-criterion observer predict human performance much better than a

template-matching (TM) observer with a single criterion. The same results were also found for natural backgrounds. The results reveal the value contrast normalization under real-world conditions where target priors are low.

Secondly, human performance was measured under simultaneous target scale and target orientation uncertainty. I also describe an efficient simulation method. Simulations revealed that the maximum-template-response (MAX) observer only approximates the ideal observer when template responses are normalized by the energy of templates (ENM observer). Maximum likelihood fits reveal the ENM and the ideal observer explain human performance much better than the simple MAX observer. Furthermore, human performance was also measured under low uncertainty. I found that humans are more efficient under high uncertainty. The efficiency difference can be accounted for by incorporating intrinsic position uncertainty into the model observers. The results reveal the value of energy normalization and the importance of intrinsic uncertainties for understanding the visual detection under various levels of uncertainty.

This thesis provided insights into visual processing under uncertainty. It exemplifies the fruitfulness of studying detection under simultaneous multiple dimensions of uncertainty, within a principled framework. The methods developed in this this should be useful in future experiments.

Table of Contents

List of Figures

Chapter 1: Introduction

Almost every day, we, humans, spend considerable time searching for keys, earphones, and various other objects. If one is looking in a direction near to the object of interest, detecting it is effortless most of the time. However, it has been a challenge for scientists to understand how the human visual system can match the patterns of light that fall into the retina with the object of interest. The difficulty resides in the fact that substantially different visual patterns (the light intensity patterns) correspond to the same object. There is substantial variation (i) in the properties of the target (like its orientation, 3D location) and (ii) in the properties of the background around the object. Because of the variation, the observer is uncertain about the properties of the target and background prior to detection and thus must pay attention to all possible visual patterns. The aim of the thesis is to better understand how the human visual system detects objects while being uncertain about the properties of the object and its background (i.e., being uncertain about the exact visual pattern to look for). In this chapter, I will discuss my conceptual approach to the problem. In the next chapter, I will review the literature and specify a generic visual detection task covering an extensive sub-set of visual pattern variations that naturally occur in the real world. In the rest of the thesis, I will focus on building computational models of visual processing for this generic task. Moreover, I will measure human performance under high levels of uncertainty and test these computational models against the data.

The behavior of complicated biological systems results from a long chain of processing which is frequently broken down into sensation, perception, cognition, and decision. The measured behavior is challenging to understand because, most of the time, it is possible to produce equivalent explanations by focusing on different stages of the processing. So, most of the time, many possible explanations and alternative computational

models explain the measured behavior equally well. To limit the number of possible explanations, introducing additional constraints based on normative principles is a fruitful approach. Normative principles describe the goals of the biological system (cost and benefits). Additionally, they have the potential to be general principles because the same normative principles can be applied to a wide range of behaviors. Thus, they can be effective for dealing with the complexity of biological systems. Therefore, developing normatively-motivated explanations and computational models is a promising approach for understanding visual processing, given the complexity of biological vision systems.

I will describe and use a normative approach based on the principle of natural selection. The visual system is at least partly designed by natural selection. Therefore, behaviors (and computations underlying them) were selected based on their contribution to the survival and reproduction of the animal (for example, see Geisler & Diehl, 2002; Geisler & Diehl, 2003). Consequently, the visual processing devoted to an important task for the overall survivability (even though success in a single task does not imply survival) should be selected based on its success in performing the task. It is almost impossible to know which tasks were most relevant to the survival and reproduction of the animal through evolutionary history. However, a task that the animal face in its natural environment (naturalistic tasks) is an informed guess of what was relevant. Therefore, for a naturalistic task, the normative approach favors the explanations (or computational models) motivated to maximize the success (motivated by the best performing models) when many explanations are equally good in explaining the measured behavior. This additional constraint provides an excellent and principled starting point for the analysis: the analysis of the optimal performance (the maximum success one could achieve). In sum, the principle of natural selection (normative approach) suggests that, given a naturalistic task, starting from the ideal observer (whose performance defines the optimal performance)

is a good starting point in the search models and explanations of the measured behavior. Furthermore, principled sub-optimal models can be obtained by replacing parts of the ideal observer with sub-optimal / heuristic computations.

1.1 Naturalistic Tasks

The normative approach starts with finding a naturalistic task. The more naturalistic the task, the more likely it is to be relevant for survival throughout evolutionary history. Thus, for more naturalistic tasks, the approach is more likely to yield new insights about how the visual system operates (Geisler & Ringach, 2009). I will first review how studying naturalistic tasks was fruitful even without quantifying the optimal performance (maximizing the success). Moreover, I will discuss why it is appropriate for research on visual detection under uncertainty to move toward more naturalistic tasks and why my thesis exemplifies a step in this direction.

Multiple components of a task could be naturalistic: input (stimuli), output (behavior), and task structure. Previous research that focused on the naturalistic version of each component successfully led to new findings and insights. For example, there have been studies focused on naturalistic inputs. The visual system evolved to operate with natural scenes, so using natural scenes as the input has revealed (a) previously undiscovered properties of primary visual cortex neurons (David et al., 2004; Felsen et al.,2005; Touryan et al., 2005) and (b) lawful relations between behavior and properties of natural scenes (Sebastian et al., 2017). Furthermore, asking how natural scenes are best processed to satisfy general principles like sparse coding (Olshausen & Field, 1996), encoding of independent features (Bell & Sejnowski, 1997), or maximizing the accuracy in a specific task (Burge & Geisler, 2011; Burge & Jaini, 2017) has produced results that resemble the visual processing either in neural or behavioral level suggesting that the visual

system might have evolved to satisfy these principles. However, working with natural scenes make it essential to characterize their properties carefully, and it is important to remember the advantages of completely controlled stimuli for testing hypotheses (Rust & Movshon, 2005; Martinez-Garcia et al., 2019). In my thesis, I used natural scene backgrounds as input to test whether the results I found with white noise backgrounds still hold for natural scene backgrounds in Chapter 2.

Some studies of natural tasks are primarily focused on naturalistic outputs (behaviors). In laboratory conditions, the range of behavior allowed by the experimenter is generally very constrained, but the behavior selected by natural selection is much more complex and richer. Therefore, measuring more realistic behavior is more likely to provide insights into how and why the visual system operates in a certain manner. Indeed, measuring naturalistic behavior has provided important new insights into visual processing (Greenwald & Knill, 2009; Hayhoe et al., 2009: Land, 2009; for review, see Hayhoe, 2017).

Lastly, task structure can be naturalistic. For example, generally, the experimenter arbitrarily decides the prior probability of visual features presented in the experiment. However, the visual system evolved to exploit the prior probabilities of the same visual features in the natural environment (naturalistic priors and also known as natural scene statistics), which are likely to be learned through evolutionary history. Thus, the visual system might only partially exploit a specific prior that the experimenter arbitrarily sets. Therefore, the measured behavior might seem reasonable under the assumption that natural priors are being used to process it instead of priors that the experimenter sets (for a review, see Geisler, 2008). In my thesis, I will also use a more naturalistic structure (design of the experiment) of the visual detection task by introducing simultaneous multiple dimensions of uncertainty. Previous literature on visual detection under uncertainty (that will be covered in Chapter 2) generally considers only a single dimension at a time. However,

multiple dimensions of uncertainty are always present in the real world. When properties of target and background are exactly specified in an experiment, the visual system may not use that information appropriately because it evolved with multidimensional uncertainty. Principles found in these naturalistic conditions might help us understand unexplained or odd behaviors in simpler conditions (with low uncertainty). The vast literature on simpler conditions with only a low or modest extrinsic uncertainty makes this literature a good target for pushing toward more naturalistic tasks. Overall, the evidence covered here also shows that research on more naturalistic tasks is indeed fruitful. Therefore, detection behavior under simultaneous multiple dimensions of uncertainty might yield important insights about how the visual system operates in the real world and help us better understand simpler conditions (with low uncertainty). Moreover, having a naturalistic task in hand makes it more likely that the task was relevant for survival. Thus, it is more likely that performance in the task is maximized due to natural selection, so starting with the ideal observer is even more appropriate for the task.

1.2 The Ideal Observer Analysis

The next step for the normative approach is to derive the best possible performance, optimal performance (limited only by the information available), since natural selection is expected to maximize success (at least partly) in a naturalistic task. The ideal observer analysis aims to describe how to make the best possible decisions (minimizing cost, maximizing benefit), thus producing the best performance possible. First, it is essential to define (i) the properties of the environment (the possible world states and their mechanics), (ii) the information (sensory input) provided to the agent, and (iii) what counts as success (cost function for the possible outcomes). For my thesis, I will cast the problem within a probabilistic Bayesian framework, as commonly done. In the Bayesian framework, the

posterior probability distribution quantifies the probability of world states given the sensory input using the prior distribution of world states and the likelihood of the sensory input given a world state (Bayes Rule). The behavior is decided based on the posterior distribution and the cost function, which quantifies the benefits of various possible outcomes. The Bayesian framework is quite general, making it possible to conceptualize most problems. Apart from the ideal observer being a good starting point for modeling a naturalistic task, it also provides a benchmark for performance. Deriving and simulating the ideal observer reveals essential computational principles for the task. Moreover, it has been shown that the ideal observer explains human performance in various tasks and provides important insight about visual computations (for reviews, see Geisler, 2003; Geisler, 2011; Burge, 2020).

The previous literature on visual detection under uncertainty also frequently utilizes ideal-observer analysis (the literature is reviewed in Chapter 2). It has been shown that the ideal observer accounts for the measured human performance under uncertainty, and it explains the relationship between various levels of uncertainty. However, as the task becomes more naturalistic, derivation and simulation of the ideal observer become more difficult. My thesis deals with tasks that are complicated enough to result in new findings but still simple enough to derive and simulate the ideal observer. Overall, the ideal observer analysis is a fruitful framework that consistently leads to important insights about visual processing under uncertainty and is likely to be important for understanding how the visual system functions under uncertainty.

In sum, the normative approach based on the principle of natural selection has been useful in studying the visual system (both using naturalistic tasks and ideal observers were useful). Moreover, the research on visual detection under uncertainty is also appropriate for such a normative approach. Therefore, in the following chapters, I will consider

naturalistic tasks that involve high levels of simultaneous multiple dimensions of uncertainty and start the modeling with the derivation of the ideal observer.

Chapter 2: Visual Detection Under Uncertainty

2.1 Visual Detection

Throughout evolutionary history, finding specific objects was essential to survival. For example, finding food, noticing predators, and finding a potential mate are crucial tasks that include the sub-task of finding specific objects. Moreover, the same sub-task remains relevant for most everyday tasks because we are involved in many tasks requiring us to find specific objects (keys, earphones, pens, etc.).

Like most animals, humans rely heavily on visual information to search for a specific object. Light reflected from surfaces first passes through the optics of the eye. When it reaches the back of the eye (retina), the intensity of light in each location is encoded by the photoreceptors, which translate the intensity into electrical and chemical signals. These signals leave the eye through the optic nerve. The optic nerve carries signals to the lateral geniculate nucleus (LGN) in the thalamus. Most of the signals are then carried to the occipital lobe, where the primary visual cortex is located. A substantial amount of the cortex is devoted to visual processing.

The visual detection task is commonly defined as deciding whether the object of interest (target) is present or absent in the image that falls into the retina at a given moment. The challenging part of visual detection is the existence of substantial variation in the visual patterns classified as the target being present. The main source of variation is external (i.e., in the stimulus). This variation makes observers uncertain about what visual pattern to look for; thus, the variation is called extrinsic uncertainty. Extrinsic uncertainties are better conceptualized by considering two types because any visual pattern consists of a target (if the target is present) and a background (always present). The first type of uncertainty is category uncertainty (a category is the set of targets) which occurs because it is impossible

to know which specific target to look for. There are two reasons for the substantial category uncertainty in the real world. (i) Many objects fall into the same category (for example, the food category includes apples, bananas, etc.) (ii) The same specific target object often appears at different 3D locations and orientations under different lighting conditions. The second type of uncertainty is context uncertainty (context is the set of backgrounds). It is driven by the fact that the background properties are also highly variable. All these dimensions of extrinsic target and background variability make the detection of specific objects difficult under natural conditions.

There are also intrinsic uncertainties (the source of these uncertainties is intrinsic). Even if observers are clearly informed about the visual patterns that could possibly be presented in a certain situation, they will still be uncertain about the properties of targets and backgrounds to some extent. The first reason is that the observer has only limited ability to memorize and represent exact specific visual patterns that are needed to be searched. Secondly, there is internal noise in the representation of the eye, head position, etc. The observer's relation with the environment determines the visual pattern to be detected in the cortex (not just external variability). When the relation is not exactly known, the observer also needs to search for various visual patterns covering the range of potential visual patterns that might fall into the retina due to the different eye and head positions. Thus, the observer is intrinsically uncertain about the visual pattern. For these two reasons, observers generally search for small variations around the specific visual pattern they are looking for and are likely to decide that the target is present even when they detect a pattern slightly different from what they are looking for.

In the rest of this chapter, I will first review how human performance is measured and modeled in the literature so far under various levels of uncertainty. Then, I will discuss methodical challenges for studying more natural conditions that include high levels of

uncertainty. Finally, I will introduce a generic task, which is the focus of my thesis, and layout a plan for studying it.

2.2 Measurement of Human Detection Performance under Low Uncertainty

Under natural conditions, the visual detection task includes a vast amount of extrinsic target and background uncertainty along multiple stimulus dimensions. To be precise, I define any variation in stimuli (either in target or background) across trials as an extrinsic uncertainty. Unlike the natural conditions, in the laboratory environment, experiments are often conducted under strictly controlled conditions with relatively little extrinsic uncertainty. My experiments and most of the literature I cover in this Chapter are similar to a canonical laboratory visual detection task. This task involves an additive windowed sine-wave target with fixed amplitude (amplitude determines the signal strength). Generally, the target is presented at the center of the screen. The target appears only in half of the trials, and in each trial, participants are asked to report whether the target is absent or present. If the target is presented against a uniform background, there is nothing varying across trials except the target presence (basic detection task). This is the condition where there is no extrinsic uncertainty at all. However, there are many experiments in which the target is presented against a white noise background, generated by independently sampling each pixel's luminance from a gaussian distribution with a fixed standard deviation. In most of these experiments, a new sample of white noise is sampled and presented as a background in each trial. The variation across trials generates an extrinsic uncertainty which is due to a random sampling of noise backgrounds.

Note that random sampling of noise backgrounds generates millions of different backgrounds that could be presented in a trial. It is possible to conceptualize the same uncertainty with other dimensions. For example, there will be small variations in contrast

and mean luminance of the background, so there will be uncertainty about these dimensions. This example illustrates that any dimensions of uncertainty can be conceptualized with different variables. To be precise, I will define the dimension of uncertainty as the dimension which is varied to generate the variation in visual patterns across trials in the first place. Another conceptual issue is quantifying the amount of uncertainty because there are also multiple ways of doing it. Here I will only roughly quantify the amount of uncertainty by comparing the variation dimension generated to the totality of the extrinsic uncertainty present in the real world. For example, even though a random sampling of noise backgrounds generates millions of different backgrounds, many other dimensions of target and backgrounds in the real world vary across fixations (trials). Thus, the variation introduced by random sampling of noise backgrounds only covers a tiny fraction of variability in the real world. Therefore, the random sampling of noise backgrounds alone will be classified as low uncertainty.

Under low uncertainty, data collected in these canonical experiments consists of hit and correct rejection rates (which compose the overall percentage correct). The participant's response in each trial falls into one of four categories. When the target is present, and if the participant reports the target is present, it is a hit (if the participant reports the target is absent, it is a miss). When the target is absent, and if the participant reports the target is absent, it is a correct rejection (if the participant reports the target is present, it is a false alarm). To measure the sensitivity of human observers independent of any response bias (being too liberal or conservative in their decisions), hit and correct rejection rates (these two determine all rates) are transformed into two other independent parameters. Signal detection theory provides a method for transforming hit and correct rejection rates to criterion and d-prime (Green & Swets, 1966) with relatively robust assumptions (e.g., the internal decision variable of the participant is Gaussian distributed). D-prime (d')

quantifies the sensitivity independent of the criterion. It is the number of standard deviations of separation between the distribution of the decision variable with target present and absent. The location of the criterion accounts for any response bias.

So far, a single block of experiment has been discussed, but most of the experiments consist of multiple blocks. Properties of target and background are often varied across blocks (not across trials). However, no further extrinsic uncertainty is introduced as long as participants have prior information on which block they will be running (and the properties of the target and background associated with the block) because they will only concentrate on the relevant properties for that block. It is possible to measure how d-prime changes as a function of some variable across blocks. For example, one can vary the signal strength (target amplitude) to measure a psychometric function. It is possible to summarize the psychometric with a more stable parameter: threshold. Thresholds are extracted by fitting a function to d-primes. The threshold is generally defined as the value of the parameter (that is varied across blocks) that gives the d-prime of 1.0 (a d-prime of 1.0 is arbitrary; any level of performance can be defined as a threshold). Furthermore, a threshold can be measured for each level of a variable (thresholds as a function of the variable), where individual thresholds are measured by varying relatively simple variables like target amplitude.

Human observers are usually compared to (modeled with) the Bayesian ideal observer for this canonical task. The ideal observer, which quantifies the maximum achievable performance (optimal performance), can be described by the process of template matching (Peterson et al., 1954; Green & Swets, 1966; Burgess et al., 1981) in white noise backgrounds. The ideal observer uses the known target as a template and, in each trial, computes the dot-product between the template and input image at the known target location (target region). Then, it compares the dot-product to an optimally placed

criterion to make a decision. It has been previously shown that human thresholds and d-primes follow a similar trend as template matching observer (e.g., Sebastian et al., 2017; for reviews, Cohn & Lasley, (1986); Geisler, 2003). Even though the pattern of thresholds and d-primes are similar in shape, template matching observers (including cases when they are implemented as ideal observers) usually perform better. The overall difference is generally explained by a free parameter (multiplicative parameter on d-primes or thresholds). There could be multiple sources of human inefficiency. For example, neural noise (Tolhurst et al., 1983), decision noise due to criterion placement (Wickelgren, 1968), sub-optimal templates (Beck et al., 2012), motor noise, and attention variability (or some combination) could explain the difference. In conclusion, methods for measuring performance under low uncertainty are well-developed, and for canonical detection experiments in noise, the pattern of results is generally consistent with ideal-like observers.

2.3 Measurement of Human Detection Performance under Modest Uncertainty

In an experiment with modest levels of uncertainty, properties of the target and/or background vary within a single block. Trials that were separated into different blocks in the previous section are now interleaved or mixed. The extrinsic uncertainty is also described as "signal and/or background known statistically" (SKS). In contrast, an experiment with no extrinsic uncertainty (or low levels of uncertainty) is described as "signal and/or background known exactly" (SKE). When single dimensions of uncertainty are not combined with a random sampling of background, the amount of uncertainty is similar to low uncertainty experiments because they both essentially include a single dimension of uncertainty. However, most single dimensions of uncertainty generate systematic variations. In contrast, a random sampling of backgrounds does not involve any

systematic variation, and performance is generally averaged over the variation. Therefore, I will also consider single dimensions of uncertainty that are not combined with a random sampling of noise backgrounds as modest uncertainty because they are methodologically more similar to single dimensions of uncertainty with a random sampling of backgrounds. The experiments that are reviewed here cover both experiment types (single dimensions of uncertainty with and without a random sampling of backgrounds).

There have been many studies of modest levels of uncertainty. For example, often in visual search experiments (which, by definition, involve target position uncertainty), there are distractors that differ from the target by a single visual feature (like color or orientation; for review Eckstein, 2011). However, these experiments can be cast and modeled in the same framework as classic position uncertainty (search) tasks involving signals in noise backgrounds (Solomon et al., 1997; Foley & Schwartz, 1998; Eckstein et al., 2000; Nachmias, 2002). Furthermore, it is possible to measure the effect of extrinsic uncertainty for various tasks. Until this point, I described a simple yes/no task, but different tasks can be designed to collect different behavioral measures. For example, there are many visual search experiments where the observer's reaction time is measured and modeled (Wolfe, 2015). Also, the effect of extrinsic uncertainty is measured for orientation discrimination and depth discrimination thresholds (Westheimer & Eric, 1996), disparity modulation thresholds (Georgeson et al., 2008), and velocity discrimination thresholds (Chen et al., 1998). Here, however, I will not comprehensively cover all this previous literature. Specifically, I will focus on detection or recognition experiments in which participants are asked about the presence of the target, and human performance is measured as the number of correct responses (accuracy) in homogenous (and statistically stationary) noise or uniform backgrounds. I will consider both the Yes/No task and two or more interval (spatial or temporal) forced-choice experiments. Moreover, I will report findings

from the discrimination experiments in which participants are asked to identify which target is presented (that does not necessarily have a target-absent response option).

The most common method for understanding/modeling the effect of a single dimension of uncertainty is comparing the performance in a blocked condition (low or no uncertainty) to an unblocked condition (modest uncertainty). If participants are uncertain about target amplitude, it has been shown that the performance is not very different from the low uncertainty condition where participants have information about target amplitude (Davis et al., 1983). For target-orientation uncertainty, the performance in the modest uncertainty condition is either slightly worse (Ukkonen et al., 1995) or not different (Doehrman, 1974) than in the low certainty condition. In contrast, Sekular and Ball (1977) found that detection performance is considerably worse than the low certainty condition if participants are uncertain about either direction or speed of motion of the target. Similarly, uncertainty about the phase of the grating target has been found to decrease the performance (Burgess & Ghandeharian 1984a; Howard & Richardson, 1988). It has been relatively uncommon to vary the properties of the background. Ahumada and Beard (1997) and Beard and Ahumada (1999) showed that human performance decreases in the forced-choice experiment if each interval has a different background (thus, each trial has a different background) than if the background is fixed for all intervals and all trials.

When a single dimension of uncertainty is introduced, the observer must consider all possible targets and backgrounds, so multiple channels (like multiple locations) must be monitored to make a decision. A random noise might interfere with processing in any of these channels (a sample of noise in all independent locations might look like a target). Because of that, the probability of making a false alarm increases when trials are intermixed, which makes the task harder. It is widely appreciated that human performance is expected to decrease, and it is possible to quantify the expected decrease in the d prime

of human observers. The uncertainty effects are better understood when the expected effect is compared to the measured effect. There are many studies that also quantify the expected decrease in performance with the introduction of uncertainty. If human performance decreases at the expected rate, the observer will have the same efficiency in both conditions relative to the optimal performance. The efficiency is defined as the ratio of the square of the observer's d prime to the square of the ideal observer's d prime. An observer with the same efficiency in both conditions will perform worse in the higher uncertainty condition, so comparing the efficiency of human observers in both conditions ensures that the expected effect is accounted for. If human performance decreases more than expected, the efficiency will be lower in the higher uncertainty condition. If human performance decreases less than expected, the efficiency will be higher in the higher uncertainty condition.

These three possibilities are associated with well-known hypotheses in the literature (same efficiency, higher efficiency, and lower efficiency). However, note that each efficiency comparison may generally involve multiple forces acting opposite or synergistically with each other. First, if human performance decreases more than expected, the efficiency will be lower in the higher uncertainty condition. The most common hypothesis for lower efficiency is computational imperfections (like not properly integrating template responses). Specifically, assuming there is a limit on the attentional capacities is the most common idea which can be implemented in many different ways as computational inefficiency (Lee et al., 1999; Lu & Dosher, 1998). Attentional capacity limits suggest that when participants are uncertain, they cannot properly attend to all possible properties because there are limited resources for processing. Note that this hypothesis suggests that attention enhances sensitivity when there is less extrinsic uncertainty. Secondly, if human performance decreases at the expected rate, the efficiency

will be the same in the higher uncertainty condition. No change in efficiency suggests that there is no capacity limit, so every alternative is processed in parallel. In this case, better performance is measured when there is less uncertainty because the visual system is hypothesized to use prior information to reduce uncertainty by only concentrating on relevant channels (uncertainty reduction hypothesis, Tanner, 1961; Pelli, 1985.). There is evidence for both hypotheses under different circumstances (Carrasco et al., 2000; Gould et al., 2007), and the issue is reviewed extensively for visual search (Luck et al., 1996; Carrasco, 2011). Lastly, if human performance decreases less than expected, the efficiency will be higher in the higher uncertainty condition. The standard hypothesis for measuring higher efficiency in the higher levels of uncertainty is the existence of intrinsic uncertainties (Tanner, 1961; Pelli, 1985). If an observer has intrinsic uncertainty, the human performance will decrease considerably in lower uncertainty conditions compared to having no intrinsic uncertainty. However, extrinsic uncertainty will dominate in higher uncertainty conditions, and the performance will be less affected by intrinsic uncertainties.

It is possible to classify the results of the experiments that compared different levels of uncertainty in terms of efficiency comparison. For example, uncertainty about the temporal location of the target is shown to decrease the performance (Lowe, 1967; Earle & Lowe, 1971). Later careful quantification of the decrease revealed that the effect was consistent with the predictions of signal detection theory (Lasley & Cohn, 1981). Thus, humans have the same efficiency for both tasks (consistent with humans appropriately using the prior information with no further constraints). When the task is detecting the change in color at the target location, uncertainty about the direction of the color change decreases the performance consistent with having the same efficiency for both conditions (Greenhouse & Cohn, 1978). The effect of target position uncertainty is generally found to be consistent with having the same efficiency (Cohn & Lasley, 1974; Swenson & Judy,

1981; Davis et al., 1983; Cohn & Wardlaw, 1985). Some studies found no change in performance (Mertens, 1956; Ukkonen et al., 1995) with the introduction of position uncertainty. However, these studies did not quantify the expected difference in their experimental settings, which differed considerably from a standard setup. Recently, Walshe and Geisler (2022) found that the incorporation of intrinsic uncertainty into modeling is necessary to predict human performance under modest uncertainty from the low uncertainty condition. Therefore, the efficiency is higher in the higher uncertainty condition when the search area is very large (thus, the uncertainty about the target's spatial position is large). The effect of size uncertainty (when spatial frequency is fixed) is consistent with humans being more efficient in the modest uncertainty condition because only a small decrease in the performance is found, which is less than the expected decrease in the performance (Judy et al., 1995; Judy et al., 1997). More recent work has also found only a small effect of uncertainty, but they did not quantify the expected effect (Meese et al., 2005; Foley et al., 2007; Meese & Summers, 2012). The simple effect of spatial frequency uncertainty (when the size is fixed) is generally consistent with the same efficiency (Davis & Graham, 1981; Davis et al., 1983; Yager et al., 1984). However, Hübner (1996a,1996b) quantified how the effect varies with the target amplitude (measured psychometric function) and found the data can only be explained with an observer model that is uncertain about the phase which calculates the energy (and he pointed out other types of intrinsic uncertainties might also explain the data). This result suggests that the efficiency is higher in the higher uncertainty condition. Later, these results were challenged by Ohtani et al. (2002). They measured inconsistent psychometric functions under similar conditions and showed their data is consistent with attentional enhancement, thus implying lower efficiency in the higher uncertainty condition. There are some studies in which the target or background is not varied along any intuitive dimension.

For example, Burgess (1985) randomly picked the target from the set of orthogonal patterns (Hadamard set). They found that the efficiency is the same for low and modest levels of uncertainty, so participants are able to use prior information in the low uncertainty condition to reduce their uncertainty (consistent with the uncertainty reduction hypothesis, with no intrinsic uncertainty). (Also see Kramer et al., 1985). In sum, most experiments found that the effect of single uncertainty dimensions (modest uncertainty) is either consistent with humans being similarly efficient in both levels of uncertainty or consistent with humans being more efficient in higher uncertainty conditions.

There are few examples of directly fitting model observers to the data rather than focusing on the comparison between various levels of uncertainty. For example, Burgess and Ghandeharian (1984b) measured the localization performance of observers under position uncertainty. They varied the amount of uncertainty and target amplitude across blocks. (There is no extra extrinsic uncertainty created by that because it is blocked). Rather than generating predictions based on low uncertainty conditions, they directly fit the d-primes of the ideal observer to the d-primes of participants and found the ideal observer well explained all the data with a single efficiency parameter. More recently, Meese and Summers (2012) used data collected under the modest size uncertainty and low size uncertainty to differentiate between different model observers of spatial pooling (integration) by directly fitting observer models to the data. They found evidence for an observer model with a non-linear transducer and linear spatial pooling and showed that none of the versions of the classic MAX observer they tested accounted for the data.

2.4 Measurement of Human Detection Performance Under High Uncertainty

There are only a few studies of multiple simultaneous dimensions of uncertainty. Some of them did not quantify the expected decrease in performance. For example, Howarth (1966) compared the high uncertainty condition, where there is uncertainty about the target's location, size, and temporal location (total of 200 alternatives), to the low uncertainty condition, where participants have all prior information about all of these dimensions. He found no difference in performance between the two conditions. Sekular and Ball (1977) found a decrease in the performance when there is combined uncertainty about the target's motion velocity and direction compared to the low uncertainty condition, where both were fixed within blocks. Eckstein et al. (1997) measured the performance under the combination of position uncertainty (4 locations) and background uncertainty. When there is background uncertainty, all four locations contain a different background, but if there is no background uncertainty, all of them contain the same background. They compared the high uncertainty condition to the modest uncertainty condition where position uncertainty is still present and found that human performance decreased with the introduction of background uncertainty.

There are some more recent studies that quantified the expected decrease in performance due to uncertainty. Eckstein et al. (1996) measured the performance under combination position uncertainty (4 locations) and uncertainty about the target’s phase of temporal modulation. They compared the high uncertainty condition to the modest uncertainty condition where position uncertainty is still present and found that a decrease in human performance is consistent with the expected decrease given the uncertainty reduction hypothesis. Therefore, the efficiency of human observers is similar for both conditions. Wilson & Manjeshwar (1999) did a very similar experiment with a more

complex background. They found that a decrease in human performance is smaller than predicted by the ideal observer. Thus, humans are more efficient in the high uncertainty condition, consistent with the existence of some intrinsic uncertainty. Eckstein and Abbey (2001) measured the human performance under combined target size and shape uncertainty and compared it to the condition where participants exactly know both. They found that humans are more efficient in the high uncertainty condition. Park et al. (2005) measured the human performance under the combined target position (128^2 number of potential locations) and background uncertainty. The background uncertainty is introduced in lumpy backgrounds by randomly sampling the number of lumps. They compute the efficiency of participants in the high uncertainty condition and the low uncertainty condition that specifies the location of the target, but background uncertainty remains intact. They found that humans are more efficient in the high uncertainty condition (which is similar to what Walshe and Geisler (2022) found). In sum, most experiments found that the effect of multiple dimensions of uncertainty (high uncertainty) is consistent with humans being more efficient in higher uncertainty conditions.

Some studies of multiple dimensions of uncertainty differentiate between different model observers based on how well they explain the data rather than comparing different levels of uncertainty. Eckstein and Whiting (1996) and Bochud et al. (2004) measured the performance under both target amplitude (signal contrast) and target position uncertainty for various blocked levels of target position uncertainty. They tested the assumption that the internal decision variable of participants is Gaussian distributed. They generated the predictions of signal detection theory based on different assumptions about the distribution of the internal decision variable and fit these models to the data. They found that the Gaussian assumption explains the data fairly well, but better-fitting models deviate from the assumption. Castella et al. (2008,2009) measured human performance under

simultaneous shape and size uncertainty and various low uncertainty conditions. They fitted various model observers (like the channelized Hotelling observer, the non-prewhitening observer with an eye-filter, the dense difference of Gaussian channels model) to the data and compared them. They showed that human performance is only considerably decreased when there is simultaneous size and shape uncertainty. They found that all model observers have pros and cons, and there is no clear best model. Han and Baek (2020) measured human performance under simultaneous background and target uncertainty. Background uncertainty is generated by changing the anatomical background in each trial, and signal uncertainty is generated by randomly sampling from eight possible signals. They did not quantify human performance under any other uncertainty condition. They simply used the data collected under high levels of uncertainty to evaluate various observer models to see which one best explains the data, including a CNN type of model. They found that the three-layer CNN network trained for this task better predicts the data than classical model observers (like the channelized Hotelling observer, the non-prewhitening observer with an eye-filter, the dense difference of Gaussian channels model).

2.6 Challenges For Measuring and Modeling Human Detection Performance Under High Uncertainty

There are only a few studies on multiple simultaneous dimensions of uncertainty (high levels of uncertainty), even though real-world tasks always include many dimensions of uncertainty. Studying multiple simultaneous dimensions of uncertainty is a reasonable next step for understanding real-world tasks. Although many dimensions of uncertainty have been studied in isolation and their effects demonstrated, it is crucial to understand how these effects combine. Recent studies focus more and more on combined dimensions of uncertainty but never get close to the high levels of uncertainty present in the real world.

However, they showed that behavioral data is rich enough to differentiate between models and test various hypotheses, so studying higher levels of uncertainty will likely provide a rich dataset. Studying more naturalistic tasks also aligns with the conceptual approach of the thesis. They are likely to be relevant through evolutionary history, because the visual system has evolved to handle uncertainty along all the dimensions that occur in natural stimuli.

Some challenges faced when studying high and modest levels of uncertainty are likely to worsen with even higher levels of uncertainty. I identified three challenges. Firstly, when higher levels of uncertainty are introduced, a single block of uncertainty experiment requires more data collection to obtain stable hit and correct rejection rates (there will be many alternatives, and all of them should be presented many times). Running multiple blocks for measuring a psychometric function becomes even more taxing. Furthermore, comparing the data with the low uncertainty condition also becomes harder since there will be more pairs of target and background conditions for which performance should be measured.

Secondly, for modest and high levels of uncertainty, most of the reported metrics of the data and experiment structures (measuring psychometrics, thresholds) are borrowed from the well-developed methods for the experiments under low uncertainty. Under modest levels of uncertainty, some of the studies just reported the overall percentage correct (Sekuler & Ball, 1977; Hübner, 1996b; Han & Baek 2020). Psychometric functions of overall percentage correct are measured by running multiple uncertainty blocks (method of constant stimuli) or implementing a staircase method in the same block (Mertens, 1956; Ukkonen et al., 1995; Hübner, 1996a; Ahumada & Beard, 1997; Beard & Ahumada, 1999; Meese et al., 2005; Foley et al., 2007; Meese & Summers, 2012). On the other hand, some studies made use of hit and correct rejection rates. From these rates, researchers focused

on sensitivity measures and accounted for the bias either by calculating d-prime or correcting for interval biases in forced-choice experiments (Lowe, 1967; Earle & Lowe, 1971; Cohn & Lasley, 1974; Doehrman, 1974; Greenhouse & Cohn, 1978; Lasley & Cohn, 1981; Swenson & Judy 1981; Davis & Graham, 1981; Davis et al., 1983; Yager et al., 1984; Howard & Richardson, 1988; Judy et al., 1995; Judy et al., 1997; Eckstein & Abbey, 2001; Castella et al., 2008; Castella et al., 2009). Furthermore, sensitivity measures were measured for each level of a variable using psychometric methods (Burgess & Ghandeharian 1984a; Burgess & Ghandeharian 1984b; Cohn & Wardlaw, 1985; Eckstein et al., 1996; Eckstein et al., 1997; Wilson & Manjeshwar, 1999; Park et al., 2005). However, the appropriateness of these metrics and experiment structures for higher uncertainty conditions is questionable. For example, the calculation of d-prime assumes the internal decision variable of participants is Gaussian distributed. However, the assumption is shown to be violated (Eckstein & Whiting, 1996; Bochud et al., 2004). Moreover, if participants are similar to ideal observers, the assumption is expected to be violated because the assumption is only true for an ideal observer under special conditions (orthogonal signal and backgrounds). Therefore, calculating d-primes with the same methods used in low uncertainty experiments might not reflect actual sensitivity in the modest level of uncertainty experiments, and these assumptions will likely be violated even more drastically with the combination of more dimensions of uncertainty. There are further complications in calculating d-prime for each target level from the individual hit rate corresponding to that target and the overall correct rejection rate. Because of these complications, comparing d-primes of each target levels across different levels of uncertainties might be misleading. These problems in calculations of d-primes might mask the similarities between participants and observer models and might mislead the efficiency analysis, which has been very fruitful in the past.

In general, the appropriateness of any method borrowed from a low extrinsic uncertainty experiment is questionable because the structure of the raw data collected in a detection experiment with no or low extrinsic uncertainty differs from the data collected in a detection experiment under modest and high levels of uncertainty. Similar to the experiments with low or no uncertainty, the overall percentage correct can be broken down into the hit and correct rejection rates. However, in contrast to having a single correct rejection and single hit rate for a single block, there is a matrix of hit and correct rejections under modest levels of uncertainty. Specifically, for each pair of different target pattern and background property, there is a distinct hit rate. And, for each property of background, there is a distinct correct rejection rate. For example, when there are 8 levels of target amplitude and 8 levels of background contrast, there are 64 distinct hit rates and 8 distinct correct rejection rates. Note that it is possible to collect 64 distinct hit rates and 64 distinct correct rejection rates with 64 blocked experiments under low uncertainty. In this case, however, the matrix of hit and correct rejections is composed of independently sampled units (each sampled in a blocked condition). These units are put together to create a matrix. Under high uncertainty, different rates are not independently sampled, so it might be even more appropriate to say that the matrix of hit and correct rejection rates is independently sampled (measured). Therefore, these two matrices are fundamentally different. In addition to concerns about the appropriateness of methods, there is a lack of methodical tools designed to measure and model the data under high uncertainty by exploiting these fundamental differences.

Thirdly, so far, the result of finding higher efficiency in the high uncertainty condition did not attract much modeling effort. However, as more studies with higher levels of uncertainty are conducted, finding higher efficiency in the high uncertainty conditions might become more prevalent. Recent experiments that are covered above measuring

human performance at higher levels of uncertainty are already reporting higher efficiency in higher uncertainty conditions most of the time. There are multiple reasons behind these findings. First, humans are already shown to be more efficient in the higher uncertainty for dimensions like target size uncertainty when they are studied in isolation. Combining these dimensions with others is likely to produce similar results. Secondly, since intrinsic uncertainties are likely to be part of the visual processing in the visual system, as extrinsic uncertainty increases, the intrinsic uncertainty will have less and less effect on the higher uncertainty condition than its effect on the low uncertainty condition. Thus, the existence intrinsic uncertainties also imply that humans are likely to be more efficient in the higher uncertainty condition. Therefore, understanding and modeling higher efficiency in uncertain conditions might be more important as the uncertainty increases. Simple laboratory experiments aim to generalize their results to more real-world conditions, but if human performance is consistently better than predicted, it is important to address this systematic failure. Therefore, it is essential to formulate new hypotheses about higher efficiency by making conceptual and theoretical analyses and testing new hypotheses about the source of higher efficiency (in addition to intrinsic uncertainties). Furthermore, it is important to simulate or calculate predictions of model observers having intrinsic uncertainties, which is not straightforward with current methods.

All these three challenges faced in the literature can largely be resolved with explicit simulations of model observers and, specifically, the ideal observer. By explicit simulations, I mean constructing image processing model observers that take the image in each trial and make a decision. These simulations will result in predictions for hit and correct rejection rates (and matrices). Firstly, directly fitting the model observers to the human data (hit and correct rejection matrices) also helps resolve the data collection problem. It makes it possible to formulate another general approach to uncertainty

experiments: differentiating between model observers with the data collected in a single block of uncertainty experiment without collecting data for the low uncertainty condition or different blocks of the same uncertainty condition. With a higher level of uncertainty, the hit and correct rejection rates will provide a richer data structure, making this approach even more plausible. Secondly, since varying an efficiency scalar factor and bias factor for the ideal observer will directly affect the predicted hit and correct rejection rates, it is possible to fit an ideal observer's hit and correct rejection rates to participants' patterns of hits and correct rejection rates by varying these two parameters. Estimated efficiency scalar factor and bias factor might be used instead of sensitivity and bias estimated with classic signal detection theory. Having these parameters that do not suffer from weaknesses of signal detection theory estimation will resolve the second challenge. Thirdly, having explicit simulation methods will make it possible to replace some computations of model observers with more biologically plausible computations such that the sub-optimal model observer will predict higher efficiency in uncertainty conditions. Thus, it is possible to formulate and test more hypotheses about the higher efficiency. Also, it is easy to incorporate intrinsic uncertainty in model observers and test these model observers against the data. Therefore, having explicit simulation methods for the ideal observer allows us to rigorously test hypotheses about higher efficiency and formulate new hypotheses, which will resolve the third challenge.

Apart from addressing these issues related to the literature, the ideal observer analysis and explicit simulation methods are aligned well with the conceptual approach presented in Chapter 1. Since detection under higher levels of uncertainty is a naturalistic task, it was likely to be relevant for the animal's survival through evolutionary history. Therefore, the accuracy in the task is expected to be maximized up to some extent, so starting modeling with the ideal observer, which determines the best possible performance,

is appropriate. In the next sections, I will start with the derivation and formulation of the image processing model observers, and specifically the ideal observer.

2.7 A Generic, More Naturalistic But Tractable Visual Task

Consider a sine-wave target windowed by a raised cosine (a Gabor-like pattern). Like any real-world target, its three-dimensional position and orientation could vary in the environment, and depending on the lighting conditions, the luminance pattern of the target could vary. There are multiple ways of formulating these variations. The variation in the target's location in depth generates target scale uncertainty, which is part of the generic task considered in this thesis. However, I will not include uncertainty related to the two other spatial position dimensions. Two-dimensional position uncertainty (visual search) makes the task much more computationally intensive and requires modeling to involve known properties of the visual system. For example, visual processing at peripheral locations is vastly different due to ganglion sampling. Since the sine-wave target is only defined to be two-dimensional, three-dimensional orientation variation is projected (approximately) as the variation in the two-dimensional orientation, which is also included in the generic task (orientation uncertainty). The variation due to lighting conditions can have a complicated effect on the luminance pattern of the target location. However, I will not focus on these variations because of their complexity and because they can be estimated to some extent for each environment. Lastly, there are variations related to the signal strength of the target. Some of these could be related to lighting conditions and some to variation in reflectance. I will consider a simple signal strength variation will focus on additive targets for tractability. Specifically, the target shape is normalized to a peak of one, and the target amplitude is defined as the single scale factor multiplying the target shape. For each pixel,

the luminance of the background is added to the luminance of the target. Thus, the variation regarding the target amplitude is included in the generic task (target amplitude uncertainty).

Consider a white noise background. The luminance of each background pixel is independently sampled from a gaussian distribution with zero mean and non-zero standard deviation (white noise background). In each trial, a new sample of white noise background is presented. Because white noise is already dependent on only two parameters, only multiple properties can be varied to introduce uncertainty. I decided to focus on the most critical dimension: contrast (standard deviation of gaussian). Even the most homogenous environments in the real-world generally include variations in local contrast, so the visual system will be uncertain about the local contrast in the real-world.

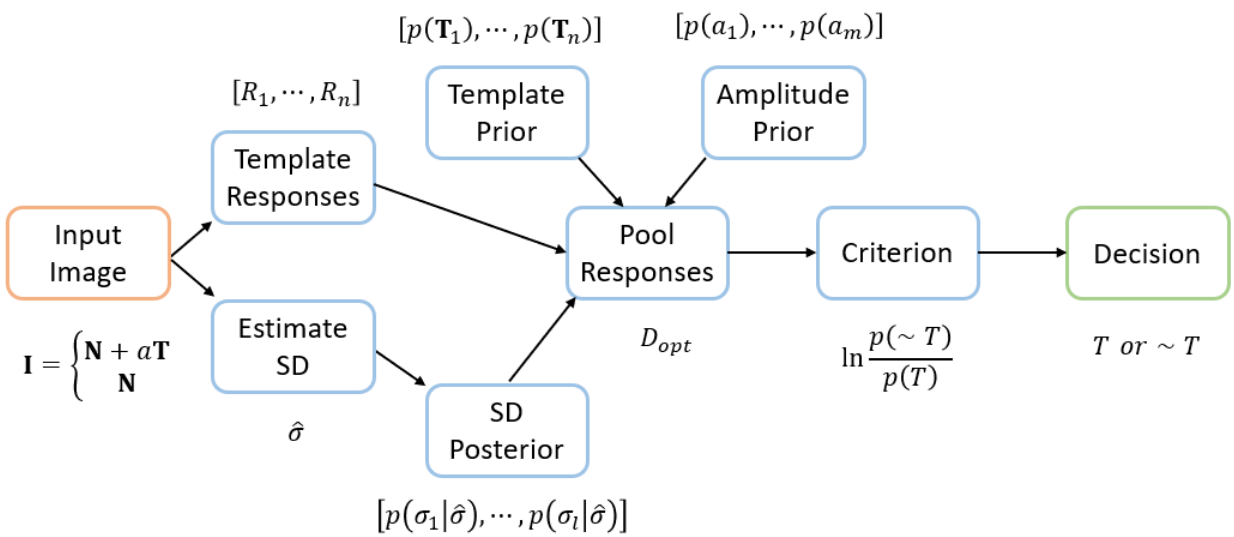
Taken together, the generic task analyzed in this thesis is detecting sine-wave targets windowed with a raised cosine added to white noise backgrounds. In each trial, participants judge whether the target is present or absent (the target prior probability determines the likelihood that a target will appear in a given trial). I consider cases where the participants may be uncertain about the target's scale, orientation, amplitude, and the background contrast. In the most general case, there is combined uncertainty about all these dimensions. The variation in this general case captures a considerable amount of variation involved in the real-world visual detection tasks and almost all the possible dimensions for an additive two-dimensional target in white noise. The generic task is still simple enough to derive the ideal observer and achieve the aims of the thesis. Another advantage of the generic task is that extensive research (both with uncertainty and without uncertainty) has already been done along these stimulus dimensions. Established results in the literature will help me to better reason about the results measured in the experiments.

These simplified settings most resemble detecting targets in medical images, where medical experts are required to detect various targets (like tumors) that might be present.

These targets come in various shapes and sizes, which generates uncertainty about the properties of targets. Also, the backgrounds vary in a complicated manner. Most importantly, because of the nature of imagining devices, targets are approximately added to the backgrounds. Therefore, the findings of this thesis might potentially have some practical consequences. However, the main aim of this thesis is to learn about how visual system operates under uncertainty.

The ideal observer is straightforward to describe as a template matching observer because the target is additive, and the background is white noise (see Figure 2.1). I assumed that the prior probabilities of background contrasts (required for computing posterior of the response standard deviations), target amplitudes (amplitude prior), and different targets (template prior) are available to the observer. The input image can be described as either noise (**N**) or noise plus target (**N** + a**T**) with amplitude of a. For all the possible target templates, the template responses can be computed by taking the dot product of the target region in the image (**I**) and template ($\mathbf{T}_i$). Note that for the generic task, variations in templates are generated by orientation and scale variations, but the theory applies in general for any set of targets. The background is generally larger than the target, so estimating the contrast (or standard deviation) associated with white noise background is possible. Given the estimated value (Estimate SD), the posterior probabilities over the possible true standard deviations can be computed (SD Posterior). The decision variable of the ideal observer (likelihood ratio, D_{opt}) is calculated by appropriately pooling these variables with respect to prior probabilities of each amplitude and target. Pooling is done with the equation provided in Figure 2.1. The decision is made by comparing the decision variable with the ratio of priors which can be conceptualized as a fixed criterion (assuming that prior

probabilities of target-present and absent are known). If the decision variable exceeds the criterion, the observer responds as target-present; otherwise responds as target absent.



$$D_{opt} = \ln\left(\sum_{j=1}^{l} p(\sigma_j|\hat{\sigma}) \sum_{k=1}^{m} p(a_k) \sum_{i=1}^{n} p(\mathbf{T}_i) \exp\left[\frac{a_k \mathbf{I} \cdot \mathbf{T}_i - 0.5 a_k^2 \mathbf{T}_i \cdot \mathbf{T}_i}{{\sigma_j}^2}\right]\right)$$

Figure 2.1 The ideal observer for the generic task.

Simulations of model observers under high levels of simultaneous multiple dimensions of extrinsic uncertainty are computationally intensive. Also, working on the combined effect of multiple dimensions without understanding the effects of subsets of these combinations makes it harder to disassociate the effects. Therefore, I split the generic task into two subtasks and studied them individually for this thesis. The first subtask includes simultaneous target amplitude and background contrast uncertainty. These dimensions fit well together because they do not introduce any pattern variation (there is only a single template). Consequently, the second subtask includes target orientation and scale uncertainty, all dimensions in the generic task that generate variations in the template

pattern. In the end, it is possible to put everything back together and reason about the generic (general case) task. The methods are derived for these subtasks can be used in the generic task to make it easier to analyze and simulate various observer models.

In the next chapter (Chapter 3), I will measure human performance and evaluate the ideal observer and various sub-optimal observers under simultaneous target amplitude and background contrast uncertainty. In the following chapter (Chapter 4), I will do the same for detection under simultaneous target orientation and scale uncertainty. In the last chapter (Chapter 5), I will discuss the results of these two subtasks and their relation to the generic task. I will further elaborate on how this thesis contributes to the previous uncertainty literature and discuss potential new directions for studying how the visual system operates under high levels of simultaneous multiple dimensions of extrinsic uncertainty.

Chapter 3: Simultaneous Target Amplitude and Background Contrast Uncertainty

3.1 MOTIVATION

In general, the detection of target objects depends on the specific properties of the targets, the specific properties of the backgrounds, the random variation of the target and background properties (i.e., the target and background uncertainty), and the prior probability of the target being present in the image. This chapter is concerned with two stimulus properties (dimensions): the strength (amplitude) of the target and the contrast of the background. These dimensions are highly variable under natural conditions and have a large effect on the discriminability of the target.

There are multiple sources of variation in the target amplitude. Sometimes within the same target category (like the category of apple), different individual members share the same visual pattern but vary in amplitude. Secondly, the same individual member's amplitude might also vary because of the illumination and its placement in the environment. Therefore, even when searching for a fixed visual pattern, the visual system almost always faces uncertainty about the target amplitude. Variation in background contrast is also common in natural scenes. It has previously been found that the contrast at one location is largely uncorrelated with the contrast in a location couple of degrees away (Frazor & Geisler, 2006). Therefore, the visual system almost always faces uncertainty about the background's contrast too. Lastly, the prior probability of a target being present varies drastically from one environment to another. In general, the likelihood of a specific target being present in any random fixation is extremely low (a fork being present in any random fixation location in a random scene is almost zero). However, the likelihood might be higher for some environments (if you are in the kitchen) and even higher under some specific conditions (when you know you are looking at the kitchen table). These differences

in prior probability are likely to be taken into account in computations performed by the visual system.

I will begin the analysis with the derivation of the ideal and sub-ideal observers for white noise backgrounds and additive targets. I will describe simulations of these models to understand the effect of various computational principles on the performance and identify differences between models. Under simultaneous background contrast and target amplitude uncertainty, it turns out that the ideal observer is a relatively simple extension of the classic template-matching observer. It still applies a fixed template, but it dynamically varies the decision criterion based on the estimated background contrast. A plausible sub-optimal model candidate is a classic template-matching observer that does not dynamically vary its criterion but has a single fixed criterion applied regardless of background contrast. It is the standard observer for explaining performance when there is low uncertainty. Thus, it embodies the simplest hypothesis: the underlying visual mechanisms under high uncertainty are the same as those under low uncertainty. Another plausible sub-optimal candidate can be constructed by normalizing the template response by the estimated level of background contrast. Normalizing responses with estimated contrast (i.e., contrast gain control) is already a known property of neurons in the primary visual cortex Albrecht & Geisler, 1991; Heeger, 1992; Carandini & Heeger, 1994; Carandini et al., 1997; Cavanaugh et al., 2002). Therefore, simple decoding of these neurons might suffice to implement the contrast-normalization computation in the brain. These three model observers will be the focus of this Chapter. In light of the simulation results of these computational principles, I will describe the psychophysical experiments.

The primary aims of this chapter are measuring and modeling the human ability to detect targets in white noise and natural backgrounds, when background contrast and target amplitude are randomly varying over a wide range, for different prior probabilities of target

present. Previously, the uncertainty about target amplitude was found to almost have no effect on human performance compared to when the participant is certain about the target amplitude (Davis et al., 1983). There are no studies on background contrast uncertainty to the best of my knowledge except the recent paper from our group (Sebastian et al., 2017) that measured the effect of combined target amplitude and background contrast uncertainty with natural backgrounds. However, it is not possible to derive and test the exact ideal observer for natural scenes. Here, with white noise backgrounds, it is possible to derive the ideal observer and test various heuristics and sub-optimal computational strategies against the human data. I will compare the detailed predictions of the three model observers with the psychophysical data measured in white noise backgrounds. Then, I will test the same model observers for the detection in natural backgrounds (although none of them are strictly ideal) and compare the predictions with the psychophysical data measured with the natural backgrounds.

3.2 Theory of Detection Under Amplitude and Background Contrast Uncertainty

Recall that the target is normalized to a peak of one, and the amplitude is defined as the single scale factor multiplied with the target. The gray level of each background pixel is independently sampled from a gaussian distribution with zero mean and non-zero standard deviation (white noise background). In each trial, a new sample of white noise background is presented. For each pixel, the gray level of the background is added to the gray level of the target. The generic case introduced in Chapter 2 is simplified here because there is only a single template (target pattern) for the theory of background contrast and target amplitude uncertainty. Thus, there is only a single template response, and the template prior probability is one for the single template (Figure 3.1). Background contrast

is randomly sampled from a pre-defined range with a stationary probability distribution, and for target present trials, both target amplitude and background contrast are randomly sampled from pre-defined ranges with stationary probability distributions. The target amplitude uncertainty is illustrated in the upper panels of Figure 3.2, and the background contrast uncertainty in the lower panels. The task is to indicate whether the target is absent or present.

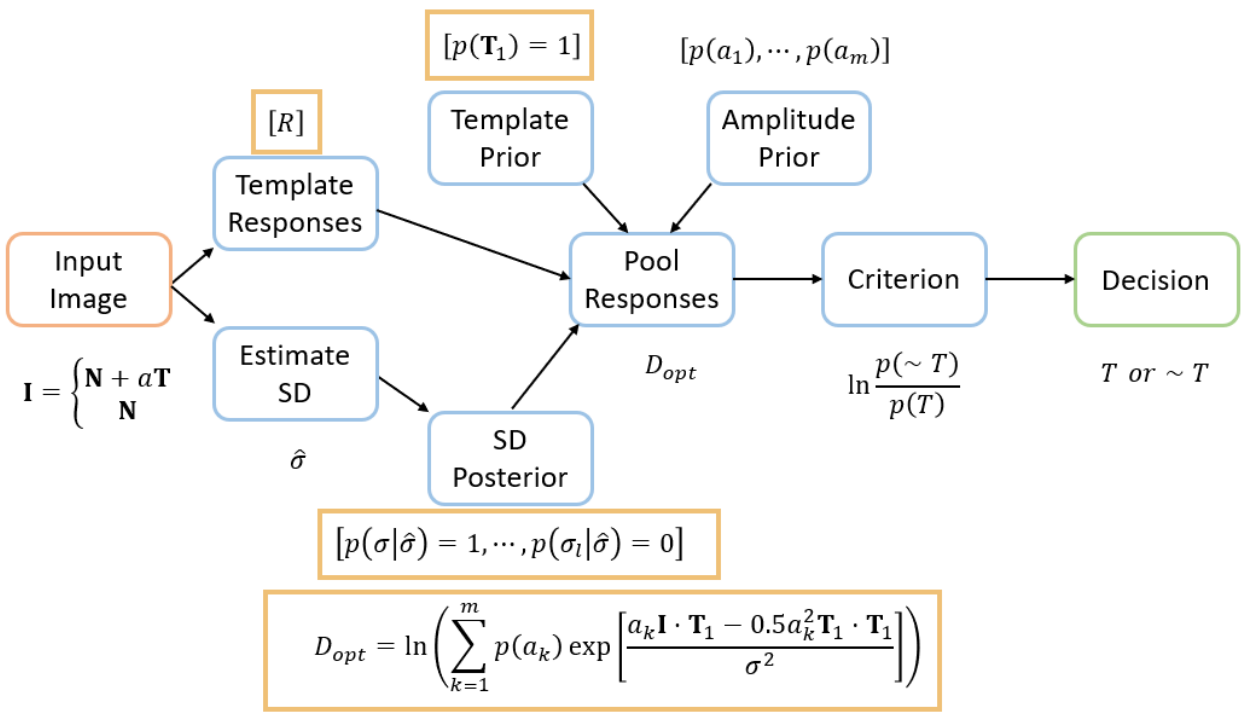


Figure 3.1 The simplification of the ideal observer.

For the theory of background contrast and target amplitude uncertainty, the ideal observer for detecting an additive 2D object in the environment.

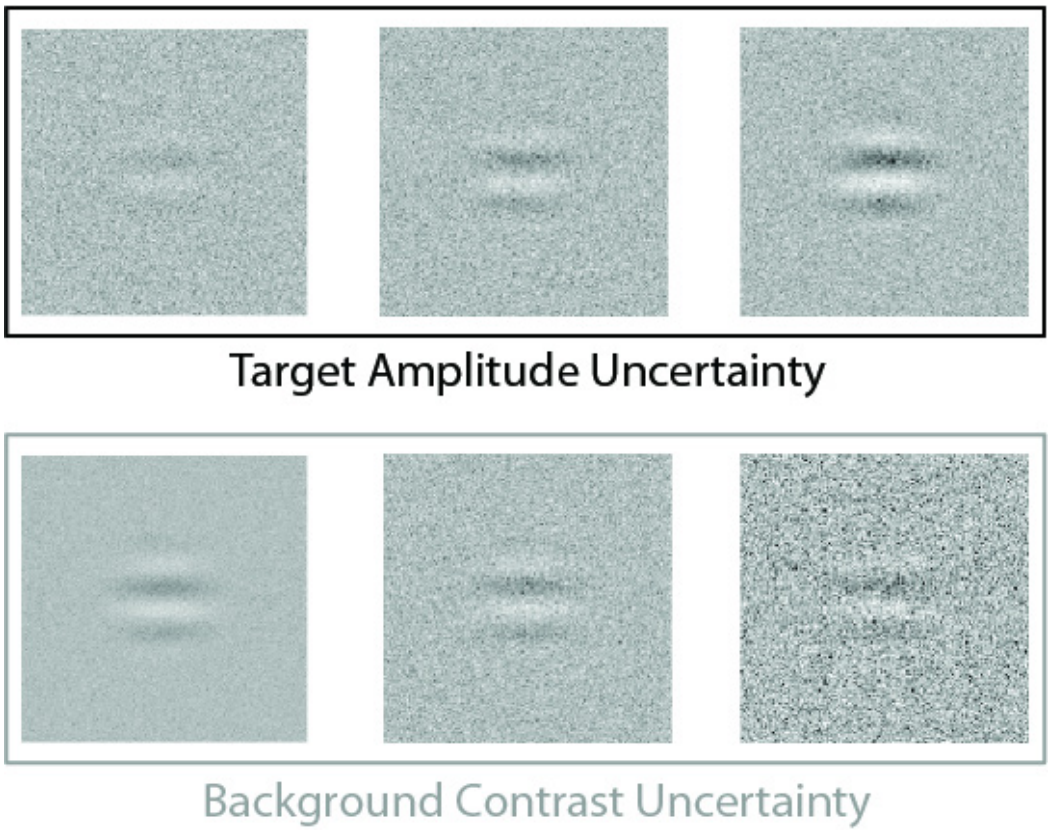


Figure 3.2 Target amplitude and background contrast uncertainty.

For target present trials, target amplitude is randomly sampled from multiple pre-defined levels that cause participants to be uncertain about the strength of the signal prior to detection (target amplitude uncertainty). For every trial, background contrast is randomly sampled from multiple pre-defined levels that cause participants to be uncertain about the background contrast prior to detection (background contrast uncertainty).

3.2.1 Models

3.2.1.1 The Ideal Observer

I assumed that the prior probability of the target being present and prior probabilities of target amplitudes and background contrasts are known to the ideal observer. Furthermore, the ideal observer is assumed to know the background contrast because the background patch is big enough to precisely estimate it. The target location is also assumed to be known (no position uncertainty).

The ideal observer responds target-present if the posterior probability of target present is more than target absent. I showed that the ideal observer can be implemented by extending a simple template-matching observer. In each trial, a simple template-matching observer computes the dot product of the template and the target region of the image and then compares the dot product (template response) to a criterion (γ) to make a decision. The template is defined to be the target, normalized by its total energy. If there were only amplitude uncertainty, comparing the template response to an optimally placed criterion would be equivalent to a maximum posterior estimation (the ideal observer) because the template response is a sufficient statistic for the likelihood ratio, and it is strictly increasing with the ratio (Appendix 1). The optimal criterion depends on the background contrast, but since the background contrast is assumed to be known, the ideal observer can be implemented by comparing the template response to a dynamic criterion based on background contrast (Figure 3.3A). Therefore, cognitively adjusting the criterion based on the background contrast is an optimal strategy (heuristic) for extending simple template-matching observer under target amplitude and background contrast uncertainty.

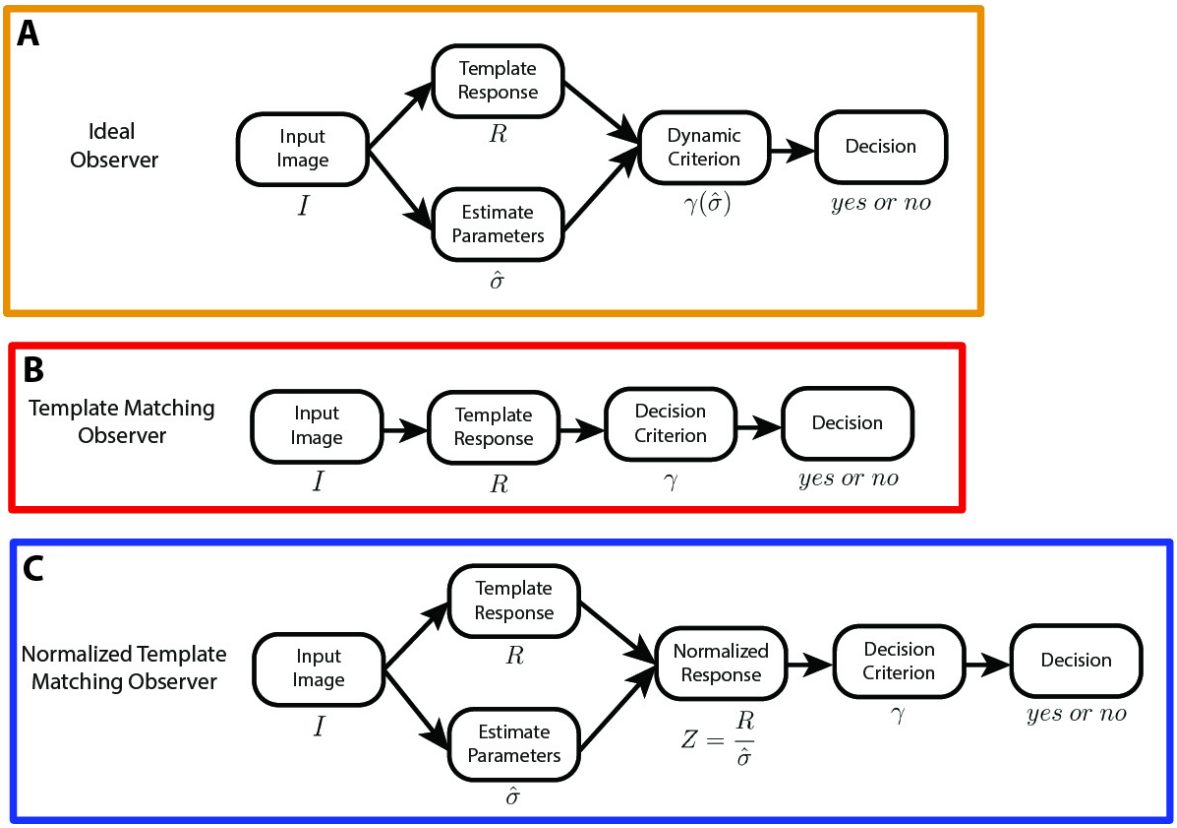


Figure 3.3 Three Observer Models.

A The ideal observer is implemented with template-matching and dynamic criterion. In each trial, the template response is computed by taking the dot product of the template and the target region of the image. Also, the background contrast is estimated, and the template response is compared to a dynamic criterion based on the estimated background contrast. **B** Simple template-matching observer. It compares template response to a single fixed criterion for all background contrast levels to produce a decision. **C** Normalized template-matching observer. It estimates the background contrast and calculates the standard deviation of template responses from it. Template response is normalized by the standard deviation of template responses for each background contrast level. The normalized response is compared to a single fixed criterion for all background contrast levels to produce a decision.

Instead of simulating model observers, I derived analytical expressions for hit and correct rejection rates to generate more precise predictions faster. I begin by deriving the hit (or correct rejection) rate for a given amplitude and background contrast pair because these pairs are disjoint and exhaustive. Thus, the overall hit (or correct rejection) rate is the

sum over amplitude and contrast levels with respect to their prior probabilities. For each amplitude and background contrast pair, in the case of white noise background and additive targets, the distribution of template responses ($R_b + a_i$, template response can be divided into response to background and response to target) is Gaussian. Means correspond to the amplitude levels of the target (a_i, 0 if the target is absent). Standard deviations (σ_j) are background contrasts (standard deviation of white noise) scaled by a constant. Note that the exact correspondence depends on how the template is defined in relation to the target (Appendix 1). A model observer makes a hit if the template response is higher than the criterion and the target is present. Therefore, the overall hit rate of the ideal observer can be derived by integrating the probability of template response exceeding the criterion for every target-present condition. For each background level, there is a unique criterion (γ_j).

$$\begin{aligned} p_h(\gamma_j) &= \sum_j \sum_i p(R_b + a_i > \gamma_j | a_i, \sigma_j) p(a_i) p(\sigma_j) \\ &= \sum_j \sum_i \Phi\left(\frac{a_i - \gamma_j}{\sigma_j}\right) p(a_i) p(\sigma_j) \end{aligned} \tag{3.1}$$

A model observer makes a correct rejection if the template response is lower than the criterion and the target is not present. The correct rejection rate of the ideal observer can be derived by integrating the probability of template responses that do not exceed the criterion for every target-absent condition. Since all target-absent distributions have an equal mean of zero, integration over the different contrast levels gives the correct rejection rate.

$$p_{cr}(\gamma_j) = \sum_j p(R_b < \gamma_j | 0, \sigma_j) p(\sigma_j) = \sum_j \Phi\left(\frac{\gamma_j}{\sigma_j}\right) p(\sigma_j) \tag{3.2}$$

The optimal criteria can be found via optimization using these equations.

3.2.1.2 Template-Matching Observer

Unlike the ideal observer, simple template-matching compares template responses to a single fixed criterion for all contrast levels. The hit and correct rejection rate of the template-matching observer can be expressed similarly, and the optimal single fixed criterion can be found via optimization (Figure 3.3B).

$$p_h(\gamma) = \sum_j \sum_i p\left(R_b + a_i > \gamma \middle| a_i, \sigma_j\right) p(a_i) p\left(\sigma_j\right)$$
$$= \sum_j \sum_i \Phi\left(\frac{a_i - \gamma}{\sigma_j}\right) p(a_i) p\left(\sigma_j\right) \tag{3.3}$$

$$p_{cr}(\gamma) = \sum_j p\left(R_b < \gamma \middle| 0, \sigma_j\right) p\left(\sigma_j\right) = \sum_j \Phi\left(\frac{\gamma}{\sigma_j}\right) p\left(\sigma_j\right) \tag{3.4}$$

3.2.1.3 Normalized Template-Matching Observer

The responses of neurons in the primary visual cortex are known to respond in a fashion consistent with contrast normalization (contrast gain control, Albrecht & Geisler, 1991; Heeger, 1992; Carandini & Heeger, 1994; Carandini et al., 1997; Cavanaugh et al., 2002). Under background contrast uncertainty, the normalization may allow a simpler cognitive strategy (a single fixed criterion) that reasonably approximates the ideal observer. Thus, I decided to incorporate contrast normalization into the template-matching observer. The NTM observer normalizes template response with the standard deviation of template responses (which is the background contrast scaled by a single constant). The normalized value is treated as a normalized template response $\left(Z_b + \frac{a_i}{\sigma_j}\right)$ and compared to a single fixed criterion (Figure 3.3C).

The standard deviations of the template responses (for all amplitude and contrast levels) are scaled to be one by the normalization. However, the mean is the initial mean

divided by the standard deviation of the template responses $\left(\frac{a_i}{\sigma_j}\right)$. Given the mean and standard deviation, the hit rate can be written in a similar form.

$$p_h(\gamma) = \sum_j \sum_i p\left(Z_b + \frac{a_i}{\sigma_j} > \gamma \middle| \frac{a_i}{\sigma_j}, 1\right) p(a_i) p(\sigma_j)$$
$$= \sum_j \sum_i \Phi\left(\frac{a_i}{\sigma_j} - \gamma\right) p(a_i) p(\sigma_j) \tag{3.5}$$

To calculate the correct rejection rate, integration over these equally likely identical normal Gaussian distributions can be simplified in a single-step computation.

$$p_{cr}(\gamma) = \sum_j p(Z_b < \gamma | 0,1) p(\sigma_j) = \Phi(\gamma) \tag{3.6}$$

3.3 Simulation Methods

The target is described in the experimental methods below. The ranges of target amplitudes and background contrasts were picked to give approximately 75 percent correct in the psychophysical experiment. Luminance is reported in gray level values so that the same metric is used for both amplitude and contrast. The amplitude of the target was randomly sampled from fifty, logarithmically spaced, and equally likely levels of target amplitudes between 4 and 19 gray levels (the amplitude is defined to be the maximum gray level of the presented target). The standard deviation of the Gaussian white noise was randomly sampled from fifty, linearly spaced, and equally likely levels between 5.12 and 12 gray levels. Corresponding RMS contrast of the backgrounds varied between 4 and 25 percent.

3.4 Simulation Results

I evaluated the performance of three model observers with the optimal criteria (picked to maximize the percentage correct) for various levels of amplitude range scalars. An amplitude range scalar is a single scalar value multiplied with all of the amplitudes in the amplitude range. I assumed that prior probability distributions of target amplitude and background contrast levels are uniform and the probability of the target being present is 0.5. However, I found that for the target prior of 0.5, the overall percentage correct of observer models is almost the same. It is because, for different levels of background contrast, the optimal criterion is approximately the same. Consider the left panels in Figure 3.4B, when there is only background contrast uncertainty. The distribution of template responses is shown in red when the target is absent and shown in yellow when the target is present. The template responses for the target-present case have a mean equal to the target amplitude and a standard deviation that increases with the background contrast. If the target prior is 0.5, the optimal criterion (that produces the highest percent correct) for all background contrast levels is exactly the same, and it is the target amplitude divided by two. However, this is not the case in general. If the target's prior probability is different than 0.5, the optimal criterion would be proportional to the standard deviation of template responses, and therefore proportional to background contrast (see right panels in Figure 3.4B). Consequently, one would expect that when the target prior is not 0.5, the predictions of the model observers will be distinguishable, and the TM observer will perform worse. Hence, along with various levels of amplitude range scalars, the target prior is also varied for the simulations. I found that for most of the low target priors, the NTM observer is better than the TM observer regardless of amplitude range scalar. However, the TM observer performs slightly better when both the amplitude range scalar and the target prior are high. More detailed results of these simulations can be found in Appendix 2.

Even though the target prior of 0.5 is efficient for the data collection in the laboratory environment, it is an unrealistic target prior under natural circumstances. Within a single fixation of the eye, the prior probability of a specific target being present in the fixated location is probably low. Therefore, I focused on low target priors and scaled the amplitude range so that three observer models are approximately 75 percent correct when the target prior is 0.5 (amplitude range scalar: 0.06). Under these conditions, the difference in performance between NTM and TM observers is the largest (Figure 3.4A). When the target prior is 0.2, the NTM observer is better than the TM observer by 3 percent correct. The relative performance measure is also defined, which quantifies the importance of percent correct differences. The relative performance measure is calculated by scaling the performance such that the range between best performance (given by the ideal observer) and baseline guessing performance (given by the prior probability) is scaled to range between one and zero. Therefore, the relative performance measure represents the fraction of the maximum performance increase from the chance that the model observer captures. When the target prior is 0.2, the NTM observer capture more than 99 percent of the maximum increase, and the TM observer captures 40 percent of the maximum increase.

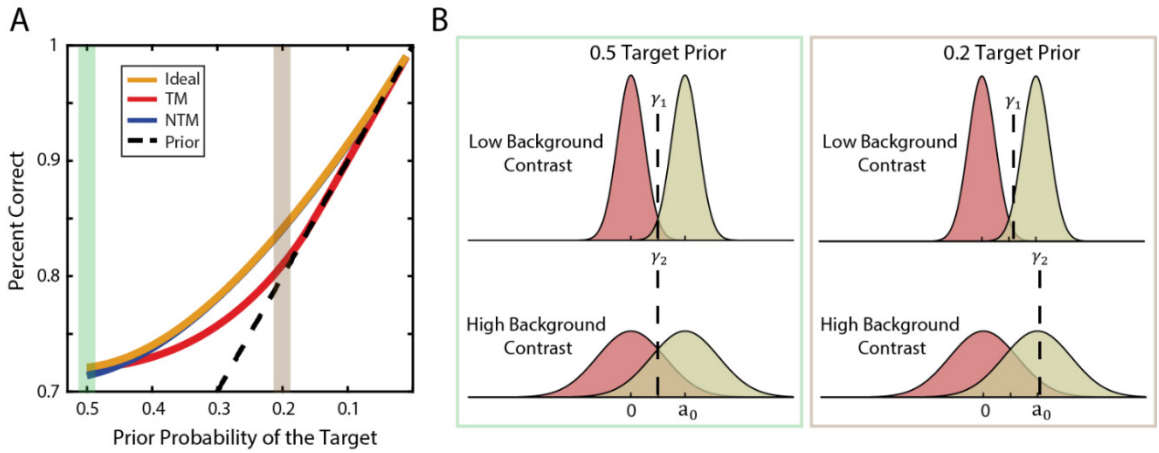


Figure 3.4 Simulation Results.

A The overall percent corrects of three observer models as a function of the prior probability of target being present. The amplitude range is scaled down with a single scalar that gives approximately seventy-five percent correct when the target prior is 0.5 (threshold efficiency). When the target prior is low, the NTM observer well approximates the ideal observer and performs better than TM observer that is almost only good as guessing based on prior probability only (shown with dashed a black line). **B** The illustration of the relationship between ideal criterion and target prior when there is only contrast uncertainty. The distribution of template responses when the target is absent is shown in red, whereas the yellow distribution corresponds to the case when the target is present. The mean of distributions when the target is present is equal to the target amplitude, and the standard deviation increases with background contrast linearly. When the target prior is 0.5, for all background contrast levels, the optimal criterion is exactly the same as shown in the first panel; however, this is only a special case. In general, the optimal criterion (γ) depends on the standard deviation of template responses, thus the background contrast.

The overall percentage correct in the uncertainty experiment is a single measure of the performance. Even if two model observers have similar overall percentage correct, their computations and thus hit and correct rejection rates might differ substantially. Examining these differences could be more efficient for model discrimination because it is relatively time-consuming to run the same uncertainty experiment for enough target prior levels to measure a curve like the one that is shown in Figure 3.4A. When the target prior is 0.5, and the model observers' efficiency is scaled to be at the threshold as in Figure 3.4A, I found that hit and correct rejection rates are fairly similar (Figure 3.5A). However, when the target prior is 0.2, TM observers' predictions are fairly different (Figure 3.5B). Thus, I

expect that I would be able to distinguish these three observer models when the target prior of 0.2, but most likely not able to distinguish them when the target prior is 0.5. Thus, I decided to run the psychophysical experiment for the two target-prior levels of 0.5 and 0.2.

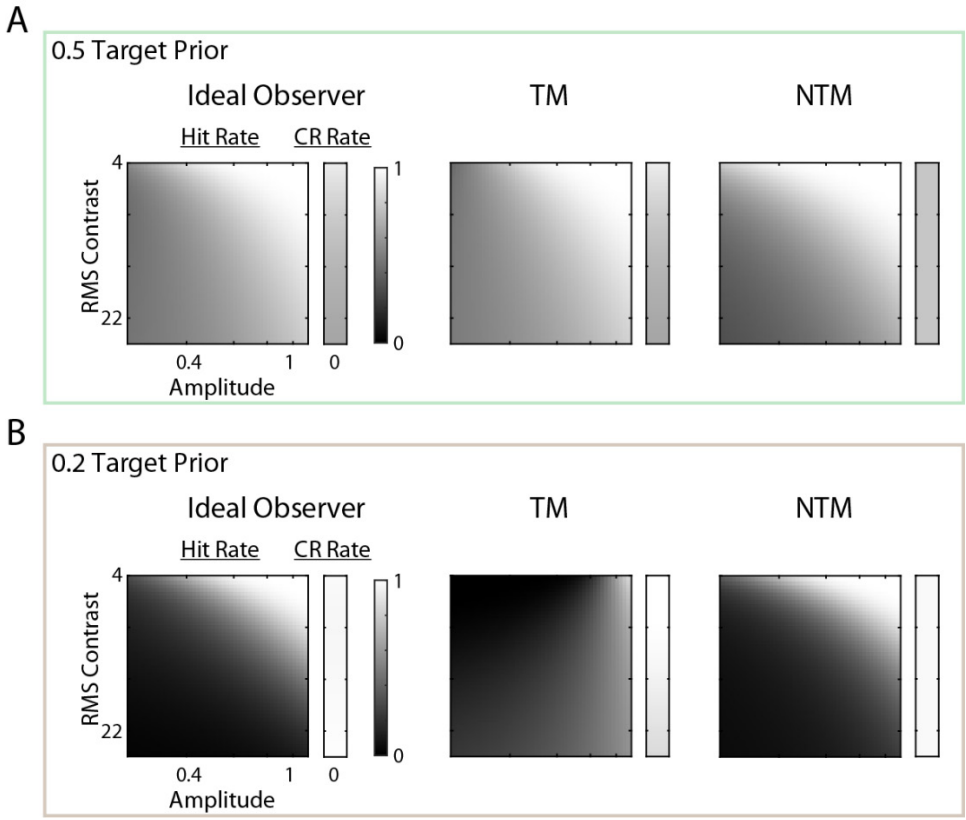


Figure 3.5 Hits and correct rejection rates of three observer models.

Model observers are simulated at threshold efficiency (as shown in Figure 3.4A). **A** Hits and correct rejection rates are shown when the target prior is 0.5 for fifty levels of target amplitude and background RMS contrast. The brightest white corresponds to a hundred percent hit (or correct rejection) rate, and the darkest black corresponds to zero percent. For each level of background contrast, there is a correct rejection rate shown in the smaller rectangle for each model observer. A fifty-by-fifty hit rate matrix is shown as the larger square for each observer model. Three observer models have similar hit and correct rejection rates. **B** Hits and correct rejection rates are shown when the target prior is 0.2. TM observer has lower correct rejection rates when background contrast is high, and the pattern of its hit rates is visibly different.

It is possible to evaluate the performance of three model observers when there is either low uncertainty or only single dimensions of uncertainty (modest uncertainty). When there is low uncertainty, these model observers are identical and optimal because template matching is optimal for additive targets in white noise. Under only target amplitude uncertainty, as I have shown, a single criterion is optimal; therefore, all observer models perform optimally and indistinguishably. Under only background contrast uncertainty, results are very similar to the combined case, except the difference between NTM observer and TM observer shrinks in favor of TM observer in general (at the threshold level, for target prior of 0.2, NTM is better than TM by 2.5 percent correct where NTM observer capture 94 percent of the maximum increase whereas TM observer captures 64 percent). Therefore, even though amplitude uncertainty on its own makes no difference, its combination with background contrast uncertainty further emphasizes the importance of contrast normalization. Therefore, one may expect that introducing more dimensions of uncertainty will even further emphasize the importance of contrast normalization. (See Appendix 2 for details).

Finally, the computation of the optimal performance allows us to characterize the effect of uncertainty and the effect of combining two uncertainty dimensions. In the case of target amplitude and background contrast dimensions, this analysis does not provide many insights because background contrast is assumed to be known (or can be precisely calculated given our stimuli), so there is no effect of background contrast uncertainty on the information (the ideal observer has the same performance regardless of uncertainty). The amplitude uncertainty has only a small effect, and this small effect is the same as the combined effect of uncertainty (on average, less than 1 percent correct difference and 11 percent of the maximum increase). If the target prior is low, the effect stays small (less than 1 percent correct and less than 10 percent of the maximum increase), but as the target prior

increases, the effect becomes bigger (up to 1.5 percent correct and 35 percent of the maximum increase). In other words, under low target prior, the ideal observer under combined two-dimensional uncertainty almost performs the same as the ideal observer under low uncertainty. Since NTM approximates the ideal observer under combined two-dimensional uncertainty, there is also only a negligible effect of amplitude uncertainty on the NTM observer when the target prior is low. (See Appendix 2 for details).

3.5 EXPERIMENTAL METHODS

3.5.1 Stimuli, Subjects and Procedure

The target was a 4-cycles-per-degree (cpd) horizontal sine wave, windowed with a radial raised-cosine function whose diameter is 0.8 visual degrees. (A visual degree was 120 pixels.) The background was a square patch of white noise whose side was 3 visual degrees (361 pixels). The target was presented at the center of the display together with the white-noise background.

The stimuli and the experimental conditions were adjusted for the psychophysical experiment. First, the white noise background consists of 4 x 4 pixel squares (1 x 1 minute), whose gray level was independently sampled from a Gaussian distribution[1]. Using a virtual pixel size of 1 x 1 minute approximately matched the noise energy to the optics of the eye, thus allowing the noise to have a larger effect on performance. Second, there were only eight levels of target amplitude and background contrast instead of fifty levels. I confirmed that computational results extend to this sparser sampling. The eight logarithmically spaced levels of target amplitudes were 4,5,6,8,10,12, 15, and 19 gray levels (with a mean gray level of 128). The eight level of background gray levels were 5.12, 8.96, 12.8, 16.64, 24.32

[1] The target is not downsampled to 4 x 4 pixel squares. Thus, the exact ideal observer would have exploited the difference between samplings. However, it is very unlikely that predictions of the exact ideal observer would be any different.

and 32 gray levels (corresponding RMS background contrasts are 4,7,10,13,16,19, 22, and 25 percent) (Figure 3.6A).

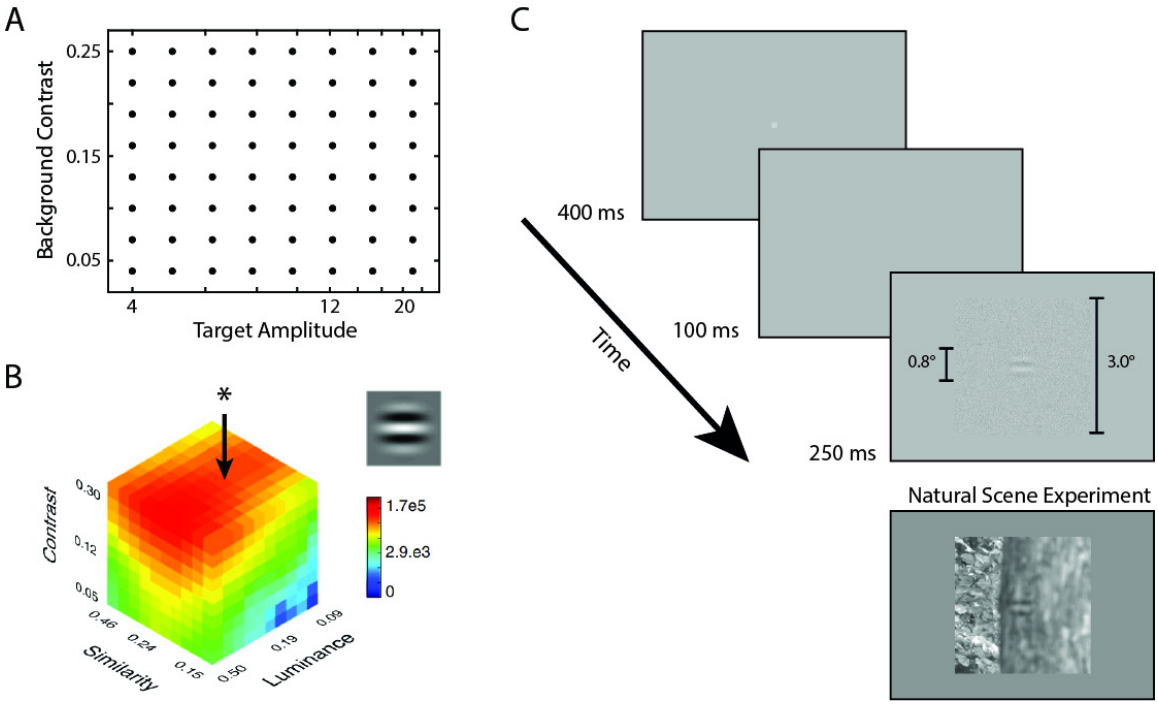


Figure 3.6 Experimental Methods.

A Target amplitude and background contrast levels in white noise experiment. For any target present trial, the stimulus's parameters are uniformly sampled from the grid of points. **B** The constrained sampling approach for the natural scene backgrounds (adapted from Sebastian et al., 2017). Natural scenes are binned along three dimensions: luminance, RMS contrast, and similarity to the target presented (cosine wave, but the measure was phase invariant, so it applied to a sine wave as well). Star and arrow point out the column containing eight bins I used in the experiment with different mean RMS contrast. The column corresponds to fixed median luminance and median similarity to the target. On each trial, a bin is randomly selected with uniform prior probability from eight bins to create background RMS contrast uncertainty. **C** The procedure of the experiment. 400 ms fixation screen is followed by a 100 ms black screen. The stimulus is shown for 250 ms with three visual degrees square background. In the natural scene experiment, the background gray level was lower, but the procedure was the same. The example target present stimulus presentation is shown.

I also conducted the same experiment with natural scene backgrounds. Sebastian et al. (2017) binned millions of natural-image patches along three dimensions: mean

luminance, RMS contrast, and similarity to the target, all computed from the target region. On average, every natural background in a bin has almost the same luminance, similarity to the target, and contrast (with minor variations). I created RMS background contrast uncertainty by sampling across eight bins, each with a different mean RMS contrast. However, the luminance (equal to the background gray level) and similarity were held at their median value (6th luminance bin, 5th similarity bin out of ten bins). The backgrounds had a mean gray level of 59. The standard deviation of the gray level of bins in the experiment were 3.25, 3.91, 6.83, 8.22, 9.90, 11.92, 14.35, and 17.28 gray levels (the corresponding RMS contrasts were 6, 7, 12, 14, 17, 20, 24, and 29 percent) (Figure 3.6B). Target amplitudes were adjusted to expect performance approximately around 75 percent correct, and they were 2, 3, 4, 5, 7, 9, 13, and 17 gray levels. Each background was sampled from appropriate bins without replacement, so participants never saw the same background twice. I made sure that the target with the smallest amplitude contained at least four different discrete gray levels and that the discretization has negligible effects on model observers' discriminability.

Four experienced observers completed the experiment with the white noise backgrounds, and three of them completed the experiment with natural scene backgrounds. They had normal (corrected) spatial acuity. Written, informed consent was obtained for all observers in accordance with The University of Texas at Austin Institutional Review Board. The stimuli were presented at a distance of 171.1 cm with a resolution of 120 pixels per degree. In the white noise experiment, the mean luminance of the background was 44.87 cd m^2. For natural backgrounds, the mean luminance was 20.7 cd m^2. Images are gamma-corrected based on the calibration of the display device (GDM-FW900; Sony) and quantized to 256 gray levels. The screen refresh rate was 85 hertz. All experiments and

analyses were done using custom code written in MATLAB, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

The procedure in both experiments was the same. Participants were asked to press a key to deliver their decision about whether the target is present or absent in each trial. Each trial starts with a dim fixation point at the center, presented for 400 ms, followed by a 100 ms blank window. The stimulus was presented for 250 ms (Figure 3.6C). Participants were given one second to indicate their decision. The subsequent trial began after an additional 500 ms. The stimulus presentation time was selected to match the average human fixation duration during a visual search. A High-frequency feedback tone was provided if the participant's response was correct, and a low-frequency if the response was incorrect responses. On each trial, the target amplitude and the background contrast were sampled randomly from the 64 possible combinations (see Figure 3.6A); however, the sampling was counter-balanced, so there was the same number of trials for each condition.

In each block of the experiment, the target's prior probability was either 0.5 or 0.2, and participants were informed before starting the block. Every participant completed three blocks for each target prior condition (one participant only joined the white noise experiment and completed two blocks). The order of the blocks was randomized. In a single block, there was a total of 640 trials. When the target prior probability was 0.5, the number of trials for each unique combination of contrast and amplitude level (total of 64 conditions in which the target is present) was 5. When the target prior probability was 0.2, it was 2.

3.5.2 Analysis Methods

The simulation methods were adjusted to changes I made for the psychophysical experiment. The background contrast and target amplitude levels were the same as those used in the psychophysical experiment (only eight levels each). For 4 x 4 pixel white noise,

the relationship between target amplitude and the mean of template responses was estimated by computing template responses for numerous samples of 4 x 4 pixel noise. I found that the mean of the distribution equals to target amplitude (as expected). The exact estimation was done to determine the relationship between the standard deviation of white noise and the standard deviation of template responses. Even though the standard deviation of templates responses is still a scaled version of the standard deviation of noise by a single constant (0.14), the constant is approximately four times the one analytically derived for a single pixel white noise. These estimated relationships were used to calculate the percentage correct with the equations provided above.

All three models were fitted to the data by estimating two independent parameters separately for each target prior probability condition and participant. These parameters are the criterion, which accounts for sub-optimal decision strategies, and the scaling parameter on the amplitudes, which allows for an overall efficiency difference between models and participants. I decided to fit the correct rejection rates (for the different background contrast levels) and hit rates (for the different pairs of background contrast and target amplitude levels) using a maximum-likelihood method. In each trial, the participant's decision is either a hit, false alarm, correct rejection, or miss (depends on whether the target is presented or not). Thus, all data can be written as a string of hits, false alarms, correct rejections, and misses. Model observers predict the probabilities of these four categories as their hit and correct rejection rates, which are determined by only two parameters. The probability of the entire data is the product of the corresponding probabilities of each response (or element in the string of data). This method is a variation of the classic maximum-likelihood method for fitting a psychometric curve to the data. I extend it for the data collected in the uncertainty experiment because it better deals with properties of the

data (number of trials being different for different conditions and high or low rates likely to be more consistent) than standard metrics like a squared error.

For natural scene backgrounds, three observer models' hit and false alarm rates must be calculated. I calculated the distribution of template responses for each natural scene bin used in the experiment for target present and absent conditions. I found that the distribution of template responses is well approximated with a Gaussian distribution consistent with earlier findings (Sebastian et al.,2017). The mean of these distributions is target amplitude, and the standard deviation is scaled background RMS contrast with a single scalar (0.13). Since these are Gaussians with known means and standard deviations, the same equations can be used to integrate the distributions, so the same fitting procedure can be applied.

3.6. Experimental Results

The fitting procedure allows model observers to use non-optimal criteria to fit the data. When criteria are not optimal, the ideal observer's dynamic criterion implementation will not reach the optimal performance, so calling it "ideal observer" will not be exactly correct. To avoid confusion, the fitted model will be called dynamic template-matching (DTM). For most figures, results for the average participant are shown (the average participant data generated by aggregating all participants' data).

3.6.1 White Noise Experiment

Correct rejection and hit rates of the average participant are shown in the left panel of Figure 3.7A for the target prior of 0.5 and in the left panel of Figure 3.7B for the target prior of 0.2. The vertical axis is the RMS contrast of the background, and the horizontal axis is the target amplitude. The grayscale color represents the hit or correct rejection rate,

where the brightest white corresponds to one, and the darkest black corresponds to 0. Model predictions are shown in the three right panels. The numbers above each panel show the negative log likelihood of the data given the model and the numbers in the brackets show the 68% confidence interval. For both target prior levels, NTM and DTM observers better explain the data than the TM observer. Note that fitted TM observers' predictions are qualitatively different from predictions shown in Figure 3.5 when the target prior is 0.2. Specifically, the orientation of iso hit rate contour is different. The difference occurs because the best-fitted criterion is very non-optimal, as shown later (see Figure 3.9); whereas the predictions in Figure 3.5 are for an optimal criterion. This result also reveals that correct rejection and hit rates are sufficient to differentiate between observer models. Negative log likelihoods associated with the fits to individual participants' data are provided in Figure 3.8, together with the estimated scaling factor (to scale down amplitudes). The confidence intervals show that the TM observer consistently fails to explain the data as well as NTM and DTM across participants (except for the first participant when the target prior is 0.2).

The estimated scaling factors mostly fall in the range of 0.4 and 0.35 and are relatively stable across participants and the target prior conditions. The scaling factor estimate for the average participant is shown with the "Av." label.

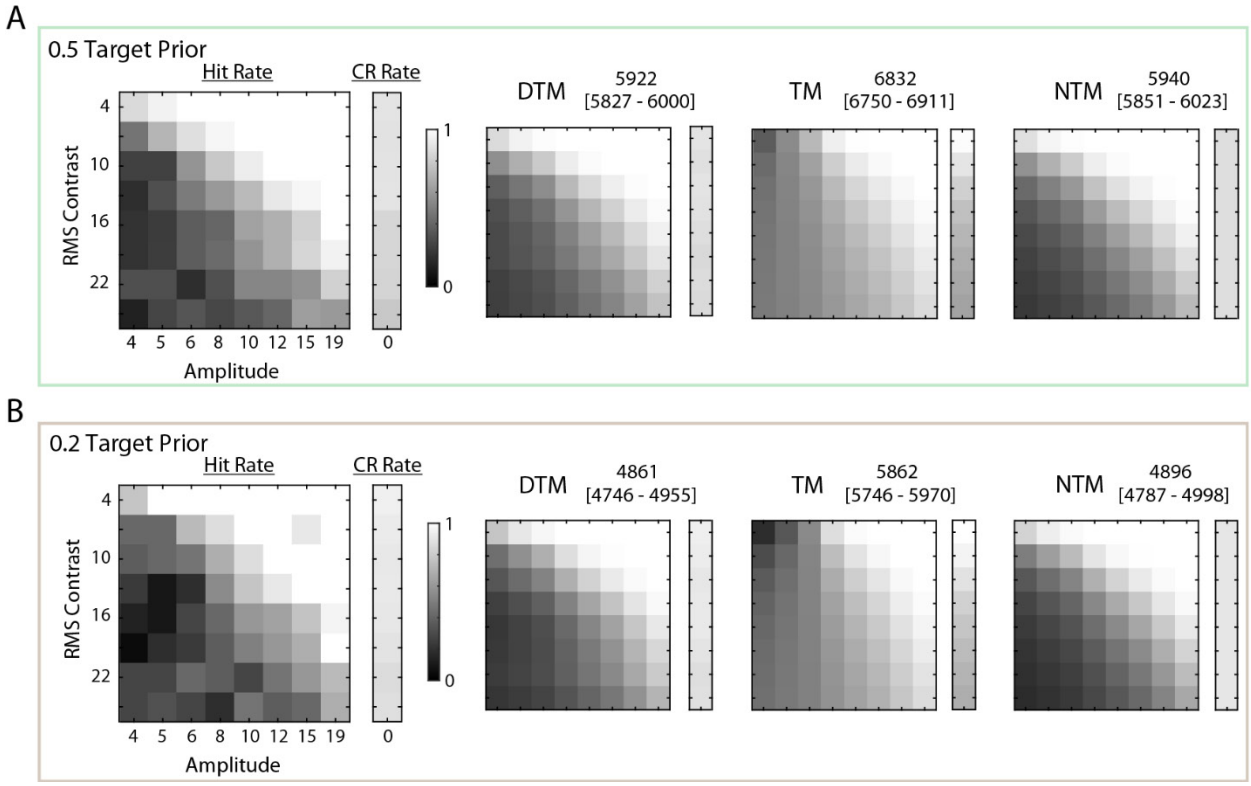


Figure 3.7 Results of white noise experiment.

A The data matrix is an eight-by-eight hit rate matrix of the average participant followed by a column of correct rejection rates when the prior probability of the target present is 0.5. The average participant data generated by aggregating the data of every participant to one. Grayscale is used to cover rates between 1 (white) and 0 (black). Best fitted observer models' hit and correct rejection rates are presented in the same structure with negative log likelihood associated with the fit and 68 percent confidence intervals around it. **B** The data matrix of the average participant when target prior is 0.2. Best fitted observer models' hit and correct rejection rates with negative log likelihood associated with the fit and 68 percent confidence intervals around it.

The scaling factors estimated for the DTM and NTM observers are very similar, but the best fitted scaling factors for the TM observer are slightly smaller. The small variation in scaling factors shows that it indeed acts as an overall efficiency difference between human participants and model observers. Thus, the scaling factor has a negligible impact on the pattern of correct rejection and hit rates. Especially, the estimated scaling

factor for the average participant is almost identical across target priors; thus, if a single scaling factor for both priors were used to fit DTM and NTM, results would be almost identical.

The other parameters estimated for each fit are the decision criteria, and they are shown in Figure 3.9 for the average participant. Recall that when the target prior is 0.5, the ideal observer and TM observer's performance was similar (when the criteria are selected to maximize percentage correct) because the optimal criterion for different contrasts is similar (Figure 3.5). However, I found that the TM observer fails to explain the data compared to DTM and NTM (Figure 3.7, 3.8), suggesting that the estimated criterion of DTM for each contrast level should be different, as shown in Figure 3.9A. The optimal criteria (in dashed yellow) are also shown in Figure 3.9A (for the estimated scaling factor), and as expected, it is much flatter compared to the best fitting criteria. The sub-optimal criteria have only a minor effect on the overall percent correct of the DTM observer compared to the true ideal observer (DTM with optimal criteria) for target prior 0.5 (0.77 vs. 0.78). However, the effect is much larger when the target prior is 0.2 (0.845 vs. 0.868). The difference is also very similar for sub-optimal observers (for target prior 0.5, NTM: 0.76 vs. 0.76, TM: 0.74 vs. 0.74; for target prior 0.2, NTM: 0.84 vs. 0.87, TM: 0.79 vs. 0.83). Note that predictions of the TM and DTM observer, when criteria were picked to maximize overall percent correct, are similar for target prior 0.5 (as shown in Figure 3.5). However, the DTM observer with sub-optimal criteria explained the data much better than the TM observer with sub-optimal criteria. Thus, the DTM observer is more flexible in fitting different patterns of hit and correct rejection rates than the TM observer because it has more parameters to vary (i.e., the criteria; see Figure 3.7).

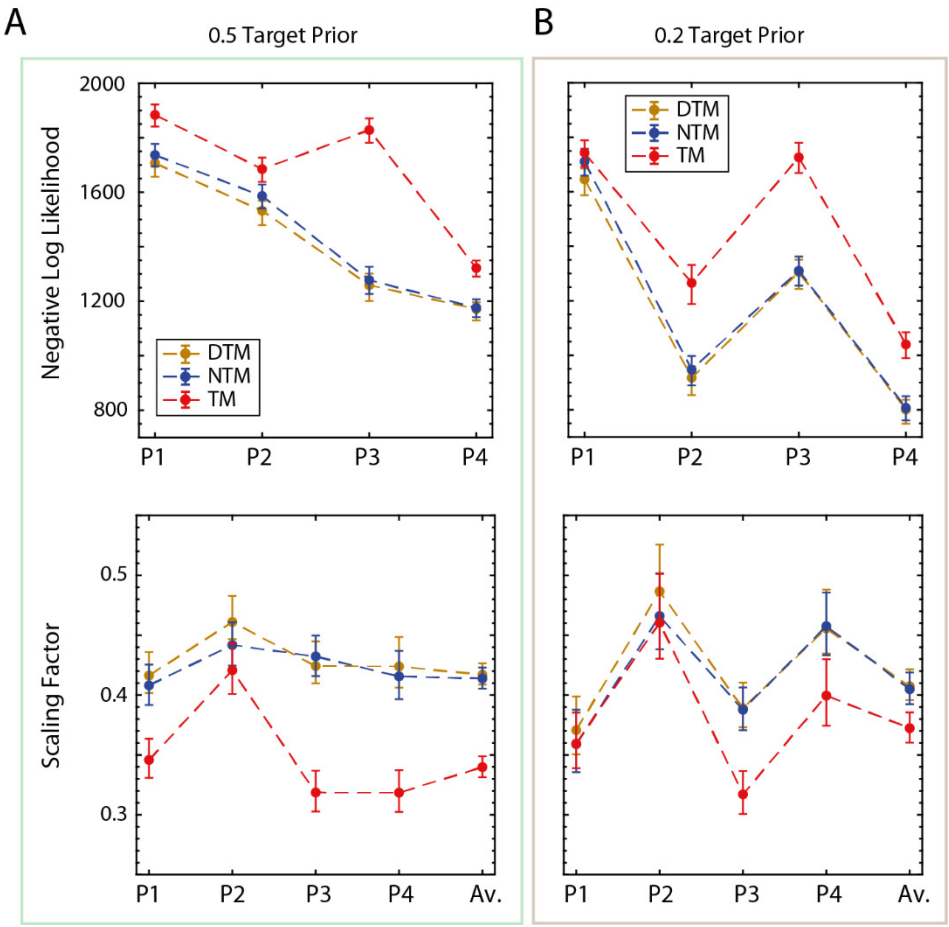


Figure 3.8 Goodness of fits and scale factors.

A Goodness of fits measure (negative log likelihood) and estimated scaling factor of each fit for individual participants when the target prior is 0.5. "Av." is used to label the average participant. Error-bars show the 68 percent confidence intervals. **B** Goodness of fits measure (negative log likelihood) and estimated scaling factor for each fit of individual participants when the target prior is 0.2.

Interestingly, the slope of estimated criteria for the DTM observer is similar for the target prior of 0.5 and 0.2, even though the ideal criteria are quite different. The estimated criteria's slope is also similar across target priors for the individual participants.

Gray lines in Figure 3.9 show the optimal criteria if there is only contrast uncertainty and the target amplitude is fixed to its minimum value in the amplitude range

used in the experiment. As already demonstrated in Figure 3.4B, when the target prior is 0.5, the criterion is the same for all background contrast levels. The amplitude uncertainty generates minor deviations from the flat criteria (the optimal criteria shown in dashed yellow is not flat). As I discussed earlier, the amplitude uncertainty, when it is combined with contrast uncertainty, has a small effect on the comparison of the observer models. However, in general, it further emphasizes the difference between TM and other observer models.

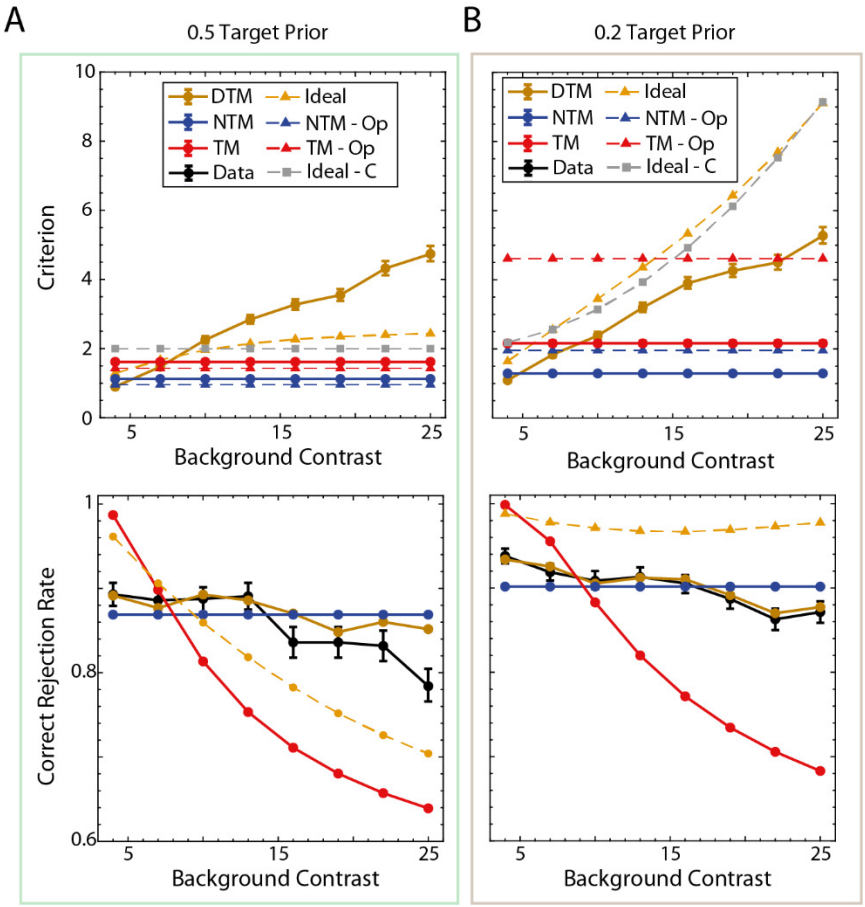


Figure 3.9 Criteria and correct rejection rates.

A Estimated criteria for three observer models associated with fits to the average participant and correct rejection rates when the target prior is 0.5. Error-bars show the 68 percent confidence intervals. For each model observer, given the estimate scale factor, optimal criteria are shown in dashed lines (DTM becomes the ideal observer with optimal criteria shown in yellow). The Gray dashed line indicates the optimal criteria when there is only contrast uncertainty and amplitude is the smallest of the range that was used in the experiment. Correct rejection rates of the average participant is shown in black with 68 percent confidence intervals. **B** Estimate criteria for three observer models associated with fits to the average participant and correct rejection rates when the target prior is 0.2.

The difference between the three observer models is best illustrated by the correct rejection rates for the different levels of background contrast. Correct rejection rates are an immediate result of the criteria since all distributions of template responses for target-absent conditions sit at zero. Figure 3.9 shows that the average participant's correct

rejection rates (shown in black) decrease slowly and are close to flat. However, the fitted TM observer fails to replicate this pattern. Even though the NTM observer has flat correct rejection rates, its predictions are still quite good, and explain the data just about as well as the DTM observer (especially given the greater number of free parameters in the DTM observer). The ideal observer's correct rejection rates are shown in yellow. Note that when the prior is 0.5, it is similar to the TM observer in terms of the slope, but when the target prior is 0.2, it is much flatter, and thus similar to the NTM observer.

Both NTM and DTM observers explain the data equally well. However, DTM uses eight criteria parameters, whereas NTM only uses one. I calculated standard model selection measures to penalize for the number of parameters and found that, for Bayesian information criteria (BIC), the NTM is slightly better than DTM on average, but the confidence intervals mostly overlap, so they still explain the data equally well. I also checked whether using dynamic criterion after the normalization would improve fits of the NTM observer. However, adding the dynamical criterion makes the NTM the same as DTM and provides no extra explanatory power. No explanatory power gain was also apparent from the estimated criteria for the dynamic criterion NTM observer because it only has a small variation (almost completely flat criteria). Lastly, I compare the BIC of three observer models to BIC calculated with an empirical model that uses the empirical correct rejection and hit rates (8+64 parameters). I found that for the average participant, the TM observer is worse than the empirical model; however, both DTM and NTM are better than the empirical model.

3.6.2 Natural Scene Experiment

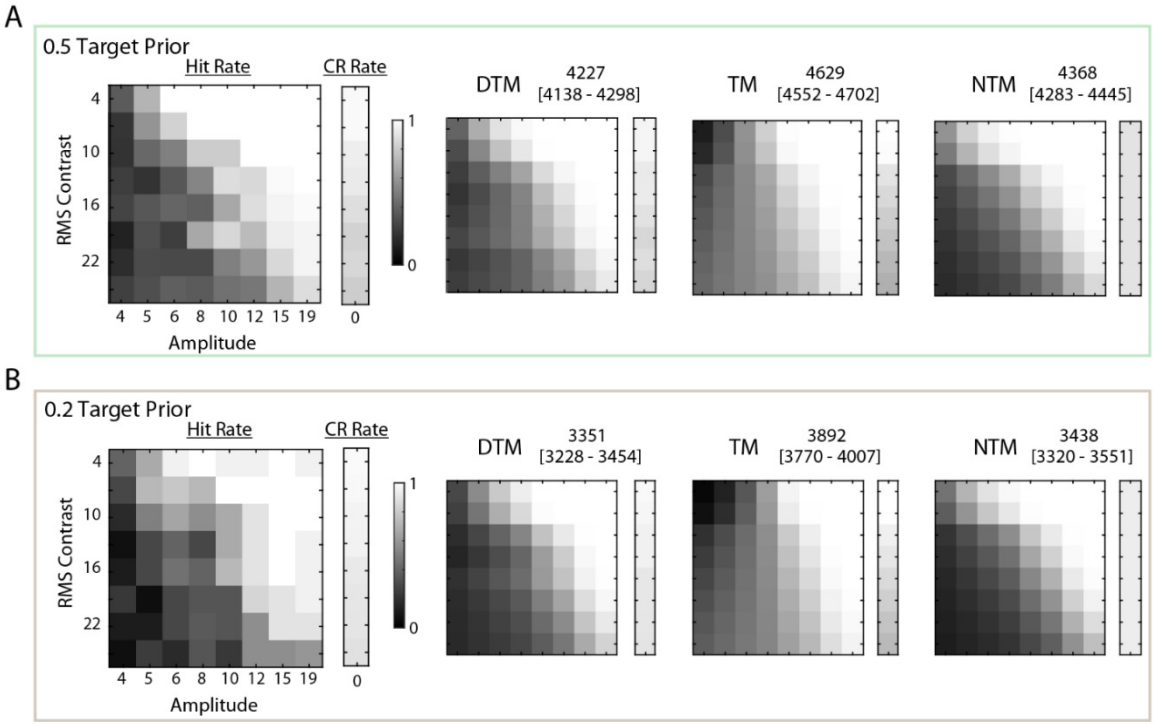


Figure 3.10 Results of natural scene experiment.

A The data matrix of the average participant when target prior is 0.5 for the natural scene background experiment. Best fitted observer models' hit and correct rejection rates with negative log likelihood associated with the fit and 68 percent confidence intervals around it. **B** The data matrix of the average participant when target prior is 0.2 for the natural scene background experiment. Best fitted observer models' hit and correct rejection rates with negative log likelihood associated with the fit and 68 percent confidence intervals around it.

For natural scene backgrounds, the results are almost the same. The data for the average participant is shown in Figure 3.10 together with model fits. The hit and correct rejection pattern is slightly changed; however, DTM and NTM observers better fit the data than the TM observer for both target prior conditions. The effect is consistent across participants (Figure 3.11), but this time for the target prior of 0.2, the effect is more apparent. In contrast, when the target prior is 0.5, the NTM observer also seems to explain the data slightly worse than the DTM observer (for the first participant, it is even worse

than the TM observer). The estimated scaling factors have more variability and are slightly higher than the white noise experiment, which is not surprising considering that template matching for natural scenes is not the ideal strategy (Figure 3.11). Also, the TM observer has very similar estimated scaling factors to those of the DTM and NTM observers, unlike in the white noise experiment. The estimated scaling factors are similar across different models and target priors, so they might be only accounting for the overall efficiency difference between human and model observers. Similar to the white noise experiment, the estimated scaling factor of the average participant is almost identical across target priors. Thus, the fits would be almost identical if a single scaling factor was used for both conditions to fit DTM and NTM.

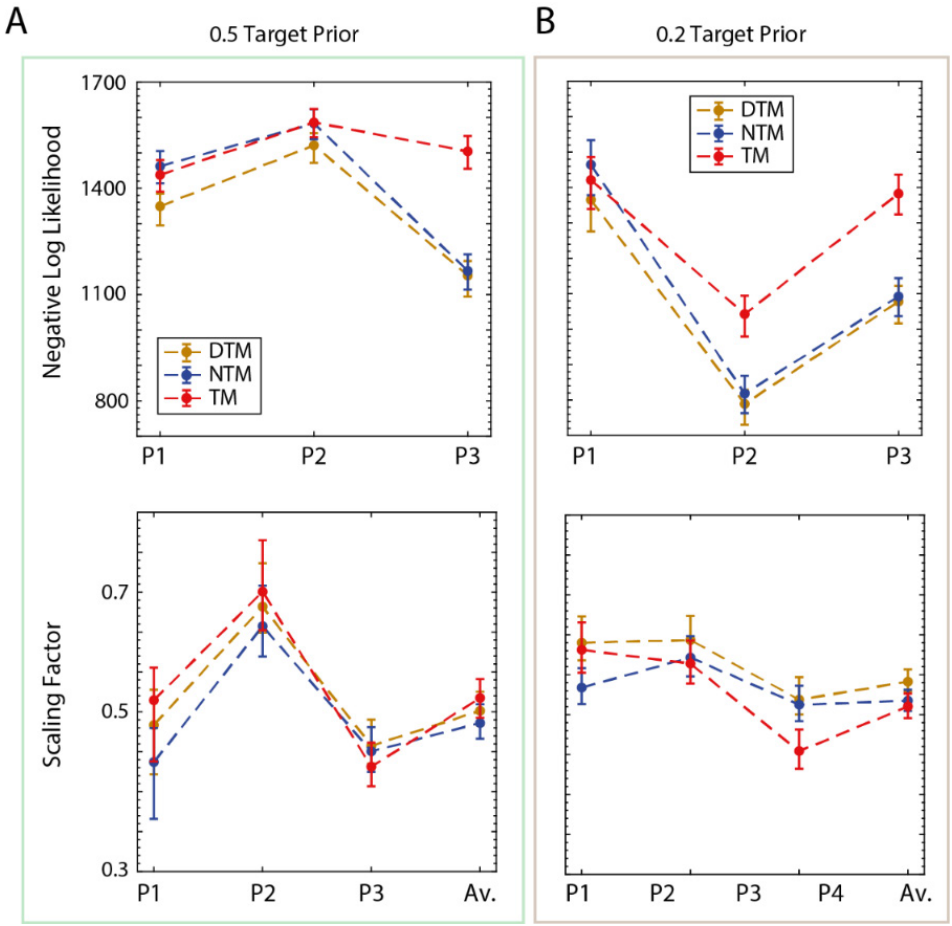


Figure 3.11 Goodness of fits and scale factors.

A Goodness of fits measure (negative log likelihood) and estimated scaling factor of each fit for individual participants when the target prior is 0.5 for the natural scene background experiment. "Av." is used to label average participant. Error-bars show the 68 percent confidence intervals. **B** Goodness of fits measure (negative log likelihood) and estimated scaling factor for each fit of individual participants when the target prior is 0.2 for the natural scene background experiment.

The estimated criteria are shown for the average participant (Figure 3.12). Especially for the target prior 0.5, the estimate criteria of DTM are flatter compared to the white noise experiment, which is why the data favors the TM observer a little more on average. However, the optimal dynamic criteria shown in yellow are flatter than the estimated ones, so the TM observer fit to the data is worse overall. The slope of the DTM criteria changes only slightly for the target prior 0.2. It is much flatter than the optimal

criteria, so the differentiation between TM and other observers is relatively similar across target priors. The effect of sub-optimal criteria is minor on the overall percent correct of the DTM observer compared to the true ideal observer (DTM with optimal criteria) for target prior 0.5 (0.78 vs. 0.78). When the target prior is 0.2, the effect is slightly larger (0.86 vs. 0.87). The difference is also very similar for sub-optimal observers (for target prior 0.5, NTM: 0.77 vs. 0.77, TM: 0.77 vs. 0.77; for target prior 0.2, NTM: 0.86 vs. 0.87, TM: 0.83 vs. 0.85). The gray lines show the optimal criteria when there is only contrast uncertainty, and the target amplitude is the smallest (gray level of 2). It provides an example of how amplitude uncertainty affects the optimal criteria when combined with contrast uncertainty. The effect is small but, in general, works against the TM observer.

The correct rejection rates illustrate why the TM observer fails to explain the data as well as NTM and DTM observers (Figure 3.12). However, especially when the target prior is 0.5, the correct rejection falls more rapidly than in the white noise experiment. A steeper slope is the primary reason why there is a trend in this experiment for the DTM observer to better explain the data than the NTM observer (whose correct rejection rate is always flat). I calculated BIC and found that NTM and DTM are indistinguishable. However, NTM and DTM observers do better than the default model, which uses the empirical correct rejection and hit rates to predict the data.

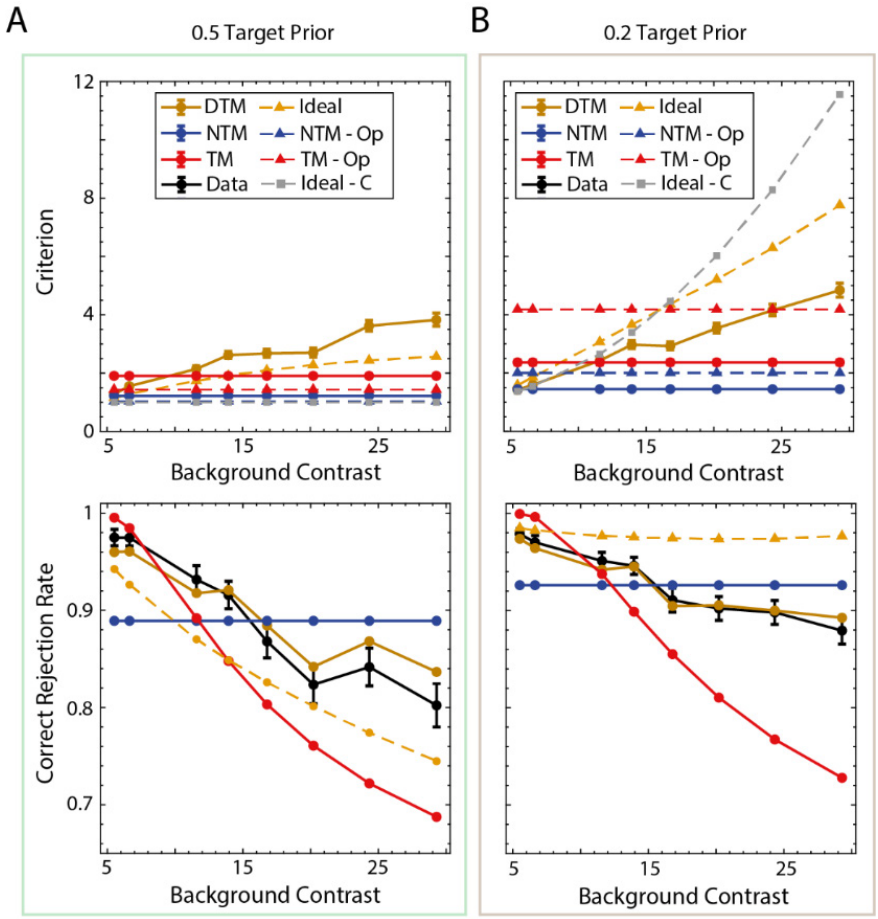


Figure 3.12 Criteria and correct rejection rates.

A Estimated criteria for three observer models associated with fits to the average participant and correct rejection rates when the target prior is 0.5 for the natural scene background experiment. Error-bars show the 68 percent confidence intervals. For each model observer, given the estimate scale factor, optimal criteria are shown in dashed lines (DTM becomes the ideal observer with optimal criteria shown in yellow). The Gray dashed line indicates the optimal criteria when there is only contrast uncertainty and amplitude is the smallest of the range that was used in the experiment. Correct rejection rates of the average participant is shown in black with 68 percent confidence intervals. **B** Estimate criteria for three observer models associated with fits to the average participant and correct rejection rates when the target prior is 0.2 for the natural scene background experiment.

3.7. SUMMARY

I have derived the ideal observer under target amplitude, and background contrast uncertainty for additive targets in white noise backgrounds. The ideal observer can be implemented as an extension of the classic template-matching observer, which is ideal when there is low uncertainty. Target amplitude uncertainty is optimally handled by appropriately placing the criterion. For background contrast uncertainty, having different criteria for different levels of estimated background contrast is an optimal computational strategy. Three observer models are evaluated: the classic template-matching observer (TM), normalized template-matching observer (NTM), and the ideal observer (which is implemented via dynamic-criterion template-matching observer, DTM). Analytical equations are introduced to calculate the performance of the three model observers. I found that observer models' performance is similar when the prior probability of target is 0.5. However, when the target prior is low, NTM is better than TM and approximates the ideal performance (DTM). The three observer models differ in their hit and correct rejection rates patterns when the overall performance is around seventy-five percent correct.

I measured the human performance under high levels of target amplitude and background contrast uncertainty for two levels of the target prior (0.2 and 0.5). I introduced a variation of classic likelihood fitting to fit hit and correct rejection levels of human data. Under high levels of uncertainty, I showed that NTM and DTM observers explain the data for both target prior levels, whereas TM observer fails to explain the data. The human detection performance was also measured for natural background under high levels of target amplitude and background contrast uncertainty. Similarly, NTM and DTM observers explain the data for both target prior levels, whereas TM observer fails to explain the data.

Chapter 4: Simultaneous Target Scale and Target Orientation Uncertainty

4.1 MOTIVATION

While searching for a target in the real world, the visual system is always almost uncertain about the multiple properties of the target simultaneously. This chapter is concerned with two stimulus properties (dimensions): the orientation of the target and the scale of the target. In a 3D environment, the target's scale is determined by the relation between the observer's 3D position and orientation (head pose) and the target's position in depth. Similarly, the 2D orientation is determined by the relation between the observer and the 3D orientation of the target. Under naturalistic conditions, the observer's 3D position and orientation are rarely the same, so the visual system rarely observes the same scene from the same viewing angle. Moreover, the 3D orientation and position of a target are rarely the same. Consequently, there is almost always simultaneous uncertainty about both the target's scale and 2D orientation.

As in the previous chapter, I will begin with the derivation of the ideal and sub-ideal observers for white noise backgrounds and additive targets. I will describe simulations of these models to understand the effect of various computational principles on the performance and identify the differences between the models. Both dimensions of interest produce uncertainty about the visual pattern, generating a set of visual patterns that might show up as the target. The ideal observer derivation is not concerned with a specific set of visual patterns; thus, the ideal observer applies to any set of uncertainty dimensions that produce uncertainty about the visual patterns. The ideal observer can be formulated in the classic template-matching observer framework by properly pooling the template responses (for each different visual pattern, there is a different template). A well-known heuristic operation for pooling template responses is called the maximum output rule,

which simply selects the maximum template response. Making decisions based on the maximum template response alone approximates the ideal observer under simple scenarios (Nolte & Jaarsma, 1967; Pelli, 1985; Graham et al., 1987) and explains the human visual detection performance under uncertainty (Davis et al., 1983; Eckstein et al., 2000). However, when the discriminability of visual patterns is different, the simple max rule might not approximate the ideal. If some templates have considerably higher energy than others, those template responses might be high enough to drive a target-present response even when the target is not present. One method to avoid the problem is normalizing either the template or template responses. I will test various normalizations to see which is the best strategy. In the light of the simulation results of these computational principles, I will then describe a psychophysical experiment under the high-extrinsic pattern uncertainty.

The first major aim of this chapter is to measure and model the human ability to detect targets in white noise when 2D target orientation and target scale are randomly varying over a wide range. When the target's scale changes (for example, when the target's position in-depth changes), both target's spatial frequency and size change in relation to each other. There are no published studies on scale uncertainty, but there are some published studies on the two components: spatial frequency and size. Human performance has been measured for target spatial frequency uncertainty and target size uncertainty while keeping the other fixed. Moreover, in these studies, the discriminability of different target levels (sizes or spatial frequencies) was equated by adjusting amplitudes for different targets. They found that the effect of size uncertainty in isolation is smaller than theoretically expected based on the data measured under low extrinsic uncertainty. In contrast, the effect of spatial frequency uncertainty in isolation seems to be consistent with the expected effect, even though it is somewhat controversial. (See Chapter 2 for more detailed discussions of the literature about the size and spatial frequency uncertainty.)

Together, it is unclear what these results predict about the target scale uncertainty. There are only a few studies on the effect of orientation uncertainty. These studies found that it has either no effect or a very small effect on human performance (see Chapter 2). However, these measurements were not compared to theoretical predictions. To the best of my knowledge, the current study is the first to examine the combined effect of these dimensions of uncertainty. I will measure the human performance and compare the predictions of the model observers with the psychophysical data measured in white noise backgrounds.

The second aim of this chapter is to measure human detection performance in a blocked experiment with the same orientation and scale levels. In a single block, scale and orientation are fixed, so; there is no uncertainty about orientation and scale (low extrinsic uncertainty experiment). I will describe the psychophysical experiment, its results, and its relation to the high extrinsic uncertainty experiment. It is common to generate predictions for high extrinsic uncertainty experiments based on data measured in low extrinsic uncertainty experiments (see Chapter 2). Similarly, I will generate predictions for various model observers and test them against the data measured under high uncertainty. The generated predictions provide a test of whether human observers are similarly efficient for both conditions. Since I derived the ideal observer for both experiments, I will also directly compare the efficiency of human observers. The efficiency of human observers under both conditions provides important insight into the computational principles underlying the relationship between various levels of uncertainty. Human efficiency could increase, decrease, or stay the same with the introduction of high levels of uncertainty. These are all associated with computational principles discussed in Chapter 2. Finally, I will test whether these computational principles can explain the efficiency results.

4.2 Theory of Detection Under Uncertainty About the Visual Pattern (Multiple Templates)

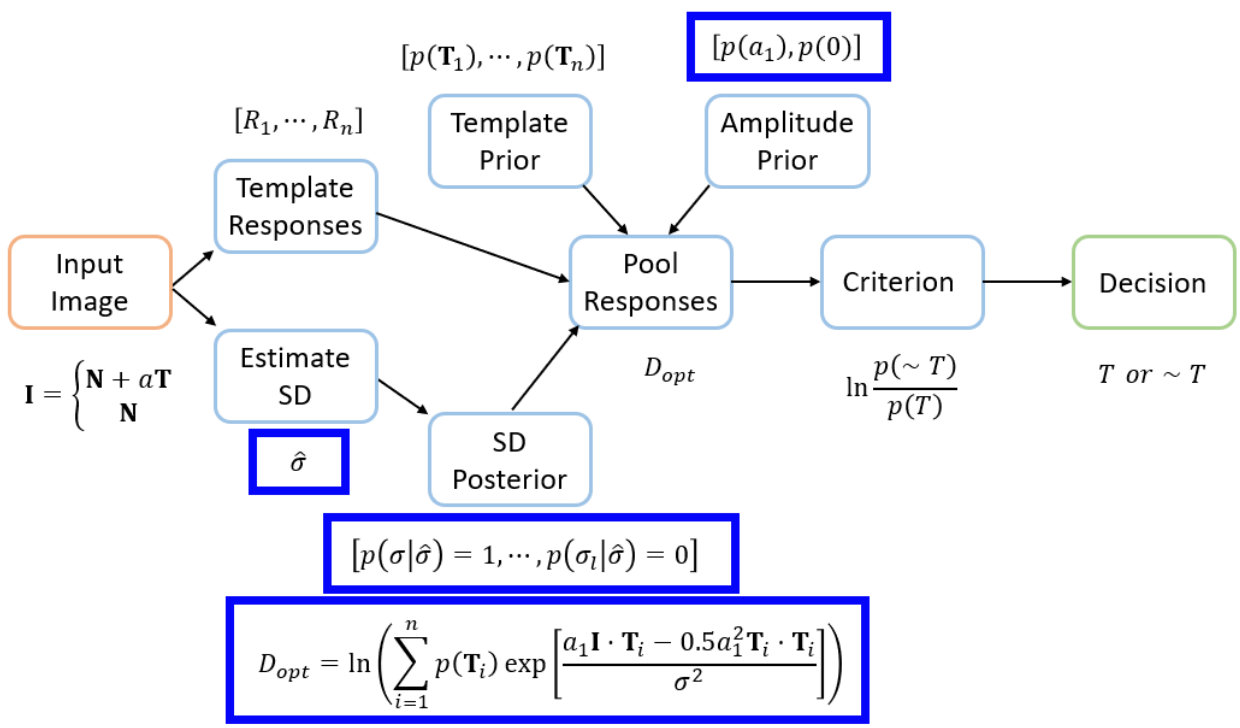


Figure 4.1 The simplification of the ideal observer.

For the theory of uncertainty about the visual pattern, the ideal observer for detecting an additive 2D object in the environment.

Recall that the target is normalized to a peak of one, and the amplitude is defined as the single scale factor multiplying the target. The gray level of each background pixel was independently sampled from a Gaussian distribution with zero mean and non-zero standard deviation (white noise background). (Because the mean gray level was fixed in the experiment, it can be set to zero in the simulations without loss of generality.) In each trial, a new sample of white noise background is presented. For each pixel, the gray level

of the background was added to the gray level of the target. The general case introduced in Chapter 2 is simplified here because there is no uncertainty about target amplitude and background contrast for the theory of uncertainty about the visual pattern. Thus, the prior probability of the known background contrast and target amplitude is one (Figure 4.1). For target present trials, the visual pattern (template/ target) is randomly sampled from a pre-defined range of visual patterns with a stationary probability distribution (this generates uncertainty about the visual pattern). The task is to indicate whether the target is absent or present.

4.2.1 Models

4.2.1.1 The Ideal Observer

The computations of the ideal observer are outlined in Figure 4.2A. I assume that the prior probability of the target being present, and prior probabilities of targets (visual patterns) are available to the ideal observer. Furthermore, the ideal observer is assumed to know the background contrast and target amplitude, which are fixed in a single simulated experiment. The target location is also assumed to be known (no position uncertainty). The ideal observer compares the natural logarithm of the likelihood ratio (D_{opt}) to the natural logarithm of the ratio of priors (criterion shown in Figure 4.2A) and responds target-present if the log-likelihood ratio is higher. The log-likelihood ratio is computed by appropriately pooling template responses (dot products of targets with the target regions in the image) weighted by their prior probability.

The computation of the target present likelihood requires summing the probability of being present for each target. Even though summation is relatively fast to compute, as the uncertainty size and number of dimensions grows, the number of potential targets grows, and computing template responses can take a substantial amount of time for each

image. However, the implementation of the ideal observer in the thesis introduces the principle of using as many pre-computed variables as possible. Specifically, the simulation procedure takes advantage of the image formation process for additive targets in Gaussian white noise background. Template responses are expressed as the combination of pre-computed responses to simple white noise background and pairwise products of all targets (for details see Appendix 3). This implementation allows various parameter changes (amplitude, background contrast, and priors) without repeating the most time-consuming part of the simulation (computing dot products). Lower levels of uncertainty can be simulated without repeating pre-computations by considering only a subset of the pre-computed dot products.

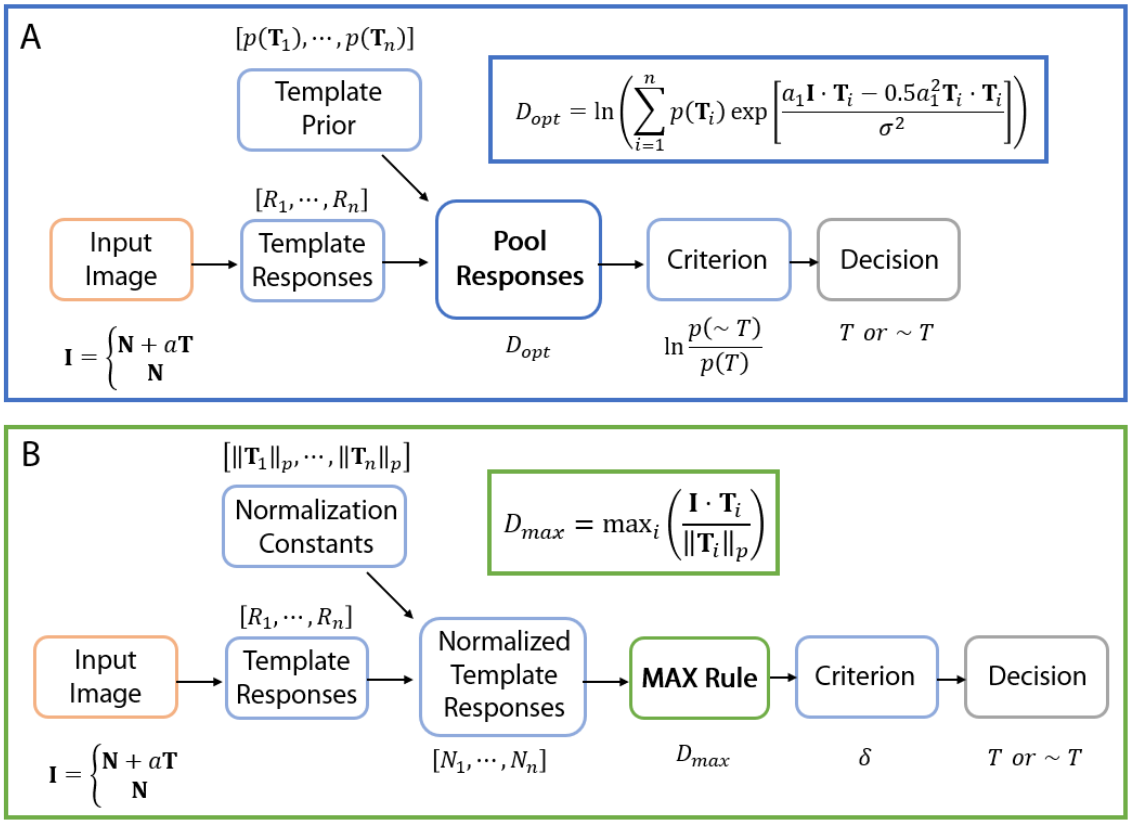


Figure 4.2 Flow chart of model observers.

A The ideal observer. First, the ideal observer computes template responses which are dot products of targets with the target region in the image. These are appropriately pooled with respect to the prior probability of each target (template). The log-likelihood ratio is compared to the logarithm of the ratio of priors which act as a criterion. If the log-likelihood ratio is higher than the criterion, the ideal observer responds as target-present. **B** The MAX observer family. Template responses are computed first. They are normalized by the L^p norm of the corresponding template. Based on the value of p, there are different versions of MAX observer. If p is infinity, the norm takes the maximum of the vector, and all normalization constants will be one. In this case, template responses will not be normalized. In the next step, the maximum of normalized template responses (max rule is applied) is compared against a fixed criterion to produce a decision.

4.2.1.2 The MAX Observer Family

When there is an extrinsic uncertainty over multiple target patterns, a common observer proposed in the literature is template-matching with a max rule (Figure 4.2B). For each target (template), the dot product between the target and target region (template

response) in the image is computed. These template responses are normalized by the L^p norm of the targets, and each p represents a different version of the MAX observer. Note that when p is infinity, the norm takes the maximum of the vector, and because all targets have the same maximum of 1.0, the normalization has no effect. The maximum template response is compared with a single fixed criterion to produce a decision. Note that the normalization of template responses is equivalent to using normalized templates. I used the same simulation procedure (expressing template responses as pre-computed variables) for this family of model observers.

4.3 Simulation Methods

The target is described in the experimental methods below. The target scale of 1 is defined as the largest target whose diameter is 3.35 visual degrees and spatial frequency is 1.0 cpd. (A visual degree is 60 pixels.) Targets are scaled by multiplying the image coordinates by one over the desired scale. The scale factor increases linearly with the target's spatial frequency, and the target diameter exponentially decays with the scale (Fig 4.3B). The target scale is randomly sampled from fifty-one, linearly spaced, and equally likely levels of scales between 1 and 6 (target scale uncertainty, Figure 4.3A). The target orientation is randomly sampled from 360, linearly spaced, and equally likely target orientations between 0 and 360 degrees (target orientation uncertainty, Figure 4.3A). Luminance is reported in gray level values so that the same measure is used for both amplitude and contrast. In each simulated experiment, the standard deviation of white noise was fixed at 5 gray levels.

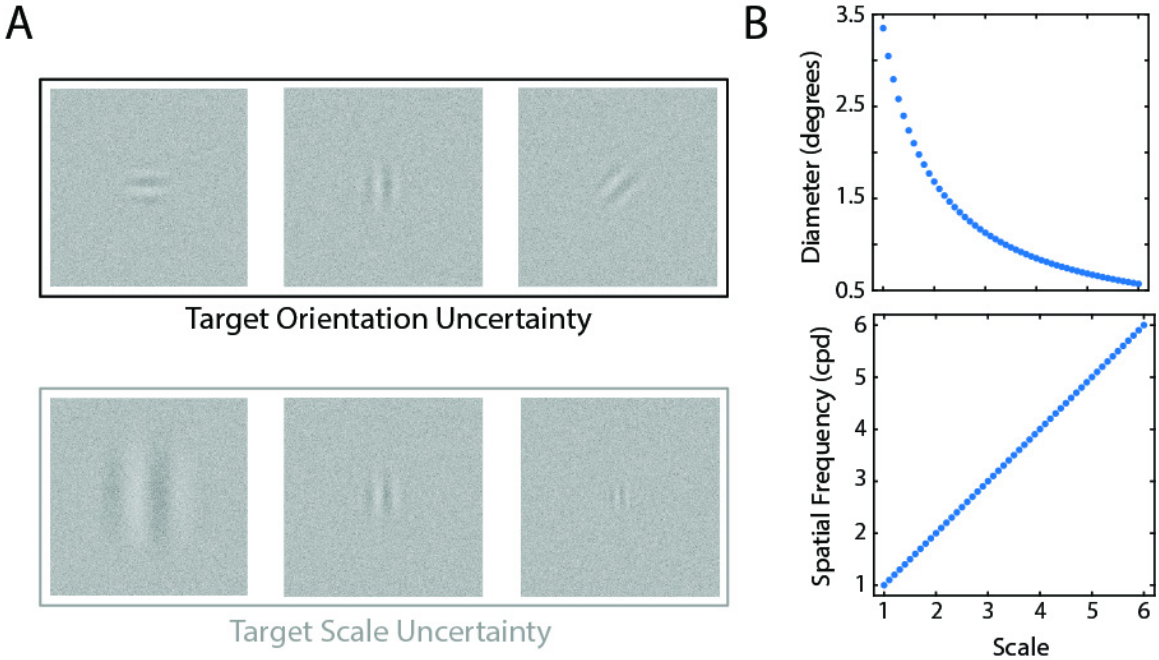


Figure 4.3 Target scale and orientation uncertainty.

A The orientations shown are 0, 90, and 135 degrees, respectively (note that at 180 degrees, the phase of the target changes since it is not symmetrical around the middle point). The scales shown are 1, 2.5, and 4.5. **B** The relationship of scale with target size and spatial frequency. The target's diameter exponentially decays with scale, and spatial frequency is linearly related to the scale.

4.4 Simulation Results

I evaluated the performance of model observers for various amplitude levels, and the criteria of the MAX observer family were selected to maximize the percentage correct. There was a total of 51 different scales between one and six. There was a total of 360 different orientations between zero and 359 (overall 18,460 different possible targets). I assumed that the prior probability distributions of target scale and target orientation are uniform and the probability of the target being present is 0.5. For each simulated experiment in which the amplitude and background contrast is fixed, there was a total of 36,720 trials (1 target present trial for each different target). When less uncertainty is

simulated (like only target-scale uncertainty), I simulated the maximum number of trials possible given the precomputed dot products.

First, dot products were computed for all possible pairing of templates and targets, and for each template, template responses to all 60,000 white noise backgrounds were computed. The process (pre-computation) takes about 25 minutes with an average office computer, and the file takes about 12 GB (when the random seed is saved rather than all individual white noise backgrounds). For thirty amplitude levels ranging from 0.01 to 3, the method I introduced (see the ideal observer section) was able to simulate the exact ideal observer in less than a minute for each amplitude level (30 amplitudes in 15 minutes approximately) with an average office computer. This is equivalent to simulating with 4 dimensions of uncertainty with 10 levels along each dimension. (Processor: Intel(R) Core (TM) i7-3770 CPU @ 3.40GHz 3.40 GHz).

I compared the ideal observer to the family of MAX observers (see Figure 4.4A). Only the MAX observer with responses normalized by the square root of the energy of templates (LP 2, Euclidian norm) approximates the ideal observer. I will call this observer the energy-normalized MAX observer. As the normalization factor increases, the detection performance decreases. If template responses are not normalized (or equivalently normalized by the LP INF norm), the performance is the worst. I will call this observer simply the MAX observer. When the normalization exponent is one (LP 1 norm), the pattern of errors is considerably different. The reason is best illustrated by considering two substantially different scales (large vs. small target, Figure 4.4B). The large target is easier to discriminate (high d prime) shown in green mist, whereas the small one is harder to discriminate (small d prime) shown in light red. The distribution of template responses computed from target absent trials (zero mean) and target present trials (non-zero mean) are shown for both scales. The target in these target present trials is exactly the same as the

template (in target present trials for the large target, the large target is presented). As shown in the first panel, when template responses are not normalized, for the large template, template responses have a larger standard deviation. When a small target is presented, the max rule is more likely to choose a response generated by the large target template. However, the same template response is also quite likely to be generated by a target absent trial (having different criteria for different scales does not help because the max rule is a problem, see Appendix 4). Therefore, small targets will be more likely to be missed. When template responses are normalized by the square root of energy, the standard deviations are equalized (note that target absent distributions are on top of each other). In this case, the primary limitation is the need for different criteria for different scales (when implemented, the performance is almost the same as the ideal observer, see Appendix 4).

When the LP norm is one, the standard deviation of the small target's template response distributions is higher than the large target's distributions. In this case, when a large target is presented, the max rule is likely to choose a response generated by the small target template, which is also quite likely to be generated by a target absent trial. Therefore, large targets will be more likely to be missed, but because large targets are highly discriminable, the effect is most prominent when the amplitude is low (red curve in Figure 4.4A). As amplitude increases, the norm of one catches up to the ideal performance because large targets become so much more discriminable than small targets (they topped to almost a hundred percent quickly). When the overall performance is close to a hundred percent, the LP 1 norm performs closer to the ideal than the LP 2 norm because target-present distributions are better aligned, and thus a single criterion works better (if different criteria are allowed for different scales, it has an almost negligible effect on the performance of LP 1 norm. See Appendix 4).

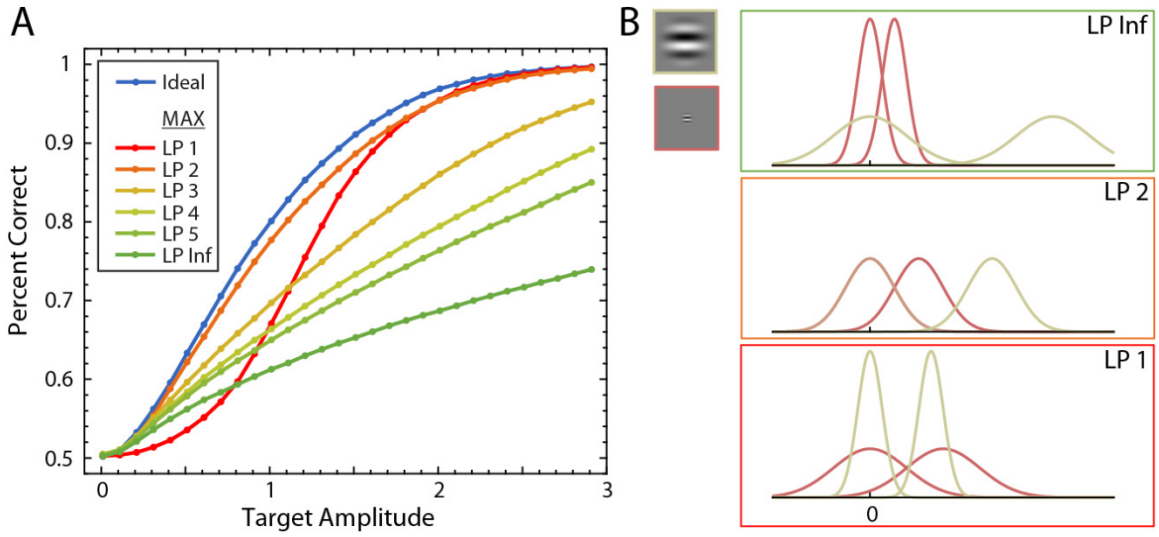


Figure 4.4 The comparison of the ideal observer with the family of MAX observers.

A For MAX observers, template responses are normalized with the L^p of targets for various levels of p. The criterion of the MAX observers is selected to maximize the overall percentage correct, and the overall percentage correct is shown as a function of target amplitude. **B** The illustration of the effect of normalization on the standard deviation of template response distributions for various levels of p. The template responses' distribution is shown for two templates: highly discriminable large target in green mist and hardly discriminable small target in light red. When the target is absent, distributions sit at zero. If the target that is the same as the template is presented, the mean is non-zero. The first panel shows the template responses when they are not normalized, the second panel shows them when they are normalized by the square root of the target's energy of the target, and the last panel shows them when the normalization norm is one (LP 1, L^1).

In the following sections, I will only focus on the ideal observer and two members of the MAX observer family: energy-normalized MAX observer (template responses normalized by the LP 2, Euclidian norm) and simple MAX observer (no normalization, LP Inf). Measuring the overall percentage correct as a function of amplitude with psychophysical experiments requires a considerable amount of data. However, the overall percentage correct only summarize the performance. Within a single uncertainty experiment that has fixed amplitude, model observers' hit and correct rejection rates for different target conditions might be sufficiently different to distinguish between observer models. At the threshold amplitude, hits and correct rejection rates are shown as a function

of the target scale in Figure 4.5. There is no difference in target absent trials, so there is only a single correct rejection rate in a single high uncertainty experiment. Moreover, theoretically, there is no difference between orientation levels (the same pattern of hits is expected), so presented hit rates are averaged over orientation levels (for each scale, there are 360 target present trials which is why the estimates are noisy). The simple max observer's hit rates are substantially lower than the other two model observers, whereas its correct rejection rate is slightly higher. The energy-normalized MAX observer closely approximates the ideal observer, and it seems possible to use hit and correct rejection rates to discriminate between model observers.

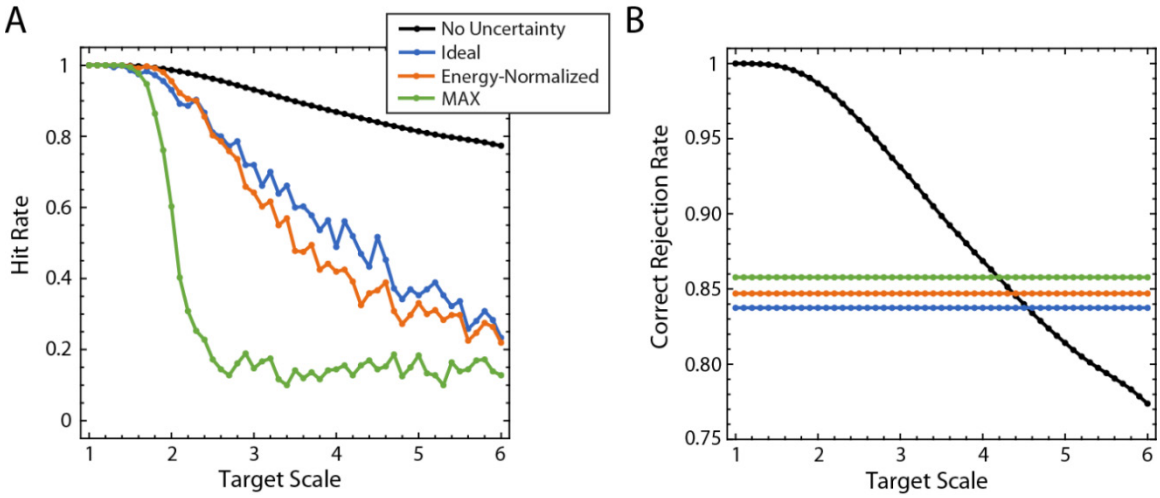


Figure 4.5 Hit and correct rejection rates.

A Hit rates as a function of target scale for three observer models under simultaneous scale and orientation uncertainty. The same rates are shown when there is no uncertainty about the scale and orientation. The target amplitude is picked such that the overall percentage correct is around 75 percent (threshold level). The expected hit rate patterns for different orientations are the same, so hit rates are averaged over orientation levels. Each point is calculated from 360 target present trials. When there is no uncertainty, all models' predictions are the same. **B** Correct rejection rates as a function of target scale for three observer models under scale and orientation uncertainty. The same rates are shown when there is no uncertainty. When there is a combined target scale and orientation uncertainty, target absent trials do not differ systematically, so there is only a single correct rejection rate for a single uncertainty experiment.

It is also interesting how distinguishable these model observers are under the sub-set of uncertainties I considered here. The three observer models are the same when there is no uncertainty about scale and orientation. When there is no uncertainty about the scale and orientation, hits and correct rejection rates are shown in Figure 4.5 (averaged over orientations for the same reason discussed previously). If there were only orientation uncertainty, then there won't be any difference between the predictions of the three model observers' model predictions (not shown here). The scale uncertainty introduces conditions where discriminability is different, which is why models have different predictions. Under only scale uncertainty, the results shown above stay the same, but the differences between the models are considerably smaller (Appendix 4). Moreover, the difference between normalizing with the LP 1 norm and the LP 2 norm shrinks considerably, and the need for a dynamic criterion for different scales becomes the major factor driving the information loss. Therefore, the addition of orientation uncertainty emphasizes the importance of normalization.

The decrease in ideal observer performance compared to the low uncertainty condition (no uncertainty about orientation and scale) characterizes each uncertainty dimension's effect and how the effects are combined. The ideal observer's performance under scale and orientation uncertainty is shown in Figure 4.6A (Scale and Orientation Unc.). Also shown in Figure 4.6A are the ideal observer performance under only scale uncertainty averaged across orientations (Scale Unc.), and under only orientation uncertainty averaged across scales (Orientation Unc.). The performance under low uncertainty is also averaged across both scale and orientation uncertainty (because each target condition is different, resulting in different overall percentages correct). Both individual dimensions of uncertainty cause a decrease in performance compared to low uncertainty, but orientation uncertainty causes a larger decrease in performance. In Figure

4.6B, the percentage correct of the low uncertainty condition is subtracted from the other three conditions to better reveal the effect of uncertainty on the overall percentage correct. The dashed blue curves show the predicted difference if the effects of scale and orientation were additive. The prediction is close to the combined effect; however, there is a systematic error pattern, so the unexplained part is likely the result of interaction between these two dimensions. Part of this interaction might be unique to the dimensions I consider here. However, the interaction might be predicted better with a more appropriate combination rule for uncertainties than simple addition. Additive predictions on the overall percentage are only shown for comparison purposes.

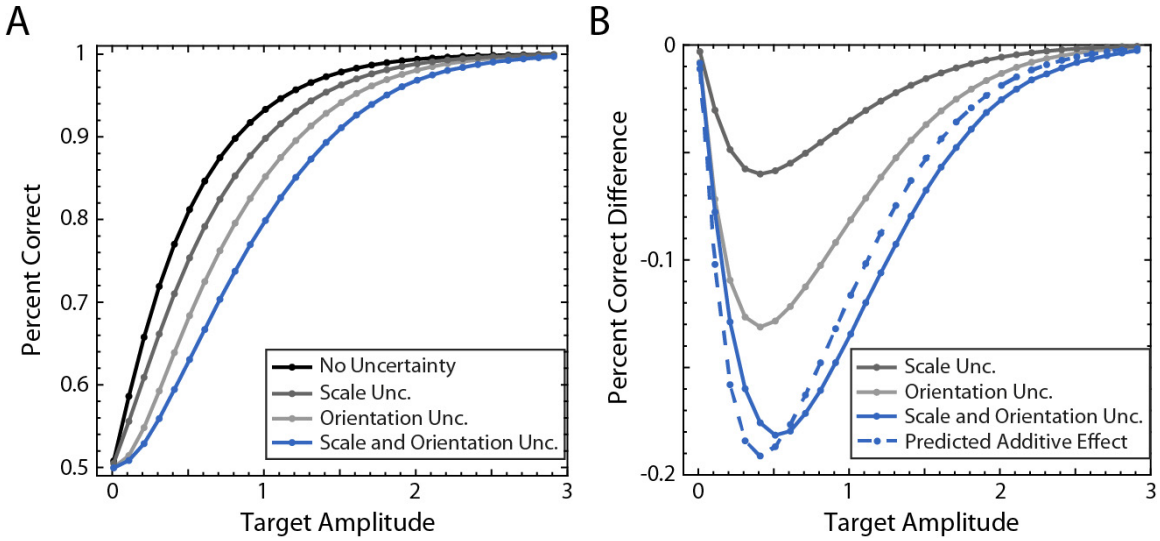


Figure 4.6 The comparison of levels of uncertainty.

A The overall percentage correct as a function of target amplitude for ideal observers when there is no uncertainty about scale and orientation, only scale uncertainty, only orientation uncertainty, combined scale and orientation uncertainty. The overall percentage correct is calculated by averaging across dimensions when there is only a sub-set of uncertainty. **B** The simple percentage difference as a function of target amplitude between the ideal observer when there is no uncertainty about scale and orientation and ideal observers when there is only scale, only orientation, combined scale and orientation uncertainty. Also, the blue dashed line shows the predicted effect of combined scale and orientation dimensions if their effect were simply added together.

4.5 EXPERIMENTAL METHODS

4.5.1 Stimuli, Subjects and Procedure

The target was a horizontal sine wave (0-degree orientation) windowed with a radial raised-cosine function. The background was a square patch of white noise whose side was 12 visual degrees (721 pixels). The target was presented at the center of the display together with the white noise background. The stimuli and design were adjusted for the psychophysical experiment. The same range of target orientations (0 to 359 degrees) was sampled with eight levels. Furthermore, the target scale range was smaller (1 to 4.5 instead of 6) and was sampled with eight levels. The eight linearly spaced levels of target orientations are 0,45,90,135,180,225, 270, and 315 degrees. The eight linearly spaced levels of target scales are 1,1.5,2,2.5,3,3.5, 4, and 4.5 (Figure 4.7A). The target amplitude and background contrast were selected to expect performance around 75 percent correct. The target amplitude was 3.5 gray levels, and the standard deviation of white noise was 15.46 gray levels (12 percent RMS contrast). The background gray level was 128.

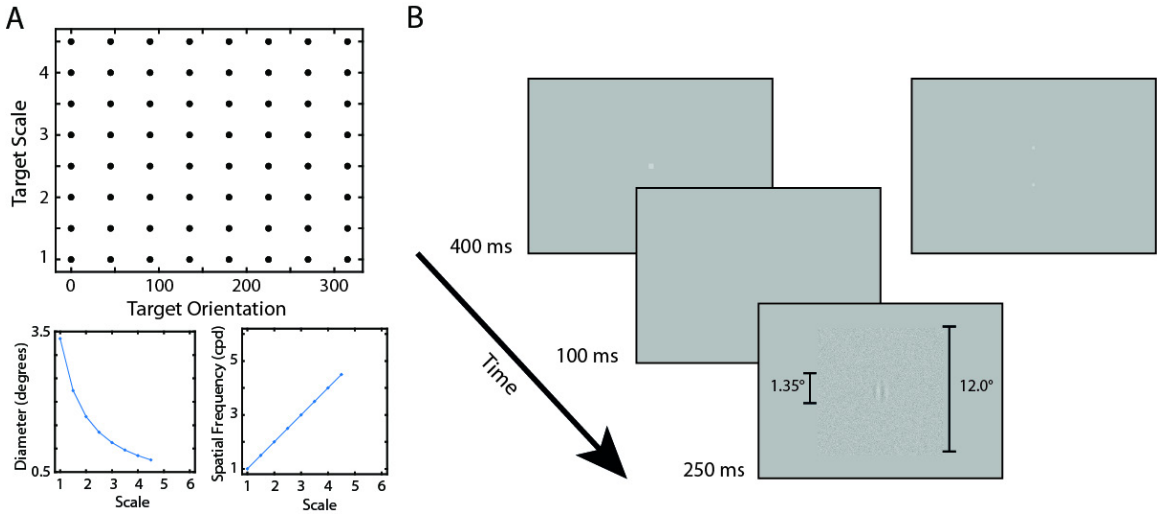


Figure 4.7 Experimental methods.

A Target scale and orientation levels in the experiment. For any target present trial, the stimulus's parameters are uniformly sampled from the grid of points. The two small figures show the relationship between the target scale and the target's diameter and spatial frequency like Figure 4.2B. Filled circles indicated the conditions used in the experiment. Note that the range of target scales is smaller. **B** The procedure of the experiment. 400 ms fixation screen is followed by a 100 ms black screen. The stimulus is shown for 250 ms with twelve visual degrees square background. In the low-uncertainty condition, the procedure was the same except for the fixation shown on the side. The fixation consists of two rectangles similarly oriented to the target and always 20 percent of the target diameter away from the target location to remind participants of the scale and orientation of the target presented in that block. The example shown here has an orientation of 90 degrees and a scale of 2.5.

The task was the same as in the previous experiment; in each trial, participants were asked to press a key to deliver their decision about whether the target was present or absent. Each trial starts with a dim fixation point at the center, presented for 400 ms, followed by a 100 ms blank window. The stimulus was presented for 250 ms (Figure 4.7B). Participants had 750 ms to indicate their decision, and the subsequent trial started after an additional 700 ms. The stimulus presentation time was selected to match the average fixation duration of the human eye. A high-frequency feedback tone was presented if the participant's response was correct, whereas a low-frequency tone was presented if the response was incorrect. On each trial, the target scale and the orientation were randomly and

independently sampled from uniform distributions over eight discrete levels of scale and orientation; however, the sampling was counter-balanced, so there was the same number of trials for each condition. The biggest and smallest target (at high amplitude) was shown on screen in the breaks to aid participants in remembering the range of scales that could be presented in the experiment.

I also conducted the same experiment when there is no uncertainty about the scale and orientation (low extrinsic uncertainty experiment). There were 64 different blocks of the experiment; at the start of each block, the target for that block (at high amplitude) was shown to make sure that participants knew which target was to be presented in that block. The fixation target also depended on the target (Figure 4.7B). The fixation target was composed of two small rectangles with a long side of 0.35 visual degrees and a small side of 0.15 visual degrees located around the target location. They were separated from the target location by 20 percent of the target's diameter (which also helped participants remember the target's scale). The fixation target was also oriented parallel to the orientation of the bars of the sinewave (which helped participants remember the orientation of the target).

To test for any learning effects through the experiment, I decided to split the experiment into two sessions. Each was a complete experiment with only half of the trials, and then I compared two sessions. In each session, participants first completed the low uncertainty condition that contains 64 blocks (different target conditions), each one with 26 trials (13 target present, a total of 52 trials with two sessions). The order of blocks was randomized. The high uncertainty condition is split into five blocks. The first three are completed in the first session, and the last two are completed in the second session. Each high uncertainty block consists of 640 trials for a total of 3200 trials.

Three experienced and one naive observer completed the experiment. They had normal (corrected) spatial acuity. Written, informed consent was obtained for all observers in accordance with The University of Texas at Austin Institutional Review Board. The stimuli were presented at a distance of 90 cm with 60 pixels per degree resolution. The luminance of the background was the mean of the display range (47.18 cd m^2). Images were gamma-corrected based on the calibration of the display device (GDM-FW900; Sony), quantized to 256 gray levels. The screen refresh rate was 85 hertz. All experiments and analyses were done using custom codes written in MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

4.5.2 Analysis Methods

The average observer's data was obtained by aggregating the data of participants. The simulation methods were adjusted to changes I made for the psychophysical experiment. The target scale and orientation levels were the same as those used in the psychophysical experiment (only eight levels). The mean gray level was subtracted from images for computational simulations. Confidence intervals were generated by resampling the data assuming the response in a trial follows a binomial distribution with a success rate of corresponding hit (miss) and correct rejection (false alarm) rates. Confidence intervals for fitted parameters were computed by fitting the bootstrapped data.

4.5.2.1 High Extrinsic Uncertainty Experiment

Three model observers (the ideal observer, energy-normalized MAX observer, and simple MAX observer) were fitted to the data collected in the high uncertainty experiment with the same likelihood method introduced in Chapter 3. The method implicitly fits the hits and correct rejection rate of the subjects by varying two parameters to maximize the

likelihood of the data. The first parameter is the scale factor on the target amplitude, which intends to account for the overall performance difference between the model observer and subject. The second parameter is a criterion intended to account for sub-optimal decision strategies (e.g., being too conservative or liberal in the decisions). The ideal observer criterion was manipulated by varying the assumed prior probability of the target absent (which effectively varies the criterion). However, the max observer family's criterion was estimated by finding a criterion value on the template-response axis that best fits the data. The units of template response are rather arbitrary and hence not very informative. A more intuitive measure is the prior probability of target absent that makes the estimated criterion optimal (this is sensible because there is a one-to-one mapping between two measures).

An adjustment was made to the likelihood method for fitting compared to that in Chapter 3. Observer models were simulated with a finite number of trials, and they sometimes reach the hundred percent or zero percent hit rate (or correct rejection rate). In this case, the likelihood of the data might become zero, and the method rejects the plausibility of any parameter, resulting in a hundred percent or zero percent hit rate (or correct rejection rate). To avoid the issue, I assumed that if I doubled the number of trials, there would be a single error and calculated rates based on this assumption.

4.5.2.2 Low Extrinsic Uncertainty Experiment

If target orientation and scale are known (as in the low uncertainty condition), then all three observer models considered here perform identically and are optimal. For 64 blocked conditions in the low uncertainty condition, rather than having 64 different scalar factors on amplitude and 64 different criteria, only a single scalar on amplitude was varied in the fitting procedure. Criteria were calculated from the data and fixed for observer models. First, signal detection criteria are calculated, and then they are converted to the

prior probability of target absent for which that criterion would be optimal for (with an equation). There was no systematic variation in prior probabilities as a function of orientation. However, there was a systematic variation for the target scale, so criteria are only averaged across the orientation levels. For each blocked condition, I used these fixed priors based only on the scale level (for the average participant: 0.16, 0.34, 0.45, 0.50, 0.56, 0.54, 0.55 respectively for scales starting from 1 to 4.5) as the assumed prior probability of target absent for the ideal observer. The rest of the fitting procedure was the same as the high uncertainty condition. When the data from both conditions were fitted simultaneously to test principles about how the low uncertainty conditions are related to the high uncertainty condition, there is only a single scalar factor was used for both conditions.

Intrinsic uncertainties about the position, orientation, and scale were simulated by incorporating new templates (templates have slightly different orientations, scales, and positions) into the computation even though there is no target presented in experimental trials corresponding to these templates. Intrinsic orientation uncertainty was implemented with one-degree orientation steps, and scale uncertainty was implemented with steps of 0.1. Position uncertainty was implemented by moving the center of the template to spatially close locations. The maximum is 12 pixels radius position uncertainty, corresponding to square 25by25 pixel square possible center locations centered around the true center. However, introducing 624 new locations is computationally expensive (note that for each new template location, there are 64 templates due to orientation and scale uncertainty, so the total number of templates would be 39,936) and not necessary because these templates are correlated. Rather than stepping over the 25 x 25 pixel region with one-pixel steps, new templates were generated for two-pixel steps, effectively creating 13 x 13 possible center locations. The smaller position uncertainties (sub-set of uncertainty) are only simulated for even pixel numbers starting from 2 to 12 with steps of 2 (because of 2 pixels stepping).

However, position uncertainty (the unscaled version) has little effect for large templates because only a tiny percentage of templates are sampling new locations. This is why I also entertained an alternative hypothesis: intrinsic position uncertainty might be fixed in terms of the number of neurons. Since larger receptive fields are separated farther away in space, the same number of neurons would cover a bigger spatial region in image space with larger steps (the scaled version). The implementation of position uncertainty stays the same as for the smallest target. However, for any other targets, both steps and range are scaled up by 4.5 divided by its scale (for example, the largest target has a scale of 1, so 12 pixel radius multiples with 4.5 and rounded to 54 pixel radius and steps are 9 pixels steps).

4.6. Experimental Results

I assumed that the participant's response in each trial was binomially distributed with a success rate of hit and correct rejection rates and resampled the data to build confidence intervals to test for the learning effect. For all experienced participants, 95 percent confidence intervals of overall percentage correct for the first session overlaps with the second session indicating there is no increase in overall percentage correct for both the high uncertainty and low uncertainty conditions. However, the percentage correct of the naïve participant slightly increased for the high uncertainty condition (0.62 vs. 0.66) because 95 percent confidence intervals are not overlapping (0.6-0.639 vs. 0.640-0.69).

4.6.1 High Extrinsic Uncertainty Experiment

The hit rates for the 64 conditions and the correct rejection rate of the average participant in the uncertainty experiment are shown in the left panel of Figure 4.8. The ideal observer, energy-normalized MAX (ENM), and the simple MAX observer (SM) were fitted to the data collected in the high uncertainty experiment. Panels on the right of Figure

4.8 show the best-fitting predictions of the three observer models in terms of hit and correct rejection rates in the same format as the data. I found that the ideal observer and ENM better explain the data than SM (Figure 4.8), and 95 percent confidence intervals for the average participant do not overlap. The estimated scalar factors on amplitude are similar across all observer models (Ideal: 0.41, ENM: 0.38, SM: 0.56). The similarity of estimated scalar factors across models is consistent with the idea that the scalar factor acts as an overall performance difference between participants and model observers. The estimated criterion, when converted to the prior probability of target absent that it would be optimal for, is also similar for all observer models (Ideal: 0.45, ENM: 0.46, SM: 0.47). The similarity across levels may suggest that the parameter only accounts for the decision strategy, which is too liberal (making too many target present decisions). It seems like both parameters only account for known sub-optimalities of human participants regardless of the model; thus, the model fits to the data are similar to the zero-parameter predictions. This suggests that the underlying computational principles of the ideal observer and ENM are well matched to the visual processing.

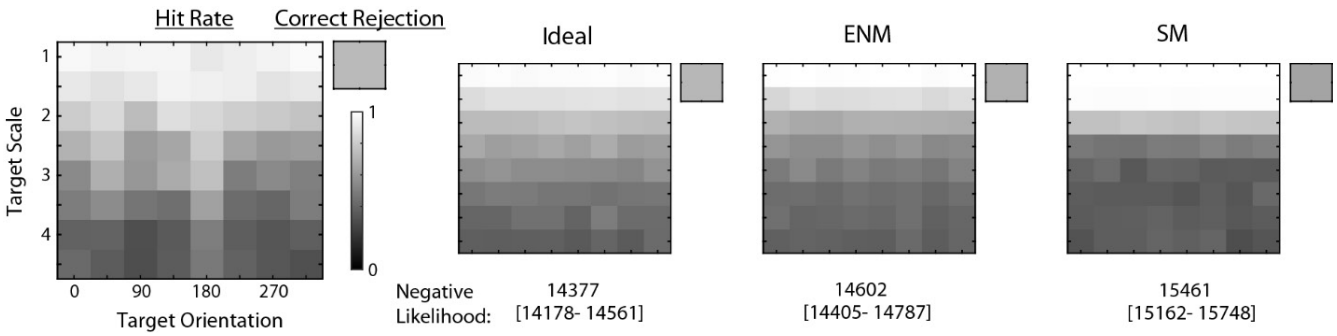


Figure 4.8 Data and model fits for experiment with high extrinsic uncertainty.

The data matrix consists of an eight-by-eight hit rate matrix and a single correct rejection rate of the average participant. The average participant data is generated by aggregating the data of every participant. Grayscale is used to represent rates between 1 (white) and 0 (black). Fitted observer models' hit and correct rejection rates are presented in the same structure with negative log-likelihood associated with the fit and 95 percent confidence intervals around it.

For some participants (P2 and P3), the data is better explained by the ideal and ENM than by the SM model (Figure 4.9A). For other participants (P1 and P4), the 95 percent confidence intervals around negative log-likelihood overlap, but there is still a clear trend in the same direction. The pattern of results is robust across participants. The quality of the fits is best illustrated with the correct rejections and the hit rates averaged across the orientation levels because there is not much variation with the orientation. For the average observer, even the best-fitted SM observer predicts a much lower correct rejection rate and a sharper decline in hit rates as a function of the target scale (Figure 4.9B). The ideal and ENM also fail systematically by predicting a slightly sharper decrease as a function of scale (compared to the approximately linear decrease of data) and a slightly lower correct rejection rate.

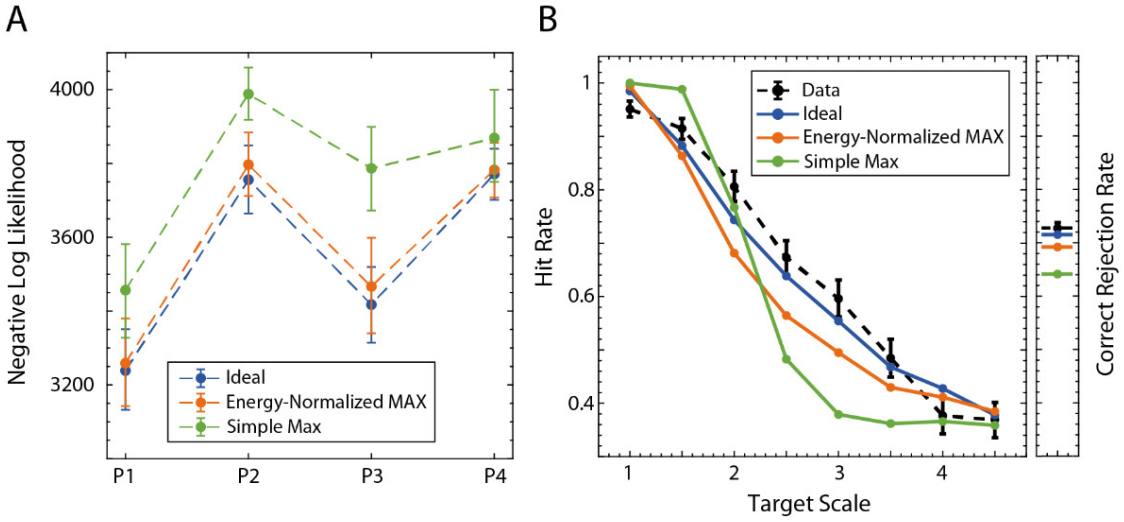


Figure 4.9 Results of models fits.

A The goodness of fit measure (negative log-likelihood) for individual participants. Error bars show the 95 percent confidence intervals that are computed by fitting the bootstrapped data. **B** Predicted hit rates (averaged across orientation) and correct rejection rate of fitted observer models to the average observer's data in the high uncertainty condition. Error bars around the data show 95 percent confidence intervals.

4.6.2 Low Extrinsic Uncertainty Experiment

A common approach to understanding the effects of extrinsic uncertainty is to compare performance in the high and low extrinsic uncertainty conditions. Understanding the relationship between these two experiments may allow us to predict the detection performance of one from the other and further predict the performance for modest levels of uncertainty. In this section, first, model observers will be fitted to the data collected when there is low extrinsic uncertainty to generate predictions about the high uncertainty condition. I will compare these predictions to the data collected in the high extrinsic uncertainty experiment to see whether the visual system performance is further limited by anything when there is high extrinsic uncertainty (like attention capacity, computational approximations like the max rule, etc.). Then, the efficiency of the average participant will

be directly compared across these two conditions by comparing the efficiency scalar factors. I will test computational principles that might allow the models to explain the differences in human efficiency between the high and low uncertainty experiments. Specifically, observer models with intrinsic uncertainties will be matched to the average d-prime of low uncertainty data to see whether predictions of these models can match the average performance in the high uncertainty condition. Lastly, observer models with intrinsic uncertainty are simultaneously fitted to both conditions to see whether the pattern of hit and correct rejection rates is better explained when the efficiency difference is accounted for.

The hit and the correct rejection rates of the average participant for the 64 conditions in the low uncertainty experiment are shown in the left panel of Figure 4.10. When the scale and orientation of the target are known, all three model observers are identical and optimal. The ideal observer of the low uncertainty condition was fitted to the data by only varying a single scalar factor on amplitude for all 64 blocked conditions. The criteria were calculated from the data and were fixed (see Methods for details).

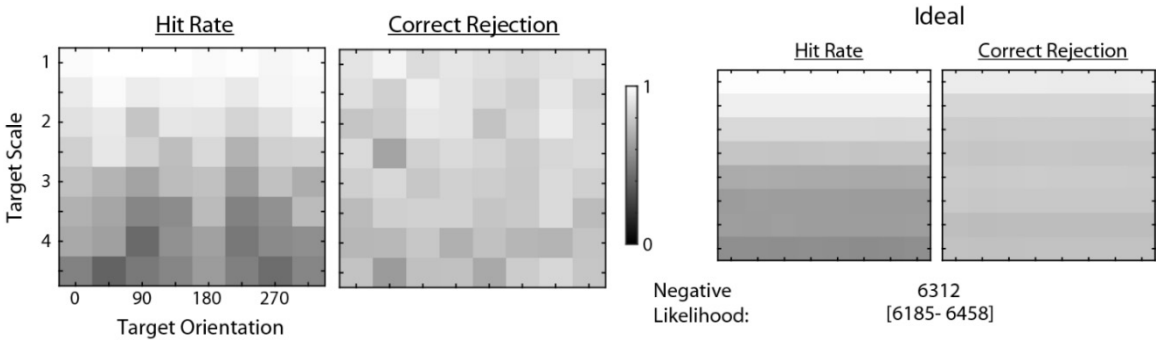


Figure 4.10 Data and model fit for experiment with no extrinsic uncertainty.

The data matrix consists of an eight-by-eight hit rate matrix and an eight-by-eight correct rejection rate matrix of the average participant. Grayscale is used to represent rates between 1 (white) and 0 (black). The fitted observer model's hit and correct rejection rates are presented in the same structure with negative log-likelihood associated with the fit and 95percent confidence intervals around it.

Panels on the right of Figure 4.10 show the best-fitting predictions of the ideal observer in terms of hit and correct rejection rates in the same format as the data. The ideal observer for the low uncertainty condition explains the most variation in the data (R^2 for hit rate matrix: 0.85, R^2 for correct rejection matrix: 0.40, Figure 4.10) with the estimated amplitude scale factor of 0.3. If the inefficiency of participants is due to internal noise, one would expect the same scale factor to translate to the uncertainty condition. This classic hypothesis follows if visual processing is not limited by attentional capacity. If there were limited attentional capacity, the performance would decrease more than predicted by the uncertainty models, because many more patterns would need to be attended/tracked in the high uncertainty condition.

I first focused on the hypothesis of no limited attentional capacity. The ideal observer's predictions in the high uncertainty condition were generated based on the scale factor estimated from the low uncertainty condition and the optimal criteria (Figure 4.11A, blue points). Clearly, the prediction is quite poor. Consistent with the same hypothesis, another possibility is that the visual system approximates the ideal computations under high levels of uncertainty with the normalized max rule (the ENM model). I generated predictions of the ENM model based on the efficiency scalar factor estimated in the low uncertainty condition with optimal criteria as well (Figure 4.11A, red points). Again, the predictions are quite poor. Versions of the hypothesis, even with optimal criteria placement, predict a considerably lower overall percentage correct (Ideal:0.6165, ENM:0.6116) than the average participant (0.6869). Therefore, assuming that the scalar factor estimated from the low uncertainty experiments is due only to internal noise cannot explain human performance in the high uncertainty conditions because it predicts performance levels below that of the human observers. Since an attentional capacity limit would only further decreases the performance, it also cannot explain the data alone.

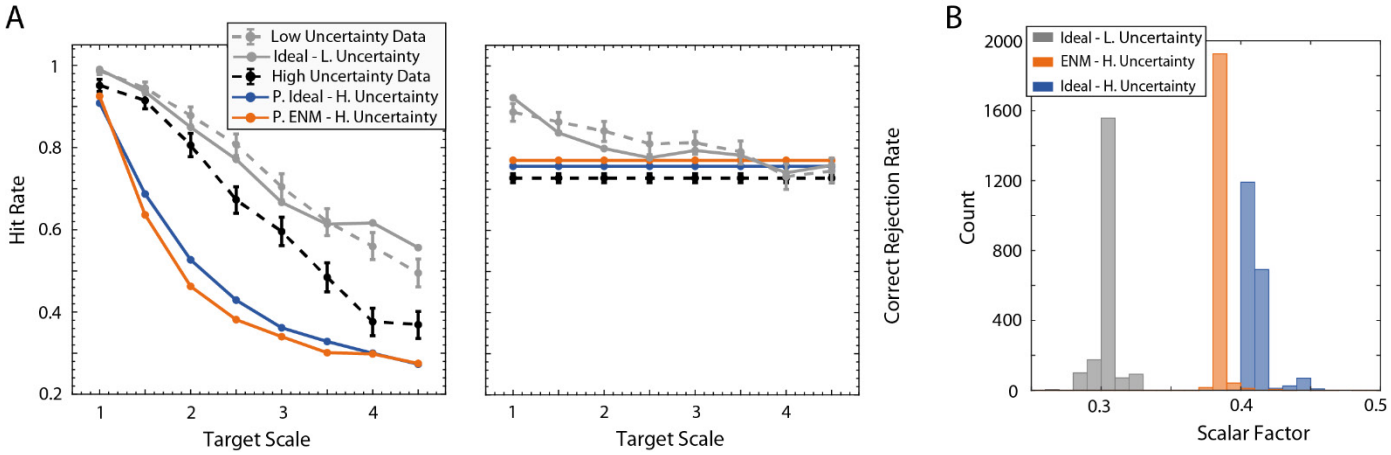


Figure 4.11 Results of model fits and efficiency comparison.

A Predicted performance in the high uncertainty condition based on the low uncertainty condition. The ideal observer for the low uncertainty condition is fitted to the average observer's data (solid gray line), and predictions for the high uncertainty condition are generated based on the estimated scalar factor and optimal criteria with ENM and the ideal observer (P. Ideal and P. ENM). Average participant data for both conditions are shown (black and gray dashed lines), and error bars around the data show 95 percent confidence intervals for both conditions. **B** The distribution of scale factors estimated by fitting the low uncertainty and high uncertainty conditions. Scale factors were estimated by fitting the 2000 bootstrapped average participant's data. For the low uncertainty condition, only the ideal observer is fitted. For the high uncertainty condition, ideal and ENM observers are fitted.

By comparing the scalar factors estimated by fitting the high uncertainty condition alone and low uncertainty condition alone, one can further show why predictions based on the scalar factor estimate from low uncertainty conditions fail to reach human performance. If the efficiency of human observers is the same for both conditions, we would expect the estimated scalar factors to be the same, which would be consistent with parallel processing of the information (no further limitation for higher uncertainty conditions due to attention, etc.). Limited attentional capacity (and various other computational limitations that are only effective for high uncertainty conditions) predict a lower scalar factor in high uncertainty condition because they predict lower efficiency in the high uncertainty. The

distribution of estimated scalar factors is computed by fitting the bootstrapped average participant data. Note that the scale factor on amplitude is estimated while criteria are being varied for the high uncertainty condition, whereas for the low uncertainty condition, the scale factor is estimated while criteria are fixed because criteria are calculated directly from the hits and correct rejections using standard signal detection theory. Regardless of the model type (ideal observer for uncertainty or ENM), the estimated scale factors for the high uncertainty experiment are higher than for the low uncertainty experiment (Figure 4.11B). This finding suggests that the visual system is more efficient in the high uncertainty condition; thus, some of the inefficiency in the low uncertainty experiment must not translate to the high uncertainty experiment.

An important factor that could explain the relationship the low uncertainty and high uncertainty conditions is intrinsic uncertainty. Even if there is no extrinsic uncertainty about the target (for example, if target orientation, scale, and position are fixed and known), the visual system may not exploit this information and still have to consider multiple levels of dimensions that are fixed. Such intrinsic uncertainty would have a considerable effect when there is low extrinsic uncertainty but would have almost no effect when there are high levels of extrinsic uncertainty. To test for the combined effect orientation and scale intrinsic uncertainty, I simulated observer models with extremely high levels of intrinsic uncertainty (+- 10 degrees of orientation and +- 1 value of scale). When the overall percentage correct in the low uncertainty condition is approximately matched (average d-prime average human participant: 1.78, ideal observer: 1.80 ENM observer: 1.79), the predictions of ideal and ENM observer with intrinsic uncertainty still fall short (0.6701, 0.6654) of human-level performance (0.6869) for the high uncertainty condition. Therefore, the relationship between conditions cannot solely be explained by these two dimensions of intrinsic uncertainty.

Intrinsic position uncertainty is more impactful than orientation and scale uncertainty on the overall percentage correct. To test whether predicted performance in high uncertainty condition reaches human performance when performance in low uncertainty conditions is matched, criteria are chosen to be optimal. For the unscaled version of position uncertainty (it does not scale with target size, see Methods for details), with a reasonable amount of intrinsic position uncertainty (8 pixels or 0.13 visual degrees radius), overall percent correct predictions (Ideal: 0.7033, ENM: 0.6953) is similar to human data (0.6869) when average d-prime is matched for low uncertainty condition (average participant: 1.78, Ideal: 1.81, ENM: 1.81). For the scaled version of position uncertainty (it does scale with target size, see Methods for details), with a reasonable amount of intrinsic position uncertainty (6 pixels or 0.1 visual degrees radius for smallest target), overall percent correct predictions (Ideal: 0.7088, ENM: 0.6954) are similar to human data (0.6869) when average d-prime is matched for low uncertainty condition (average participant: 1.78, Ideal: 1.81, ENM: 1.81).

For the rest of the analysis, the range of position uncertainty levels I tested starts from the smallest level of position uncertainty that predicts the overall performance in the high uncertainty condition of the average participant, when the d-prime is matched to the low uncertainty data (see above). The observer models with intrinsic position uncertainty were fitted to data of both conditions simultaneously. The aim was to test (i) whether the observer models that are able to account for the overall percentage correct will also be able to fit the hit and correct rejection pattern better than the baseline goodness of fit and (ii) whether a higher level of position uncertainty can better explain the pattern of hit and correct rejection rates with non-optimal criterion while still accounting for the overall percent corrects. The baseline is the goodness of fit obtained when the model without intrinsic uncertainty is fitted to both conditions simultaneously. Even though the model

observers without intrinsic uncertainty fail to capture the difference in overall percentage corrects for the two conditions, they provide a baseline to see whether intrinsic uncertainty is better able to explain the pattern of hit and correct rejection rates. I focused on the scaled version of intrinsic uncertainty here. However, all results apply to the unscaled position uncertainty, and its fits to the data are generally worse than the scaled version.

Figure 4.12A shows a trend suggesting that intrinsic uncertainty makes the fits slightly worse compared to the model without intrinsic uncertainty. On the other hand, as expected, the overall percentage corrects are better predicted with intrinsic uncertainty. For example, for the scaled intrinsic position uncertainty radius of 0.13° (for smallest target), the overall percentage corrects in both conditions are similar to the average participant (high uncertainty condition, human:0.69, ideal:0.69, ENM: 0.64; low uncertainty condition, human:0.78, ideal:0.78, ENM: 0.75). If the error between average participant's and fitted model's hit and correct rejection rates are quantified with root-mean-squared error, the model observers with intrinsic uncertainty turn out to be better (Figure 4.12A). The ENM observer with intrinsic position uncertainty always fits better than the baseline. However, the ideal observer only fits better than the baseline when the scaled intrinsic position uncertainty has a radius of 0.13° (for the smallest target), which is also the best fitting model observer overall (RMSE for high uncertainty: 0.46 for low uncertainty:0.81; Baseline RMSE for high uncertainty: 0.79 for low uncertainty:0.66). The reason behind the difference between RMSE and negative log likelihood measures is best illustrated by comparing the fitted hits and correct rejection rates for the scaled intrinsic position uncertainty of 0.13° to the fitted hits and correct rejection rates when there is no intrinsic position uncertainty (Figure 4.12B). Even though the overall difference between the high-uncertainty and low-uncertainty conditions is smaller for hit and correct rejection rates with intrinsic uncertainty, both fall sharper with the target scale than predictions without

intrinsic uncertainty (especially in low uncertainty conditions). Because of that, even though the fits improve due to better capturing of the overall relationship between high uncertainty and low uncertainty, they also worsen because the hit and correct rejection rates fall much more sharply than human data. The sharp decrease has the largest effect on the rates that are close to a hundred percent which affects the negative log likelihood much more than RMSE. I have not tested combining orientation and scale uncertainty with position uncertainty. In that case, a smaller position uncertainty would likely be enough to predict the overall percentage corrects reasonably. However, it is unclear whether the quality of fit will improve because the hit and correct rejection rates might still fall rapidly as a function of the target scale due to intrinsic position uncertainty.

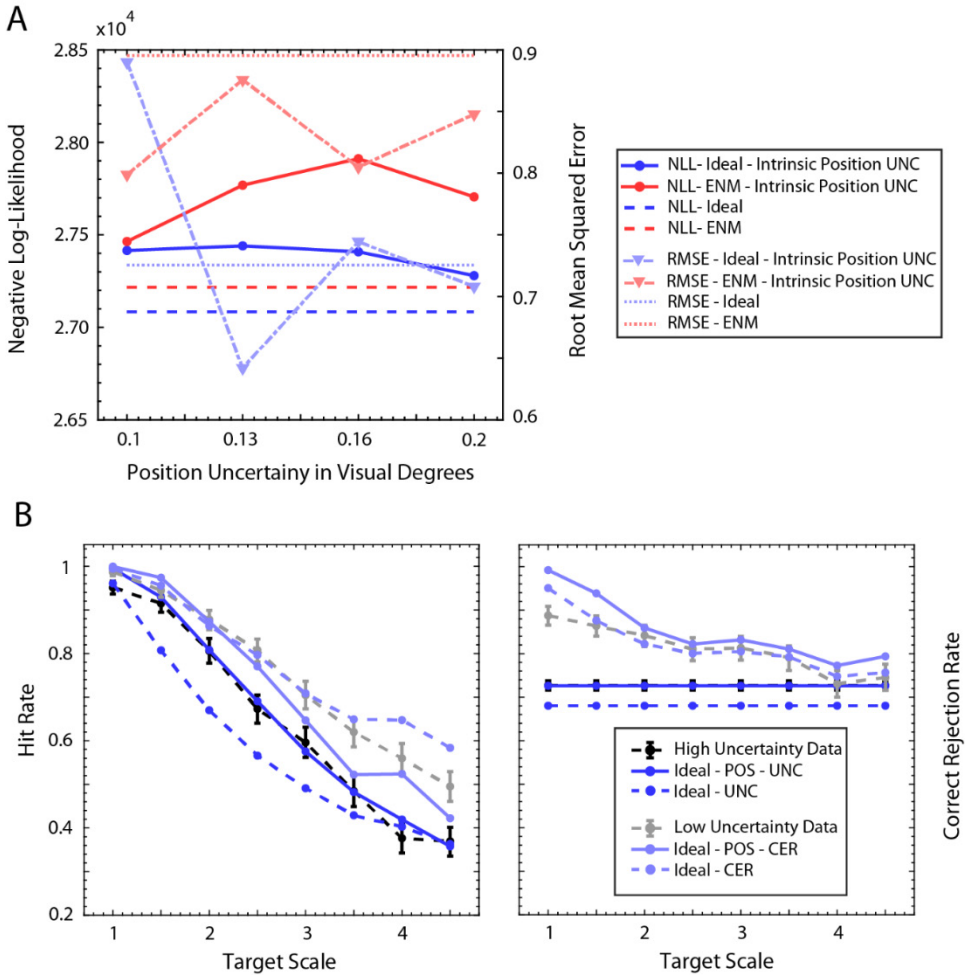


Figure 4.12 Models with position intrinsic uncertainty.

A The goodness of fit measure (negative log-likelihood and root mean squared error) as a function of scaled intrinsic position uncertainty. The observer models are fitted to both conditions simultaneously with the maximum likelihood method. Various levels of intrinsic uncertainty are fitted to the data and shown in solid lines. Dashed lines correspond to the fit of the observer models that do not include any intrinsic position uncertainty. If observer models with intrinsic position uncertainty explain the data better, solid lines should go below the dashed lines (baseline). Dotted lines and lighter colors represent the secondary y-axis which is the root mean squared error. The x-axis represents the radius of the position uncertainty for the smallest target size because this is the scaled version of the position uncertainty (see methods). **B** For the scaled intrinsic position uncertainty of 0.13 visual degrees radius (for the smallest target), predicted hit rates (averaged across orientation) and correct rejection rate of the fitted ideal observer to the average observer's data. Error bars around the data show 95 percent confidence intervals. Lighter colors show model predictions for the low uncertainty condition. The ideal observer without intrinsic position uncertainty also fitted to both conditions at the same time. Dashed blue lines indicate the best-fitted hit and correct rejection rates of the ideal observer without any intrinsic position uncertainty.

4.7. SUMMARY

I have derived the ideal observer under the combined target scale and 2D target orientation uncertainty for additive targets in white noise backgrounds. The ideal observer can be formulated as an extension of the classic template-matching observer (that is ideal when there is no uncertainty) by appropriately pooling all template responses. The derivation of the ideal observer does not depend on any properties of 2D target orientation and target scale except that they generate a set of possible visual patterns that can show up as the target. Therefore, the same ideal observer applies to any uncertainty dimensions that generate a set of possible visual patterns. Furthermore, I described an efficient method for simulating the exact ideal observer (and the maximum-template-response, MAX, observer) and demonstrated its efficiency. For a total of 18,360 template shapes, the method was able to simulate 36,720 trials for the exact ideal observer in less than a minute with an average office computer. Under high levels of 2D target orientation and target scale uncertainty, simulations revealed that the MAX observer only approximates the ideal observer if either its templates or template responses are normalized by the square root of energy of its templates. When it is not properly normalized, observer models differ in their hit and correct rejection rates patterns.

I measured the human performance under high levels of combined target scale and 2D target orientation uncertainty (extrinsic uncertainty experiment). Using the same fitting method introduced in Chapter 3, I showed that the ideal and energy-normalized MAX observers could explain human participants' hit and correct rejection rates. In contrast, a simple MAX observer fails to explain the data. The human detection performance is also measured for the same conditions when there is no extrinsic uncertainty about scale and orientation (low extrinsic uncertainty experiment). I found that predictions of model observers based on the assumption that humans are similarly efficient in both experiments

fail to predict the data measured in the high extrinsic uncertainty experiment. Then, I found that human participants are more efficient in the high extrinsic uncertainty experiment by estimating the scalar factors with the exact ideal observers. Multiple hypotheses about why efficiency is increased in the extrinsic uncertainty experiment are tested, including combined 2D target orientation and target scale intrinsic uncertainty, and sole position intrinsic uncertainty. I found that no reasonable combination of 2D target orientation and target scale intrinsic uncertainty does explain the data better. However, intrinsic position uncertainty does explain the data better in terms of root mean squared error but does not in likelihood.

Chapter 5: Discussion

5.1. Summary

Visual detection is an essential part of many everyday tasks; consequently, humans spend considerable time searching for specific targets. Complex behaviors, like visual detection, consist of layers of processes (perception, cognition, action) and are better understood within a principled conceptual approach. Chapter 1 describes a normative conceptual approach based on natural selection. The approach has two components: naturalistic tasks and ideal observer analysis. I discussed how each component already provides important insights about visual processing. Furthermore, I argued that the conceptual approach fits the current literature on visual detection under uncertainty.

Chapter 2 is concerned with reviewing how human performance is measured and modeled under various levels of uncertainty so far in the literature. The previous literature has primarily measured the effect of extrinsic uncertainty on human performance compared to conditions with lower levels of extrinsic uncertainty. Most uncertainty dimensions have been studied in isolation (modest levels of uncertainty), even though real-world visual detection involves simultaneous multiple dimensions of uncertainty. A few examples of studies that measure and model human performance uncertainty multiple dimensions of uncertainty are also reviewed. I identified three major challenges that will be faced moving forward to more naturalistic conditions and discussed a potential resolution to tackle these challenges. Finally, I introduced the generic task, which is the focus of this thesis. It is a more naturalistic visual detection task that involves simultaneous multiple uncertainty dimensions for additive targets in white-noise backgrounds. Then, I lay out a plan for studying the generic task in two parts: (i) simultaneous target amplitude and background contrast uncertainty and (ii) simultaneous target scale and target 2D orientation uncertainty.

In Chapter 3, I derived and simulated the performance of the exact ideal observer and compared it with two sub-optimal observers under simultaneous target amplitude and background contrast uncertainty. I found that the simple extension of the template-matching observer that normalizes the template response by the background contrast approximates the ideal observer. However, a simple template matching observer fails to approximate optimal performance only when the prior probability of the target being present is low. Then, I measured the human performance under two levels of the target prior (0.5 and 0.2). Surprisingly, the template-matching observer fails to explain the human data for both target priors, whereas the ideal observer and normalized template-matching observer explain the data better.

In Chapter 4, I derived and simulated the performance of the exact ideal observer under simultaneous target scale and 2D target orientation uncertainty. The ideal observer is compared to the differently normalized versions of the model observer using the maximum-output (MAX) rule. I found that the MAX observer only approximates the ideal when normalized by the square root of the energy of templates, and its performance is worst when it is not normalized at all (the simple MAX observer). I measured human performance and found that the ideal and energy-normalized MAX observers explain the data better than the simple MAX observer. Furthermore, I measured the human performance under low extrinsic uncertainty for the same levels (no extrinsic uncertainty about orientation and scale, blocked condition). I found that humans are more efficient in the high extrinsic uncertainty experiment than in the low extrinsic uncertainty experiment. To explain this result (and fit both sets of data at the same time), I simulated various intrinsic uncertainties. I found that the ideal observer and energy-normalized MAX observer with intrinsic position uncertainty could better explain the data in both conditions simultaneously up to some extent.

This chapter will first discuss results reported in Chapter 3 about simultaneous target amplitude and background contrast uncertainty. Then, I will proceed with the discussion of findings presented in Chapter 4 about simultaneous target scale and 2D target orientation uncertainty. In the last section, I will discuss how the findings of both chapters relate to the generic task introduced in Chapter 2. Then, I will consider these two experiments in relation to the research on simultaneous multiple dimensions of uncertainty in general. Furthermore, I will remind the reader of three research challenges faced by moving to more naturalistic tasks and how this thesis exemplifies a potential resolution to tackle these. Finally, I will discuss how methods introduced in this thesis can potentially guide future research and make it possible to conduct principled research on multiple simultaneous dimensions of uncertainty.

5.2 Target Amplitude and Background Contrast Uncertainty

5.2.1 Information, Model Discriminability and Target Prior

Previous research has primarily measured the effect of extrinsic uncertainty compared to the condition with low extrinsic uncertainty. Davis et al. (1983) found that target amplitude (or target contrast) uncertainty in isolation does not affect human performance. However, the uncertainty makes the task objectively harder (the maximum achievable performance decreases when uncertainty is introduced, see Chapter 2), and they did not quantify the expected effect of uncertainty. Here, I found that the expected effect of target amplitude uncertainty is very small. Thus, the effect is likely to be within the error margin of their experiments, so it is not clear how human efficiency is compared between these two levels of uncertainty. I have not measured the performance in a low extrinsic uncertainty experiment, so I cannot compare the theoretical difference with measurements.

Theoretically, for my experiment, introducing background contrast uncertainty does not make the task harder because the uncertainty can be resolved by estimating the background contrast from the surrounding region. Thus, human performance is expected to be the same with or without background contrast uncertainty. However, proper computations need to take place because sub-optimal computations will result in a decrease in performance. Therefore, it is possible to introduce uncertainty dimensions like background contrast uncertainty to distinguish ideal and sub-optimal model observers. The performance of sub-optimal observers is more affected by the further introduction of target amplitude uncertainty (when the background contrast uncertainty is already present) than the ideal observer. Therefore, even though target amplitude uncertainty has a small effect on the ideal observer, it makes model observers much more distinguishable. Because of the unexpectedly detrimental effects of target amplitude uncertainty when the processing is already sub-optimal, one should be careful collecting data while varying the target's amplitude in a block. If the visual system is already engaging sub-optimal computations for the experiment, introducing target amplitude uncertainty might result in unexpected effects.

In most laboratory visual detection experiments, the target is present in half of the trials. So, the prior probability of the target being present is 0.5. The reasoning behind this decision is to collect more informative data about the detection behavior by providing minimum prior information. When a different target prior is used in the experiment, the general expectation is that participants will change their criterion accordingly (sometimes sub-optimally). So, the change in performance is expected to be consistent with the use of prior information provided, and there is no change in discriminability. However, as shown in Chapter 3, the expectation is not satisfied when the underlying decision variable is not optimal (i.e., the decision variable is not the likelihood ratio). Analysis of the TM observer

performance provides a striking example. When the target prior is 0.5, TM observer does approximate the ideal observer. However, when the target prior is low (like 0.2), its performance is almost not distinguishable from the chance performance achieved with only prior information. Thus, this result exemplifies that for some of the sub-optimal decision variables, the efficiency of model observers' (potentially human performance as well) dramatically depends on the target prior. Therefore, manipulating prior probability can be helpful, especially for more naturalistic tasks where the true decision variable (likelihood ratio) has a complex structure. Furthermore, it might be useful to simulate model observers whose underlying decision variable is not apparent, like many black box-type models, for various target priors. Because of the black box structure, it is not clear how the efficiency of these observers changes with the target prior. Therefore, manipulating prior probability could be a more informative strategy than previously thought for understanding the behavior of model observers (and the visual system) and for distinguishing between model observers.

5.2.2 The visual mechanisms

The simple template-matching observer is shown to explain human performance in basic visual detection experiments under low extrinsic uncertainty (Sebastian et al., 2017; for reviews, Cohn & Lasley, 1986; Geisler, 2003). However, this is the first time that the observer model has been tested under simultaneous target amplitude and background contrast uncertainty, a simple extrinsic uncertainty almost always present in the real world. I found that it fails to account for the human data under these more naturalistic conditions. I tested two candidate computational principles to extend the simple TM observer and showed that both explain the measured data better. These are dynamic criterion adjustment (which turns out to be optimal) and contrast normalization (contrast gain control). In this

section, I will discuss the evidence for the existence of both computations (starting with dynamic criterion adjustment) in visual processing.

There are multiple biologically plausible implementations of dynamic criteria. First, there might be built-in mechanisms learned through evolutionary history, which are (potentially automatically) applied when making decisions about object detection under different levels of contrast. In this case, participants do not need to learn anything new about adjusting their criteria in the experiment. Consequently, there will not be a cognitive load for adjusting criteria. Secondly, participants might learn the relationship between background contrast and criteria in this experiment to adjust their criteria cognitively. It is clear that participants are cognitively aware of background contrast, and there is no doubt that humans can keep multiple criteria in mind. Widespread experimental methods (rating scales) depend on the ability to keep multiple criteria in mind (most of the time, these are confidence ratings). In rating experiments, participants are assumed to rate the stimuli based on which criterion is exceeded, assuming criteria act as a border between any of the two rating levels. Furthermore, these methods are useful in measuring the ROC curves of participants in perceptual experiments and recognition memory experiments. However, there is controversy about whether keeping multiple criteria in mind affects the measured performance. For example, the cognitive load could decrease performance compared to measured performance with an experiment that does not require multiple criteria (Benjamin et al., 2013; Tekin & Roediger, 2017; for perceptual experiments: Egan et al., 1959; Swets et al., 1961). However, an eight-point scale is shown not to affect the measured performance (Swets, 1959); thus, participants can keep eight criteria (required in my experiment, too) without any loss due to memory constraints. However, note that criteria are on the same axis in rating experiments. Although criteria can be cast along the same axis in this experiment, they do not have to be on the same axis. In other words, it might

be unnatural for humans to compare a criterion for high background contrast to the criterion for low background contrast. Moreover, in rating experiments, the relationship between criteria might be simpler (linear or logarithmic increase), whereas it could be more complicated in this experiment. Together, these issues suggest that participants might have more trouble keeping multiple criteria in mind in this experiment than in rating experiments. However, in sum, there are multiple plausible ways of implementing dynamic criteria that are consistent with what is already known about the perceptual and decision processes.

Simple cells in the primary visual cortex are generally modeled with a simple linear receptive field (similar to a Gabor-like pattern) and compute neural responses similar to template matching operation. It has been previously shown that some neurons in the primary visual cortex normalize their response by the estimated contrast (contrast gain control, Albrecht & Geisler, 1991; Heeger, 1992; Carandini & Heeger, 1994; Carandini et al., 1997; Cavanaugh et al., 2002). If the brain decodes the activation of these neurons, their activation will be already normalized by the background contrast. Therefore, the TM observer's extension with contrast normalization simply corresponds to a more biologically realistic model of these neurons. The normalization does not affect the decision when there is no variation in background contrast. However, as I showed, if there is variation, the normalization is necessary to approximate the ideal observer. It is not the first time that contrast normalization has been added to image-processing models to explain psychophysical data (Watson & Solomon, 1997; Sebastian et al., 2017). It is a parsimonious explanation of measured data based on the computations known to be already implemented by the visual system. The normalization operation is also prevalently applied in various processing levels for other variables (Carandini & Heeger, 2012). There are many benefits of normalization operation reviewed by Carandini and Heeger (2012). For

example, background contrast normalization removes redundant correlations between neural responses, thus improving coding efficiency (Schwartz & Simoncelli, 2001). The findings presented here reveal another value of contrast normalization: making it possible to accurately detect targets under background contrast uncertainty with a single criterion. This benefit might drive the visual system to implement a normalization operation because, in the real world, background contrast varies considerably (Frazor& Geisler, 2006), making background contrast uncertainty unavoidable. Furthermore, computational simulations revealed its value is highest (compared to TM) when the target prior is low. In the real world, the probability of the target being present in any random fixation is extremely low. Taken together, the need for contrast normalization to accurately detect targets might be a strong factor that drove the natural selection.

Both computations are very likely to be implemented in the brain, and the data collected in this experiment does not distinguish between them. However, there is an important difference between these two model observers for the target prior of 0.5. Note that when criteria were chosen to maximize the percentage correct, the TM observer performs very similarly to the ideal observer. However, fitting results showed that the TM observer could not explain the data for the target prior 0.5, whereas the DTM observer could explain the data by choosing sub-optimal criteria (Figure 3.9). The estimated sub-optimal criteria suggest that humans are too liberal in their decisions for low background contrast, whereas they are too conservative in their decisions for high background contrast. It is not clear that considerable and systematic sub-optimality truly exists. On the other hand, the NTM observer with a single slightly non-optimal criterion can explain the data, which does not require introducing systematic and considerable sub-optimality. It is even possible that the DTM observer picked those criteria to explain participants' usage of relatively flat criteria after contrast normalization. Thus, considerable and systematic sub-

optimality of estimated criteria by DTM might be a by-product of the fitting process and might not truly exist. Finally, these two computations are not mutually exclusive, so that participants may be using both. Participants might endorse dynamic criteria after the contrast normalization has already taken place. If criteria are placed correctly, it is the exact ideal observer too. Thus, this strategy is equivalent to choosing dynamic criteria without normalization.

5.3 Target Scale and 2D Target Orientation Uncertainty

5.3.1 Behavioral Effects

I measured the effect of combined target orientation and scale uncertainty compared to low uncertainty conditions where both of them were specified in each block. The absolute performance is worse in the high extrinsic uncertainty experiment. However, the efficiency of human observers is higher under high uncertainty, so participants' performance decreased with the uncertainty, but the decrease is smaller than theoretically expected. Previously, the effect of these uncertainty dimensions was studied in isolation. Introducing orientation uncertainty either does not affect (Doehrman, 1974) or slightly decreases (Ukkonen et al., 1995) the human performance. Even though these studies did not quantify the expected effect, orientation uncertainty generally has a large effect on theoretical predictions. Therefore, it is likely that their findings suggest humans are more efficient in the high extrinsic uncertainty condition, consistent with the findings presented here. There are no studies that measured the effect of scale uncertainty, but there are studies on components of the target scale: spatial frequency and size. The studies on the effect of size uncertainty (when spatial frequency is fixed) revealed that humans are more efficient in the high extrinsic uncertainty condition (Judy et al., 1995; Judy et al., 1997). More recent studies also found only a small effect of introducing size uncertainty (Meese et al., 2005;

Foley et al., 2007; Meese & Summers, 2012) but did not quantify the expected effect. The studies on the effect of spatial frequency uncertainty (when the size is fixed) revealed that humans are equally efficient in the extrinsic uncertainty condition (Davis & Graham, 1981; Davis et al., 1983; Yager et al., 1984). However, Hübner (1996a,1996b) measured the effect as a function of target amplitude. He found the data is inconsistent with humans being equally efficient in both conditions but consistent with models that suggest humans are more efficient in the high extrinsic uncertainty condition. Later, Ohtani et al. (2002) measured psychometrics under similar conditions, which turned out to be inconsistent with Hübner 's data. In contrast, their measured psychometrics are consistent with humans being less efficient in high extrinsic uncertainty condition. Overall, it is not clear what these results predict about the scale uncertainty since when scale varies, size and spatial frequency vary in relation to each other. However, it seems the results presented here are consistent with the literature in general because two of three associated uncertainties in isolation are consistent with humans being more efficient in high extrinsic uncertainty conditions. Therefore, it seems human performance under the combination of uncertainty dimensions that is found to decrease the performance less than expected also results in a performance decrease smaller than expected. This suggests that previous experiments on the sub-set of uncertainties that will be presented in the experiment provide partly accurate predictions for the experiment.

Some experiments measure human performance when either the target scale (Oruc & Barton, 2010; Han et al., 2020) or target orientation (Guyonneau et al., 2006) is varied. However, they aimed to test whether the visual system detects objects invariant of these dimensions (see Logothetis & Sheinberg, 1996). There are conceptual differences between invariance and uncertainty, making it hard to compare those experiments with the experiment done here (see the general discussion).

5.3.2 Visual Mechanisms

5.3.2.1 High Extrinsic Uncertainty

The ideal observer under simultaneous target orientation and scale uncertainty is an extension of the classical template-matching observer. Detection under uncertainty about target dimensions generally requires identifying multiple visual patterns as the target. There is a template for each visual pattern, and a template response is computed with the template. The first stage of processing in these observer models (encoding stage) is similar to computations done by neurons in the primary visual cortex. The primary visual cortex receptive fields resemble these templates, and neural responses resemble template responses. The ideal observer can be implemented by appropriately pooling these template responses, which could be thought of as decoding strategy from a population of neural responses. The ideal observer's pooling operation requires exponentiation which is not biologically implausible (especially static spiking non-linearity might be relevant). However, it is also possible that the ideal observer can be approximated with heuristic operations. The standard heuristic is called the "max rule". The max rule is selecting the maximum of template responses to compare it with a criterion to make a decision. The MAX observer is shown to approximate the ideal observer closely under many scenarios, and the predictions of these two observer models are indistinguishable most of the time (Nolte & Jaarsma, 1967; Pelli, 1985; Graham et al., 1987). The previous studies used the MAX observer as an approximation to the ideal observer, and naturally, data favoring one favors the other (like Davis et al., 1983; Eckstein et al., 2000).

Recently, there have been more attempts to distinguish between these two model observers. Neri (2010) focused on more advanced metrics derived from the human data than simple accuracies and showed that early static non-linearity better explains the data

(the ideal observer in that experiment). Ma et al. (2015) reviewed various experiments comparing the MAX and the ideal observers. Their work revealed that if the reliability of different template responses is different, the simple MAX observer fails to perform as good as the ideal observer and does not explain the human data as good as the ideal observer. They introduced a version of MAX observer that works with likelihoods instead of template responses to approximate the ideal observer and explain the human data. However, they showed that even this updated version of the MAX observer fails to approximate the ideal observer in some conditions. These recent findings provide evidence for the ideal observers performing considerably better and explaining the data better under more realistic scenarios (where there are different reliabilities). There are multiple similarities between these recent studies and the experiment presented in Chapter 4. Here, I also focused on a more detailed description of the measured data (correct rejection and hit rates) and tested these two model observers under a more realistic scenario when different targets have different discriminability. Consistent with recent findings, I also found that the simple MAX observer does not explain the data and does not approximate the ideal observer. The normalization is required to explain the data and approximate the ideal observer (normalizing templates or template responses with the square root of energy of targets). It makes sure that standard deviations of target absent distributions are the same. Therefore, the normalization is similar to equalizing the reliability of different templates responses, similar to applying the max rule to local likelihoods. However, note that even though the ENM observer with a single criterion performs a bit worse than the ideal observer, as discussed previously, having different criteria for different levels of scales (for different reliabilities in some sense) makes it almost indistinguishable from the ideal observer (it is only shown in Appendix 4 and not tested against the human data). Therefore, as long as the MAX observer is properly normalized (and potentially extended with

dynamic criteria), this experiment does not discriminate between the two model observers regarding their absolute performance and ability to explain the human data.

Even though these recent findings favor the ideal observer, it is still unclear how the ideal observer-type of models can be implemented in a biologically plausible fashion. Neri (2010) implemented the ideal observer with early static non-linearity and argued that the visual system could utilize the computation in general. Early static non-linearity might help implement exponentiation computations required for the ideal observer's pooling operation. Moreover, there seems to be more evidence showing that the visual system utilizes the computation. Abbey and Eckstein (2006) used the method of classification images to measure human performance for three forced-choice tasks (detection, discrimination, and identification) in white noise. They have extended the ideal observer with a non-linear transducer and implicit position uncertainty to account for the differences between the ideal observer and data. The real world almost always involves extrinsic uncertainty; if the experiment does not involve enough extrinsic uncertainty, the internal limitations might dominate the data, so they must be included in the modeling. These limitations might be implicit position uncertainty and a non-linear transducer in their case. The non-linear transducer (similar to early static non-linearity) might be a built-in mechanism to deal with extrinsic uncertainties by approximating exponentiation. Moreover, Meese and Summers (2012) have compared various model observers, including variations of the MAX observer, to see which one best explains the contrast sensitivity of human performance in low and high levels of size uncertainty. They found that the observer model with a non-linear transducer and linear pooling best explains the data, and none of the variations of MAX observers they tested were unable to account for the data. Their findings are consistent with others and provide evidence for the non-linear transducer. Therefore, these recent findings suggest that visual processing involves a non-linear

transducer (or early static non-linearity). This computation may help deal with extrinsic uncertainty by approximating the exponentiation required for the ideal observer's pooling computation. Another possible way of implementing the ideal observer computations is suggested by Ma et al. (2011). They used the probabilistic population coding method to approximate the ideal observer for the visual search task involving distractors with biologically plausible computations. They showed that it is possible to approximate the ideal observer using divisive normalization and quadratic non-linearity operations.

For the max pooling of template responses (that is required for the MAX observer), various biologically plausible implementations are proposed (Angela et al., 2002). There is evidence showing that responses of some neural responses can be modeled as they are computing the maximum of their inputs (Lampl et al., 2004; also see Serre et al., 2005). Furthermore, the MAX operation is inspired and used in many successful algorithms for difficult object detection tasks (Riesenhuber & Poggio, 1999b; Serre et al., 2007). However, since the max operation approximates the ideal observer pooling under many scenarios (almost indistinguishable), its success in object recognition and its ability to explain the neural responses might not be surprising. These might not provide discriminative evidence for the max operation being utilized by the visual system.

5.3.2.2 Relation Between Low and High Extrinsic Uncertainty

Another approach to understanding visual mechanisms underlying visual detection under uncertainty is to compare various levels of uncertainty, specifically comparing low and high levels of extrinsic uncertainty. With this approach, the relationship between levels of uncertainty is also investigated. Thus, the observer models are not simply expected to explain the performance in the high extrinsic uncertainty experiment but should also account for the relationship between two conditions. I found that generating predictions

based on the efficiency (scalar factor) in the low extrinsic uncertainty experiment fails to achieve human performance in the high extrinsic uncertainty condition even if criteria are chosen to be optimal. As this result suggests, I found that humans are indeed more efficient under high uncertainty than low uncertainty. I already discussed how the result found here is consistent with the literature above (see 5.3.1 Behavioral Effects).

Intrinsic uncertainties can explain why humans are more efficient under high uncertainty because they cause a larger decrease in performance in low uncertainty conditions than in high uncertainty conditions. The existence of intrinsic uncertainties in visual processing was proposed and demonstrated previously (Tanner, 1961; Nachmias & Kocher, 1970; Theodore et al., 1974; Lasley & Cohn, 1981; Pelli, 1985). The visual system is intrinsically uncertain about the target and background properties even if properties are specified exactly because of reasons like memory limitation, noise in the representation of specified parameters, and imperfect knowledge of eye and head position. In this experiment, the target was always presented at the center at the exact same location. However, it was previously shown that the visual system cannot exactly use the prior information about the exact position and is intrinsically uncertain about the position of the target (Hess & Hayes, 1994; Manjeshwar & Wilson, 2001; Semizer & Michel, 2017). Therefore, the addition of intrinsic position uncertainty to the model observer is sensible because it is a known property of the visual system. I found that model observers with intrinsic position uncertainty better explain the whole data in terms of root-mean-squared error computed from hit and false alarm rates and ensure that overall percentage corrects are correctly predicted in both uncertainty levels. Thus, this result further revealed that intrinsic position uncertainty is specifically important for understanding the relations between different levels of uncertainty. However, intrinsic position uncertainty does not provide a better fit in terms of log-likelihood. Therefore, some other mechanisms may also

be involved in the processing that makes the visual system more efficient in high uncertainty (see the general discussion for other potential mechanisms that might underly the higher efficiency in higher uncertainty conditions).

I also found that a particular implementation of intrinsic position uncertainty better accounts for the data, which is when the position uncertainty scales with the target size. If the intrinsic uncertainty does not scale with the target's size, the performance associated with the largest target is not affected by the position uncertainty as much as the smallest target. However, they are similarly affected when the position uncertainty scales with the target size. A plausible explanation of why position uncertainty might scale as the target size is the following. Suppose the brain decodes from neurons with larger receptive fields when the target is large. In the retina, larger receptive fields tile visual space with the same overlap as small receptive fields; thus, their centers are more apart from each other in space. It is likely that receptive fields in the cortex inherited the same tiling principle up to some extent. In that case, the same uncertainty in cortical space (decoding from a nearby neuron) for both target sizes corresponds to a larger position uncertainty in space (in the real world) for a neuron with larger receptive fields. The same reasoning (the uncertainty being fixed in cortical space) could be applied when participants need to detect a target presented in the periphery since receptive fields sizes increase with eccentricity. Consistent with the idea, intrinsic position uncertainty has been previously shown to increase with eccentricity (Michel & Geisler, 2011).

5.4 General Discussion

5.4.1 The generic task

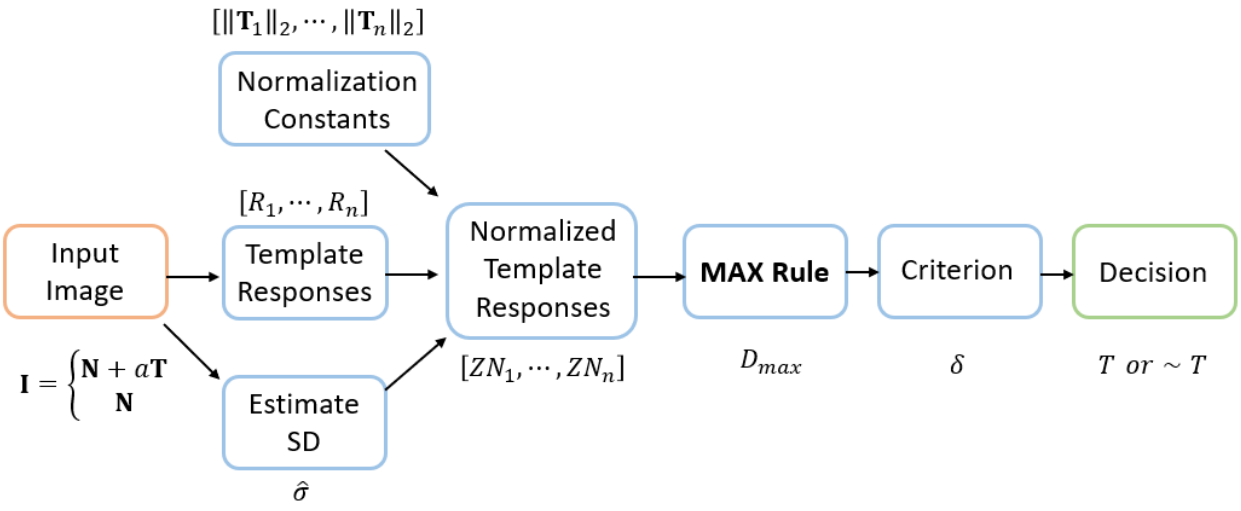
In Chapter 2, I defined the generic task as the visual detection of an additive target under simultaneous target amplitude, scale, orientation, and background contrast

uncertainty in white noise. I divided it into two sub-tasks: (i) target amplitude and background contrast uncertainty and (ii) target scale and target orientation uncertainty. The results of these sub-tasks are discussed above, and now I will focus on the generic case.

First, I will focus on the computational and informational aspects of the generic task. The ideal observer for the generic case is derived and presented in Chapter 2. Chapter 3 revealed the importance of normalization by the background contrast to approximate the ideal observer. Chapter 4 revealed the importance of normalizing the template responses or templates of the MAX observer with the square root of the target's energy to approximate the ideal observer. Thus, by combining these two heuristics, a model observer can be constructed for the generic task (Figure 5.1). Note that the dynamic criterion cannot replace background contrast normalization for this model observer because, without contrast normalization, the max operation is likely to result in a large sub-optimality. The Normalized MAX observer is likely to approximate the ideal observer that is derived in Chapter 2 (but it should be a bit worse in terms of absolute performance). Constructing a model observer indistinguishable from the ideal observer requires adding dynamic criteria to the Normalized MAX observer after the max rule operation. There should be a different criterion for each pair of background contrast and target scale.

It is interesting to compare the absolute performance of the ideal observer for the generic task to optimal performances measured in both Chapters. For example, the performance might not decrease compared to performance measured under the simultaneous target scale and orientation case because simultaneous background contrast and target amplitude only have a negligible effect on the performance. On the other hand,

this negligible effect might be amplified when it is combined with target scale and orientation uncertainty.



$$D_{max} = \max\left(\sum_{i=1}^{n} \frac{\mathbf{I} \cdot \mathbf{T}_i}{\|\mathbf{T}_i\|_2 \hat{\sigma}}\right)$$

Figure 5.1 Normalized MAX observer for the generic task.

The generic task includes considerable uncertainty, so it might not be trivial to build a very accurate model with modern computer vision tools because of its computational complexity. Convolutional neural networks recently attracted significant attention, and they sometimes fail on seemingly easy tasks (Baker et al., 2020; Reith & Wandell, 2020). This task would be one of the interesting cases where CNNs can be compared with the exact ideal observer (Reith & Wandell, 2020). Training a neural network will reveal how much and what kind of learning is required for a CNN to perform the task as well as the ideal observer (if it can ever reach that level). These types of comparisons might help us discover general principles about training neural networks and help us build a principled

understanding of the training process. The trained network might also provide insight into computational heuristics. For example, the single criterion, max operation, and energy-normalized templates are all easy to implement in a CNN. Thus, the network primarily needs to learn divisive normalization by background contrast to reach a similar performance to the Normalized MAX observer. It would be interesting to see how it will be implemented in the network. Furthermore, suppose it surpasses the normalized MAX observer performance and becomes indistinguishable from the ideal observer. In that case, its structure might provide some insight into how the ideal observer can be implemented with classic CNN-type operations.

When human performance is measured for the generic task, it is clear that the data will provide a rich structure of hit and false alarm rates that will be fruitful to discriminate between model observers. Moreover, it would be interesting to compare the efficiency (scale factor) estimated in the generic task to scale factors found in Chapter 3 and Chapter 4. Scale factors are around 0.4 (see Figures 3.8 and 4.11b) under high uncertainty. However, as shown in Chapter 4, the scale factor under low uncertainty is around 0.3 (Figure 4.11b). Since the uncertainty will be considerably higher in this experiment, it would be interesting to see whether the scale factor will be even higher. However, the computational mechanisms devoted to solving the combination of these two sub-tasks might be similar to those required for the individual components; thus, the efficiency might not change. Moreover, there is another force acting against the increase in efficiency. As computational difficulty increases, the optimal solution might become so computationally intensive that the visual system might use sub-optimal heuristics, which cause efficiency to decrease. Therefore, it is unclear how the efficiency will compare to the ones found in both chapters.

I discussed the neurological basis of heuristic computations in previous sections. Because of that, for the generic task, the normalized MAX observer seems to be a biologically plausible model. Also, as discussed previously, multiple hypotheses about how the ideal observer might be implemented are proposed in the literature. Most of the neuroscientific discussion focuses on biologically plausible computational mechanisms and their existence in the brain for understanding visual detection under uncertainty. There is less interest in recording brain activity under various levels of uncertainty. These recordings of the population of neurons might reveal how the activity changes (for example, in the primary visual cortex) for different levels of uncertainty. For example, Martens and Blake (1980) measured the behavioral contrast thresholds of cats for the detection of grating patterns under various levels of uncertainty. They found that thresholds are elevated as theoretically expected (similar to human thresholds, the performance decreases at the expected rate). However, they did not extend their research by recording brain activity.

Another example is a relatively recent experiment that recorded neurons of the rat barrel cortex when there was temporal uncertainty about the time of the event (Stüttgen & Schwarz, 2008). The task was to detect a whisker deflection, and the occurrence time of whisker deflection is varied. They found that individual neurons are unlikely to underlie psychophysical sensitivity under uncertainty, even though they might underlie the sensitivity under low uncertainty conditions. They showed that a few neurons together could reach the level of psychophysical sensitivity. However, they have not compared neurons' activity for various levels of uncertainty. As reviewed in previous sections, there are multiple hypotheses about how the prior information is used when it is specified (low uncertainty conditions). Attentional capacity limitation, uncertainty reduction, and intrinsic uncertainties might have different predictions about how brain activity should look at

various uncertainty levels. These predictions might be tested by measuring brain activity under various levels of uncertainty.

5.4.2 Challenges and Resolutions

In Chapter 2, I specifically identified three challenges for measuring and modeling detection under high levels of uncertainty. I proposed that the explicit simulations of the ideal observer and other model observers (i.e., constructing and simulating image processing models that take the image and produce a response) will help us overcome these challenges. Throughout this thesis, I developed methods for the explicit simulation of model observers and applied them to overcome these challenges. Now, I will address three challenges with respect to resolutions provided in this work and discuss methods developed in this work for incorporating image processing models into the analysis.

Firstly, the previous literature almost exclusively focused on comparing various levels of uncertainty (see Chapter 2). Even though this is a valuable approach for understanding detection under uncertainty and the relation between various levels of uncertainty, it requires an enormous data collection effort for high levels of uncertainty. This thesis provides evidence for the richness of the data measured in a single block under high or modest uncertainty to distinguish between model observers. In both Chapters, valuable insights are gained about the computational mechanisms by simply using a single block of an experiment to discriminate between model observers. Specifically, the explicit simulation of model observers (image processing models) allows us to fit single blocks of the uncertainty experiment, which have rich structures of hit and correct rejection rates. Therefore, with the explicit simulation of model observers, distinguishing between model observers with a single block of data provides another approach for studying high levels of uncertainty that does not suffer from the same data collection challenge.

Secondly, calculating d-prime (sensitivity) and criterion with the classic signal-detection framework might be misleading for a couple of reasons discussed in Chapter 2, especially when there are high levels of uncertainty. With the explicit simulation of model observers (image processing models), model observers' behavior is simulated in more detail. For example, it is possible to calculate hit and correct rejections for each level of target and background variations. Varying a bias and sensitivity parameters affect these predictions; thus, the sensitivity and bias of human observers can be estimated by fitting the ideal observer to the human data (specifically hit and correct rejection rates). In Chapters 3 and 4, the estimated scale factor and bias (which can be interpreted as the subjective prior probability of the target being present) are reported by fitting model observers to observers' hit and correct rejection rates. These two parameters are specific to the model observer, so they require a good fit for the ideal observer. Thus, these estimations are implicitly done under many assumptions about the nature of the process, unlike classic signal detection methods. However, they do not suffer from the weakness of classic signal detection assumptions. Especially, a well-fitted ideal observer can estimate efficiency and bias without suffering from the same weakness of classic signal detection theory analysis under high levels of uncertainty. Since classic signal detection theory assumptions are expected to be violated more drastically for higher levels of uncertainty, this approach to estimating these parameters could still make it possible to compare efficiency across various levels of uncertainty.

Thirdly, comparing the efficiency across levels of uncertainty might result in finding higher efficiency under high uncertainty (as discussed before) for the experiments with high extrinsic uncertainty. Even though the most common explanation is the existence of intrinsic uncertainties, the specific dimensions of intrinsic uncertainties are rarely simulated (Hübner 1996a;1996b), except for position uncertainty (Hess & Hayes, 1994;

Manjeshwar & Wilson, 2001; Michel & Geisler; 2011; Semizer & Michel, 2017). The intrinsic uncertainties are crucial to test and identify because one of the primary aims of conducting simplistic laboratory experiments is to generalize the results to more naturalistic conditions. When humans are more efficient under high uncertainty, the predictions generated from lower uncertainty conditions (even the highest performance possible based on lower uncertainty conditions) sometimes fail to reach the measured human performance (see Chapter 4, and Walshe &Geisler, 2021). Such poor predictions for more naturalistic conditions make it necessary to incorporate intrinsic uncertainties (or other hypotheses) into the modeling. In Chapter 4, I simulated model observers with intrinsic position uncertainty, but I also tested other dimensions of intrinsic uncertainty: target scale, target orientation, and their combination because the visual system could be intrinsically uncertain about these dimensions as well (Bravo & Farid, 2009). Implementing explicit simulation of observer models makes it easy to simulate many dimensions of intrinsic uncertainties because it only requires the addition of new templates to image processing models without changing the structure of simulations.

It is also essential to formulate other hypotheses to explain why humans are more efficient under higher levels of uncertainty. We can start by considering why humans are inefficient (compared to the ideal) in the low uncertainty in the first place. Among these hypotheses that explain why humans are inefficient under low uncertainty, there are good candidates that could potentially explain why humans are more efficient under higher levels of uncertainty. For example, the same amount of noise in the criterion placement (Wickelgren, 1968) might predict different magnitudes of the effect for different levels of uncertainty. Specifically, a small variation in low uncertainty conditions might be more detrimental to performance than the same variation under high uncertainty because under high uncertainty, distributions of the decision variable are likely to be wider (have a higher

standard deviation). It is also possible that the usage of sub-optimal templates might explain the effect (Beck et al., 2012). The reasoning behind it is very similar to intrinsic uncertainties. Regardless of the number of sub-optimal templates decoded to make a decision under low uncertainty, the same level of sub-optimality of the templates might have a larger effect on the performance under low uncertainty compared to the condition where there are multiple visual patterns to detect (high uncertainty). The reason is that the variation due to extrinsic uncertainties is likely to dominate the variation due to the sub-optimality of templates under high uncertainty. These alternatives could be incorporated into the explicit simulation of model observers because it is easy to replace some parts of the computations in these image processing models.

All these resolutions require constructing image processing model observers and explicit simulation of model observers' predictions (hit and correct rejection rates), especially the ideal observer's predictions. The simulation (and derivation) of model observers and directly fitting their predictions to the data is necessary to (i) use the rich structure in the data to distinguish between model observers, (ii) estimate efficiency scalar and sub-optimal criterion with the ideal observer, and (iii) implementing various hypotheses about the high efficiency in high uncertainty and testing these together with models with intrinsic uncertainties. Thus, throughout the thesis, I have derived model observers and developed faster simulations methods and fitting procedures.

In this thesis, I have developed and adapted methods for directly fitting the hit and correct rejection rates of human observers with the predicted hit and correct rejection rates of model observers. Chapter 3 describes an extension of the classic psychometric fitting method to fit hit and correct rejection rates by treating the model observers' hit and correct rejection rates as probabilities. There was a weakness of this fitting method because when a finite number of trials are used to determine the model observer's hit and correct rejection

rates, these rates might be zero or a hundred percent, making the likelihood negative infinity. Chapter 4 proposes a solution to this problem, with the assumption that there would be a single different response if the number of trials in simulations is doubled.

Because all these resolutions depend on fitting while varying multiple parameters, faster simulations are essential for the viability of these resolutions. Furthermore, as the number of uncertainty dimensions increases, the number of template responses that need to be pooled increases. Therefore, computing thousands of template responses in each trial to determine the optimal performance make the problem computationally intensive, especially because there are thousands of trials in a convincing simulation in the first place. This computational intensity posed a problem for simulating the ideal observer and sub-optimal model observers. Recent advances in computational processing definitely help, but I developed methods to speed up the simulations further. In Chapter 3, the analytical equations are derived for hit and correct rejection rates, showing that it is possible to derive the analytical formulas under some circumstances. These formulas might provide insights for deriving similar formulas for different types of backgrounds or when these two dimensions of uncertainty are combined with other dimensions. A method developed in Chapter 4 that relies on pre-computed variables for additive targets in white noise speeds up simulations significantly. The method is applicable to any uncertainty dimension that generates multiple visual patterns that should be classified as the target. It can handle four dimensions of uncertainty, where each one has ten possible levels. The method can potentially be extended to different types of backgrounds. Therefore, these methods provide a starting point for making the explicit simulation of model observers even faster. Together with fitting methods, I think methods developed in this thesis make these resolutions viable for future experiments too.

5.3.3 Invariance and Uncertainty

It is interesting to compare the visual detection task under the highest level of uncertainty (real-world conditions) with naturalistic object identification tasks. The object identification task is defined as asking participants to categorize various objects; while the task generally involves only discrimination, it could possibly include trials where there is no object presented at all. Since there are too many categories, the computational crux of the problem is to differentiate between categories. In contrast, participants are only looking for a specific target in the visual detection task. Therefore, having this aim can shape the search to maximally differentiate the object category of interest from any other visual patterns. Thus, the visual detection task does not require the recognition of other objects (even though recognizing all objects is also a viable strategy in the real world). The similarity between these experiments is that an object can be presented in various scales and orientations but still needs to be classified as the same target in both of them. Therefore, the object needs to be identified invariant of these dimensions. However, there are multiple proposals for a criterion defining when the invariance is achieved. Recently, Han et al. (2020) distinguished between two types of invariances: intrinsic invariance and example-based invariance. The uncertainty experiment will generally be classified as example-based invariance because possible variations of the target are shown before the experiment. Thus, in an uncertainty experiment, the observers know (are shown) all possible variations (for example, see the Methods in Chapter 4). Han et al. (2020) argued that this type of invariance is trivial since any recognition system with sufficient memory and large training data can solve this problem. In contrast, they are more interested in intrinsic invariance, which is the ability to recognize various versions of the object after seeing only a couple of example variations of that object. Intrinsic invariance is definitely an interesting version of invariance; however, I think the first one is not trivial to achieve. Similar to the

description of real-world uncertainty in Chapter 2, DiCarlo et al. (2012) describe the variation in the following paragraph:

> In the real world, each encounter with an object is almost entirely unique, because of identity-preserving image transformations. Specifically, the vast array of images caused by objects that should receive the same label (e.g., "car," Figure 1) results from the variability of the world and the observer: each object can be encountered at any location on the retina (position variability), at a range of distances (scale variability), at many angles relative to the observer (pose variability), at a range lighting conditions (illumination variability), and in new visual contexts (clutter variability). Moreover, some objects are deformable in shape (e.g., bodies and faces), and often we need to group varying three-dimensional shapes into a common category such as "cars," "faces," or "dogs" (intraclass variability). In sum, each encounter of the same object activates an entirely different retinal response pattern and the task of the visual system is to somehow establish the equivalence of all of these response patterns while, at the same time, not confuse any of them with images of all other possible objects (see Figure 1). (p. 417)

Therefore, there are almost infinitely many different visual patterns in the real world that need to be classified as the target; simply remembering them in a sufficient memory is not a viable computational strategy. However, it is still unclear what the criterion should be for calling a detection behavior invariant. To the best of my knowledge, there is no consensus on the exact definition of invariance in the field (see DiCarlo & Cox, 2007; Pinto et al., 2008; Bart & Hegdé, 2012). DiCarlo et al. (2012) state that object detection research aims to replicate human capabilities, but when is the visual system processing considered to be invariant? For example, if an object is far away, its image in the retina is small, making it harder to see. Therefore, one should not expect the

detectability of all scales to be the same. The same applies to a discrimination experiment, too; some views of the objects are easier to mislabel. Therefore, without knowing the optimal (expected) performance (information content), it is hard to determine whether human detection performance is affected by the variation in target and background properties or whether human detection is invariant to these variations. However, this requires derivation of the ideal observer, which is still implausible for these types of difficult tasks. The principled approach provided here also provides insight into the discussion of the definition of invariance and provides expectations about the effects of variations in more difficult tasks. Applying the same principled approach to slightly more difficult tasks could be a fruitful future direction for invariance research too.

5.3.4 Future Directions

This thesis focuses on the visual detection task that involves simultaneous multiple dimensions of uncertainty (more naturalistic conditions). Measuring human performance under more naturalistic conditions is part of a general approach outlined in Chapter 1. I showed that the data measured in a single block of an experiment in more naturalistic conditions is rich enough to distinguish between model observers. Thus, this thesis provides another good example that this approach can be fruitful in providing important insight into visual processing. Therefore, with the methods developed in this thesis, focusing on even more naturalistic conditions in the future is likely to be fruitful.

The most naturalistic experiment measures the human performance for a real-world task in which subjects are looking for a target object in a real 3D environment. The extrinsic uncertainty in a real-world task can be characterized and measured as a combination of uncertainty dimensions that generate the variation and the prior probability of variations. By the definition of uncertainty that I endorsed for this thesis (see Chapter 2), the measured

level of uncertainty in a real-world task corresponds to the highest level of extrinsic uncertainty. My definition roughly hopes to scale with the amount of computational complexity due to extrinsic uncertainty. For example, even though position uncertainty has detrimental consequences on performance, it is only a modest level of uncertainty. It is because optimal performance is achieved with relatively simple computations compared to the computations required to deal with higher levels of uncertainty involving multiple simultaneous dimensions. Thus, measuring uncertainties in real world tasks and quantifying the computational complexity associated with them could be useful for better understanding what the more naturalistic conditions are to conduct experiments under these conditions.

This thesis also applies the ideal observer analysis for the visual detection task involving multiple uncertainty dimensions. Compared to other studies that focus on more naturalistic conditions, the one distinctive feature of this work is that I have derived and tested the ideal observers that are processing images. As previously discussed in Chapter 1, the ideal-observer analysis provides a principled approach for studying naturalistic tasks because visual systems evolve toward maximizing accuracy in order to survive. Firstly, the ideal observers derived for various tasks in this thesis explained all the human data fairly well. Thus, the ideal-observer analysis continues to provide a normative approach that unifies our understanding of measured performance in various tasks. Secondly, an important part of the ideal-observer analysis is replacing components of processing with the sub-optimal ones. These biologically plausible but sub-optimal heuristics might explain the human data better because even if the processing evolves toward maximizing accuracy, there are biological limitations. Here, none of the sub-optimal observers explained the data better than the ideal observer. However, it is shown that relatively simple heuristic computations can approximate the ideal observer. By deriving the ideal observer and

simulating it, one better understands the computations required to perform the task. Thus, it is easier to simplify some computations and find heuristics that could approximate the ideal observer in general. Finally, the ideal observer marks the maximum achievable performance (optimal performance) and provides a benchmark for both the human performance (efficiency analysis) and the performance of sub-optimal model observers. In sum, the ideal observer analysis continues to be a principled fruitful approach to understanding visual detection under uncertainty.

The knowledge about optimal performance also allows us to compare the effect of different dimensions of uncertainty on the performance. For example, performance can be calculated as a function of overall target strength. These functions reveal differences between different dimensions of uncertainty. I recently started deriving analytical formulas for calculating the performance as a function of overall target strength with simple approximations (not shown here, but it is ongoing research). Even though different dimensions have a different effect on performance as a function of overall target strength, it might be useful to quantify them with a single measure. One possible strategy is to quantify them all in terms of the equivalent number of orthogonal channels that produce a similar decrease in performance. It would be interesting to see whether patterns produced by uncertainty due to the number of orthogonal channels can approximate the patterns produced by the various dimensions of uncertainty. It is also possible to fit an analytical expression like the UNI function (Geisler, 2018) to these functions. It is possible to compare fitted parameters across different dimensions of uncertainty. Effectively simulating various dimensions of uncertainty with analytical formulas and quantifying the effect of various dimensions of uncertainty with a unifying measure might provide a general framework for understanding the effects of uncertainty.

In conclusion, this thesis exemplified how the conceptual approach that combines naturalistic tasks with the ideal observer analysis (described in Chapter 1) continues to provide insight into visual processing. The findings of these experiments specifically helped us understand more about how the visual system deals with extreme levels of uncertainty in the real world. With the methods developed in this thesis, measuring and modeling human performance under high levels of uncertainty resulted in new findings and insights. These findings and methods developed in the thesis reveal the potential value of studying visual detection under high levels of uncertainty in the future. In future experiments, moving forward to more realistic conditions with a principled approach will be critical for a deeper understanding of visual processing in humans and other organisms.

Appendices

APPENDIX 1: THE IDEAL OBSERVER FOR AMPLITUDE AND CONTRAST UNCERTAINTY

Single Criterion is Optimal under Amplitude Uncertainty

I assumed that the prior probability of the target being present and prior probabilities of target amplitudes and background contrasts are available to the ideal observer. Furthermore, the ideal observer is assumed to know the background contrast because the background patch is big enough to estimate it precisely. The target location is also assumed to be known (no position uncertainty).

When the background contrast is known to the observer, the problem reduces to making an ideal decision under amplitude uncertainty because the appropriate ideal observer can be evaluated for each background contrast level. The ideal observer responds target-present if the ratio of the likelihood of target present (H_1) to absent (H_0) is more than the ratio of the prior probability of target absent to present.

$$\frac{P(I \mid H_1)}{P(I \mid H_0)} > \frac{P(H_0)}{P(H_1)}$$

Ω_s is the target region, and I_x is the gray level of the location x in the image. When the target is present, luminance in any location is the sum of the target gray level (T_x) and the gray level of the Gaussian noise background (mean of 0 and standard deviation of σ_n) that is sampled independently for each location. In a target-present trial, either one of the amplitudes (a_m) is used, so they are disjoint events and exhaustive. Therefore, the likelihood can be expressed as a sum of the probabilities of these disjoint events. The likelihood of target-present and target-absent are given by:

$$P(I \mid H_1) = \sum_m \left(\prod_{x \in \Omega_s} G(I_x \mid a_m T_x, \sigma_n) \right) P(a_m)$$

$$P(I \mid H_0) = \prod_{x \in \Omega_s} G(I_x \mid 0, \sigma_n)$$

Explicitly writing out Gaussian distributions lead to:

$$\frac{P(I \mid H_1)}{P(I \mid H_0)} = \sum_m e^{\frac{a_m}{\sigma_n^2} \sum_{x \in \Omega_s} (I_x T_x)} \, e^{\frac{-a_m^2}{2\sigma_n^2} \sum_{x \in \Omega_s} (T_x)^2} \, P(a_m)$$

In the experiment, every term is fixed except the dot product of the target region with the target ($\sum_{x \in \Omega_s}(I_x T_x)$). Therefore, the dot product is a sufficient statistic for the likelihood ratio (captures all the information in the likelihood ratio). Also, the ratio increases monotonically with the dot product; therefore, a single optimal criterion on this metric would be equivalent to the maximum posterior estimation. There are no further simplifications for the derivation of the optimal criterion, and its value depends on the background contrast. Therefore, the ideal observer can be implemented by comparing the dot product to a dynamic criterion that changes based on the background contrast level.

Mean and standard deviation of template responses

For a given pair of target amplitude and background contrast, it is possible to derive the mean and standard deviation of distributions of dot products ($\sum_{x \in \Omega_s}(I_x T_x)$). If the target is present and amplitude is a_m, the gray level of the image at any point (I_x) is distributed as Gaussian with a mean of $a_m \; T_x$ and standard deviation of σ_n. The dot product is also distributed as Gaussian with mean $a_m \sum_{x \in \Omega_s} T_x^2$ and standard deviation of $\sigma_n \sqrt{\sum_{x \in \Omega_s} T_x^2}$.

$$I_x \sim G(a_m T_x, \sigma_n)$$

$$I_x T_x \sim G(a_m T_x^2, \sigma_n T_x)$$

$$\sum_{x \in \Omega_s} I_x T_x \sim G\left(a_m \sum_{x \in \Omega_s} T_x^2 \, , \sqrt{\sum_{x \in \Omega_s} (\sigma_n T_x)^2}\right) \text{ (Variances add)}$$

If the target is absent, the gray level of the image at any point (I_x) is distributed as Gaussian with a mean of zero and standard deviation of σ_n. The dot product is also distributed as Gaussian with mean of zero and standard deviation of $\sigma_n\sqrt{\sum_{x\in\Omega_s} T_x^2}$.

A template is generated by normalizing the target with a constant (k). Note that these operations only change the location of the optimal criterion and don't change the result presented above about a single criterion being optimal. In my case, I normalized the target by the energy of the target to generate a template.

$$Tm_x = \frac{T_x}{k}$$

$$k = \sum_{x\in\Omega_s} T_x^2$$

The mean and standard deviation of template responses can be derived from the mean and standard deviation of dot products by dividing them by the normalization constant. Distributions of template responses, if the target absent is present, are Gaussian with means of a_m and standard deviations of $\sigma_n \frac{1}{\sqrt{\sum_{x\in\Omega_s} T_x^2}}$. The distribution of template responses, if the target absent is absent, is a Gaussian with mean of 0 and standard deviation of $\sigma_n \frac{1}{\sqrt{\sum_{x\in\Omega_s} T_x^2}}$.

APPENDIX 2: MORE COMPUTATIONAL RESULTS FOR SIMULTANEOUS TARGET AMPLITUDE AND BACKGROUND CONTRAST UNCERTAINTY

Comparison of three model observers

I evaluated the performance of three model observers with the optimal criteria for various levels of prior probability of the target being present that ranging from 0.05 to 0.95. The amplitude range is scaled down with a single scalar value to evaluate models for a particular amplitude range scalar level. The range of scalars (amplitude range scalars) was between 0.005 to 0.5. I assumed that prior probability distributions of target amplitude and background contrast levels are uniform. The overall performance is measured by two metrics: simple percent correct difference between the ideal observer and sub-optimal observer models (TM and NTM) and the percentage of the maximum increase (the maximum increase depends on the optimal performance and the target prior). To calculate the percent correct difference, the sub-optimal observer's percentage correct is subtracted from the percentage correct of the ideal observer in any condition (Figure A2.1).

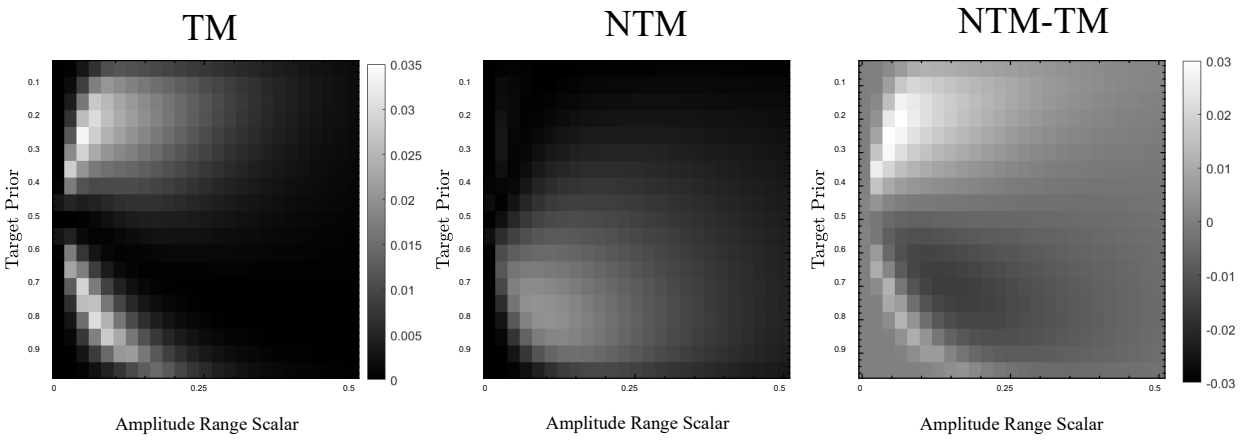


Figure A2.1 Percent correct difference from the optimal performance.

The percentage correct difference of two sub-optimal observer models from the ideal observer is shown in the first two panels as a function of the target prior and amplitude range scalar. The third panel shows the difference between the first two panels. Positive numbers correspond to conditions that the NTM observer is closer to the ideal observer than the TM observer. Note that the different grayscale is provided for the third panel.

The highest difference between the TM observer and the ideal observer is in the low amplitude range scalars when the target prior is low. On the contrary, the highest difference for the NTM observer is in relatively higher amplitude range scalars when the target prior is high. The total percent correct difference for the NTM observer (3.06) is slightly higher than the TM observer (2.78). Even though the TM observer is not substantially better than the NTM observer in any condition, for most of the high-amplitude range scalars, the TM observer is slightly better than the NTM observer. In contrast, when the target prior is low, and the amplitude range scalar is low, the NTM observer is substantially better than the TM observer (up to 3 percent difference). However, these percent differences are not enough to fully capture the difference between NTM observer and TM observer because a ten percent correct difference might be a small or a large

portion of the difference between chance and optimal performance. Therefore, the fraction of the percent correct difference to the difference between chance and optimal performance might play a key role in task performance. For example, consider an animal hunting for an insect. If the prior probability of an insect being present in a single fixation location is low, like 0.2, the chance performance is already 0.8. Suppose the optimal performance is 0.9; a ten percent increase in performance, in this case, made an animal the best predator from being the worst predator. However, suppose the target prior is 0.5 in the first place and performance is 0.8 percent correct. In that case, the increase will correspond to only a slight advantage gained against the other competing animals for the resources. Therefore, a ten percent increase under a low target prior might determine whether the animal dominates the environment or does not survive. Because under low target prior, there are a smaller number of trials that matter and are not dictated by the prior information in the first place. I calculated a second metric (relative performance measure) that represents the fraction of the maximum performance increase from the chance that the model observer captures. To calculate the percentage of maximum increase (I), I first scale percentage corrects for all target prior conditions between 0 and 1 (subtract the performance expected by prior, PP, divide by one minus performance expected by the prior). Then, the scaled percent corrects (S_s) of sub-optimal observers is subtracted from the scaled percent corrects of the ideal observer (S_i) and divided by it. Any value that is smaller than 10^-9 is fixed to be zero. Zero divided by zero, shown as zero (Figure A2.2).

$$S = \frac{(PC - PP)}{(1 - PP)}$$

$$I = \frac{S_i - S_s}{S_i}$$

Both observer models capture most of the maximum performance increase when the amplitude range scalar is high. However, the TM observer fails to capture a substantial amount of maximum performance increase when the target prior is low, and the overall percentage correct is around 75 percent. On the other hand, the NTM observer only fails to capture a significant degree of the maximum performance increase if the target prior is high. On average, the TM observer fails to capture 17 percent of the maximum performance increase, whereas the NTM observer only fails to capture 12 percent. This suggests that for most higher amplitude range scalars, which the TM observer does better in terms of percent correct, the difference in percent correct only constitutes a small amount of the maximum performance increase for these conditions. However, for almost all the conditions that NTM does perform better in terms of percent correct, the difference between models constitutes a large portion of maximum performance increase. At the threshold efficiencies (when the overall percentage correct is around 75 percent) and when the target prior is low, the three percent correct difference constitutes more than seventy percent of the maximum performance increase, so the NTM observer does capture 70 percent more of the maximum performance increase compared to TM observer.

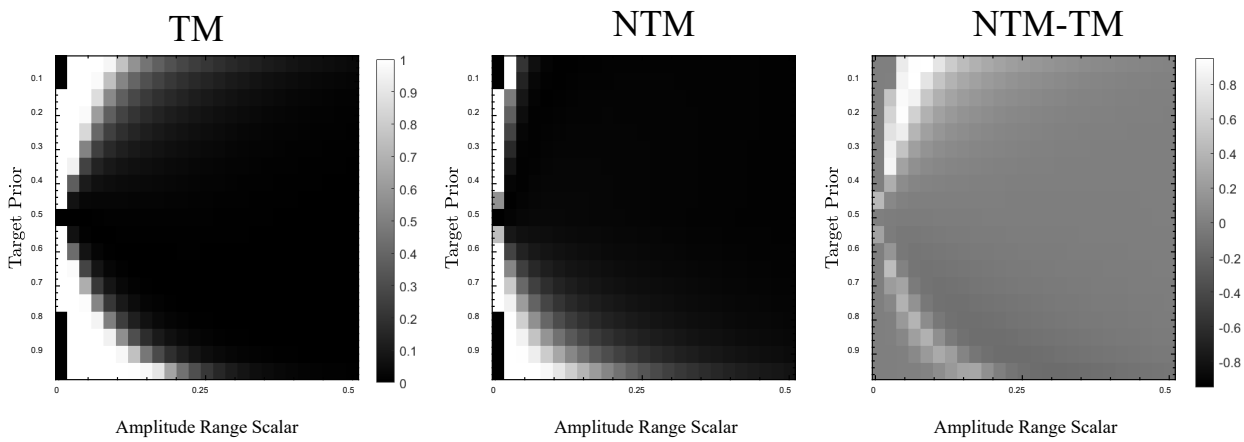


Figure A2.2 The relative performance measure for sub-optimal observers.

The percentage of a maximum performance increase for sub-optimal observer models is shown in the first two panels as a function of the target prior and amplitude range scalar. The third panel shows the difference between the first two panels. Positive numbers correspond to conditions in which the NTM is performing better than the TM observer. Note that the different grayscale is provided for the third panel.

In sum, I found that the NTM observer performs better than the TM observer for low target priors. Also, it well-approximates the ideal observer (on average, TM fails to capture 18 percent of the maximum increase, whereas NTM only does this misses 6 percent). On the other hand, when the target prior is high and models are operating with high amplitudes (high amplitude range scalars), I found that TM observer is only slightly better than NTM (maximum better by 1.5 percent correct that is 7 percent of the maximum increase, at maximum captures 15 percent more of the maximum increase where the percent correct difference is less than 1 percent correct).

Only Background Contrast Uncertainty

Both measures revealed very similar trends when there is only background contrast uncertainty. However, the difference between the NTM and TM observers shrinks in general favor of the TM observer. The total percentage correct difference (3.95) for the NTM observer is bigger than the total percent correct difference for the TM observer (2.22,

Figure A2.3A). On average, the TM observer fails to capture 15 percent of the maximum performance increase, whereas the NTM observer only fails to capture 12 percent (Figure A2.3B).

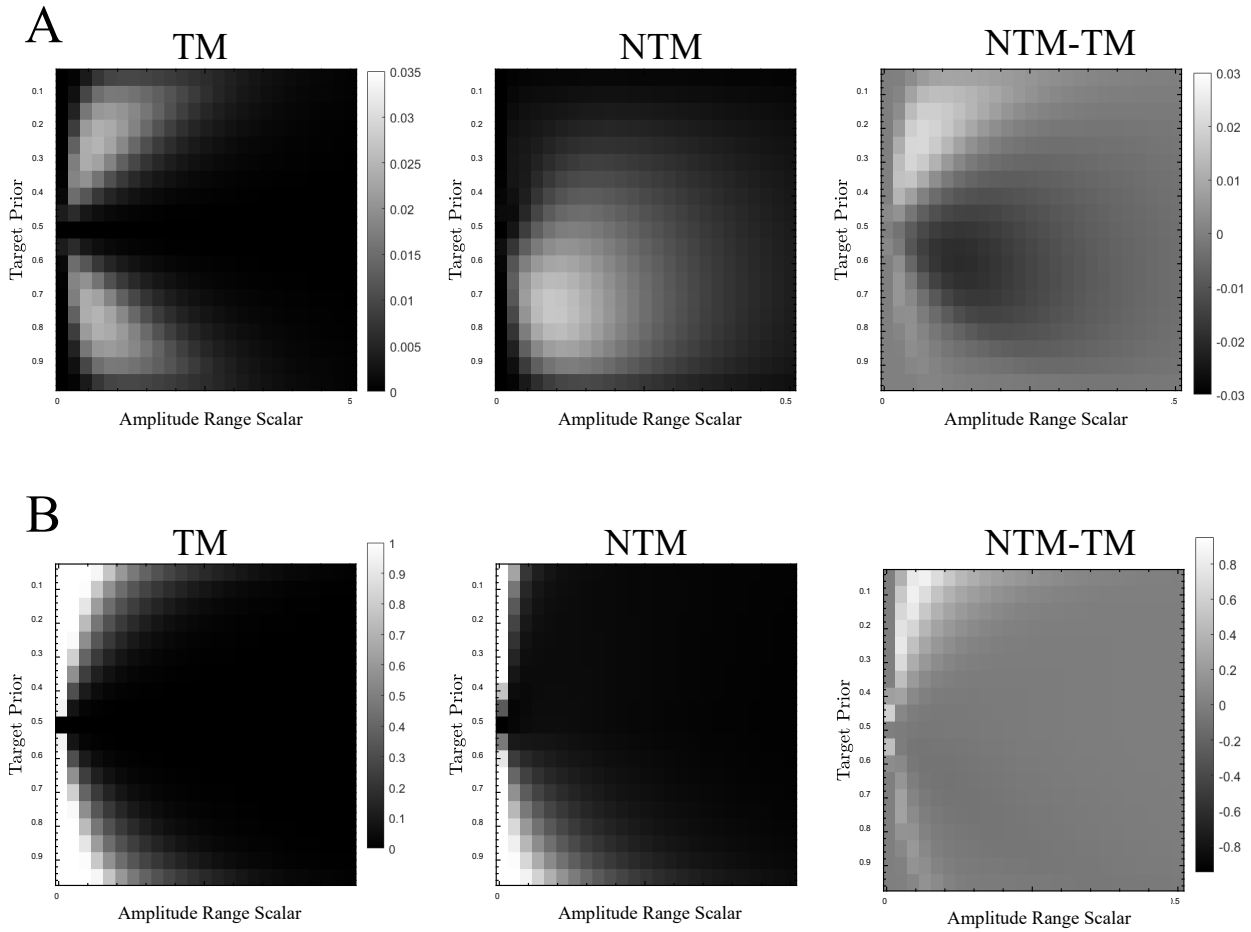


Figure A2.3 Performance under only background contrast uncertainty.

A The simple percentage correct difference between two sub-optimal model observers is shown in the first two panels as a function of the target prior and amplitude range scalar. The third panel shows the difference between the first two panels. **B** The percentage of the maximum performance increase for sub-optimal observer models is shown in the first two panels as a function of the target prior and amplitude range scalar. The third panel shows the difference between the first two panels.

The Effect of Amplitude Uncertainty

To quantify the effect of amplitude uncertainty, I calculated both metrics for the ideal observer under low uncertainty, and for the ideal observer under amplitude uncertainty. The overall effect of amplitude uncertainty is small (less than 1 percent correct difference and less than ten percent of the maximum performance increase in general). However, the effect strikingly depends on the target prior for both metrics, and it is negligible when the target prior is low, whereas when the target prior is high, it goes up to a loss of 35 percent of the maximum performance increase (Figure A2.4).

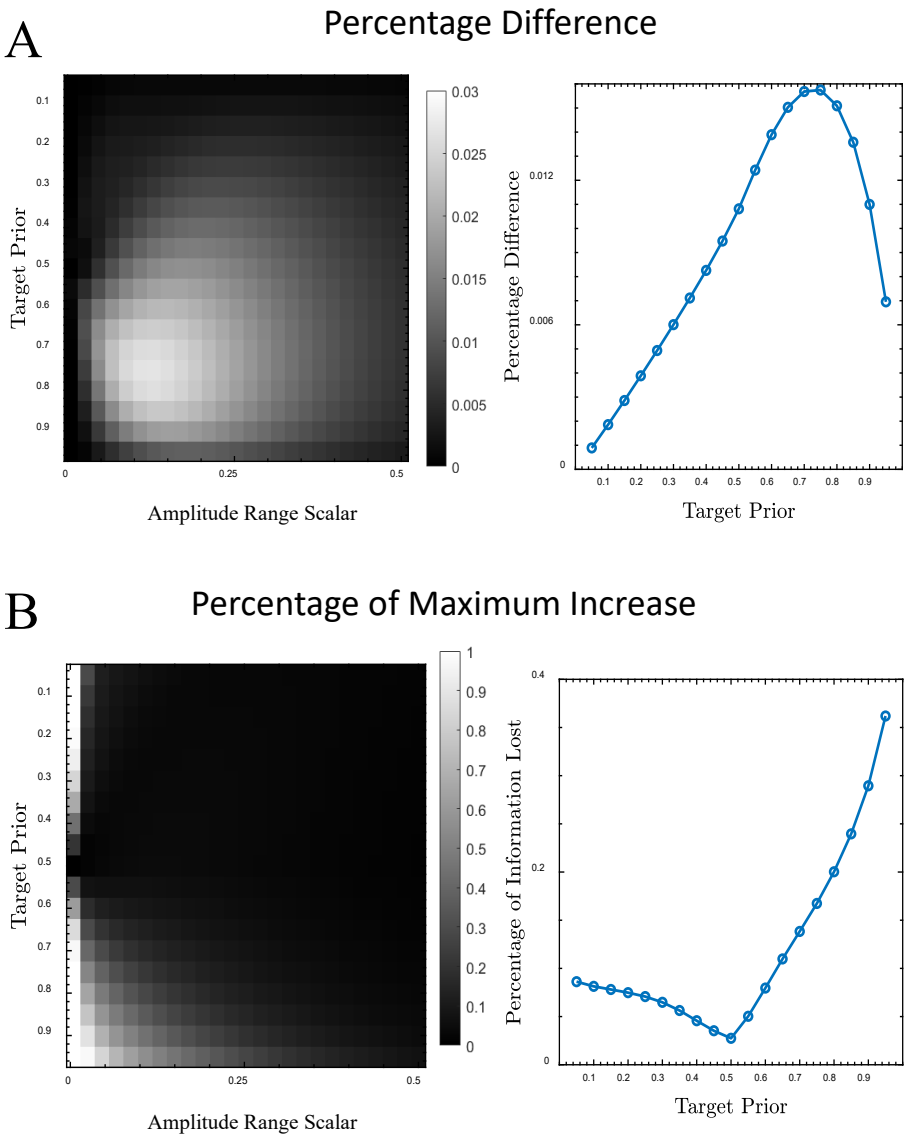


Figure A2.4 The effect of amplitude uncertainty.

Two measures of performance are calculated to compare the ideal observer when there is no uncertainty to the ideal observer when there is only amplitude uncertainty. **A** The simple percentage correct difference for each target prior and amplitude range scalar is shown in the image matrix. The plot shows the percent correct difference averaged over efficiencies as a function of the target prior. **B** The percentage of maximum increase for each target prior and amplitude range scalar. The plot shows the percentage of maximum increase over efficiencies as a function of the target prior.

APPENDIX 3: THE SIMULATION PROCEDURE OF MODEL OBSERVERS FOR PATTERN UNCERTAINTY

Derivation of the ideal observer

I assumed that the prior probability of the target being present and prior probabilities of target scales and orientation are available to the ideal observer. Furthermore, the ideal observer is assumed to know the background contrast and target amplitude fixed in a single simulated experiment. The target location is also assumed to be known (no position uncertainty).

The ideal observer responds target-present if the ratio of the likelihood of target present (H_1) to absent (H_0) is more than the ratio of the prior probability of target absent to present.

$$\frac{P(I \mid H_1)}{P(I \mid H_0)} > \frac{P(H_0)}{P(H_1)}$$

Ω_s is the target region, and I_x is the luminance of the location x in the image. When the target is present, the gray level in any location is the sum of the target's gray level (T_x^m) and the gray level of the Gaussian noise background (mean of 0 and standard deviation of σ_n) that is sampled independently for each location. In a target-present trial, either one of the targets (T^m) is presented with amplitude of a, so they are disjoint events and exhaustive. Therefore, the likelihood can be expressed as a sum of these disjoint events. The likelihoods of target-present and target-absent are given by:

$$P(I \mid H_1) = \sum_m \left(\prod_{x \in \Omega_s} G(I_x \mid a\, T_x^m, \sigma_n) \right) P(T^m)$$

$$(I \mid H_0) = \prod_{x \in \Omega_s} G(I_x \mid 0, \sigma_n)$$

Explicitly writing out Gaussian distributions lead to:

$$\frac{P(I \mid H_1)}{P(I \mid H_0)} = \sum_m e^{\frac{a_m}{\sigma_n^2} \sum_{x \in \Omega_s} (I_x T_x^m)} \, e^{\frac{-a_m^2}{2\sigma_n^2} \sum_{x \in \Omega_s} (T_x^m)^2} \, P(\boldsymbol{T^m})$$

Re-formulation of the ideal observer based on pre-computed variables

The reformulation aims to avoid the computation of dot products ($I_x T_x^m$), so I aim to express them in terms of pre-computed variables. I start by describing the image formation. As discussed before, the gray level of a location x (I_x) is the sum of the target's gray level ($a\, T_x^m$) and the gray level of the Gaussian noise background (mean of 0 and standard deviation of σ_n). Assume that target i will be presented ($m = i$), and $\mathrm{I_x^b}$ is gray level of location x for the base Gaussian noise background with a mean of zero and standard deviation of one. We can describe the image formation process (I_x) as:

$$I_x = aT_x^i + G_x(0, \sigma_n)$$

$$I_x = aT_x^i + I_x^b \sigma_n$$

The dot product of any template ($\mathrm{I_x T_m^m}$) can be described in terms of the pre-computed dot product of target with base Gaussian noise background ($\mathrm{I_x^b T_x^m}$):

$$\sum_{x \in \Omega_s} T_m^m I_x = \sum_{x \in \Omega_s} T_x^m \; aT_x^i + \sum_{x \in \Omega_s} T_x^m \; \sigma_n I_x^b$$

If we reformulate the ideal observer with this result:

$$\frac{P(I \mid H_1)}{P(I \mid H_0)} = \sum_m e^{\frac{(a^2 \sum_{x \in \Omega_s} T_x^m T_x^i)}{\sigma_n^2} + \frac{(a \sum_{x \in \Omega_s} T_x^m I_x^b)}{\sigma_n} - \frac{(a^2 \sum_{x \in \Omega_s} (T_x^m)^2)}{\sigma_n^2}} \, P(\boldsymbol{T^m})$$

All three terms can be pre-computed. The cross-products of templates ($T^m T^i$) are pre-calculated alongside the energy of targets ($T^m T^m$) as part of the pre-computed cross-product matrix. For the second term, template responses to base noise images ($T^m I^b$) are pre-calculated. Therefore, it is easy to change the amplitude, background contrast (standard deviation), and priors without recalculating template responses. It is also possible to only run a subset of uncertainty conditions since we are in control of the image formation procedure and able to decide which sub-set of targets will be presented.

Re-formulation of the MAX observer based on pre-computed variables

The MAX observer decision rule is the following: respond target present if

$$max_m \left(\sum_{x \in \Omega_s} T_x^m \, I_x \right) > \delta$$

However, I will consider a family of MAX observers based on how dot products ($I_x T_x^m$) is normalized.

$$max_m (NR^m) > \delta$$

$$NR^m = \frac{\sum_{x \in \Omega_s} T_x^m I_x}{k}$$

$$k = \left(\sum_{x \in \Omega_s} | \; T_x^m \; |^p \right)^{1/p}$$

Normalization factors (k) are the L^p norms of target vectors (T^m). When p is infinity, taking L^p norm equals taking the maximum of the vector, and since all targets have the peak of one, the decision rule effectively becomes the first decision rule. If we express the decision rule in terms of pre-computed variables, it can be rewritten as:

$$max_m \frac{\left(\sum_{x \in \Omega_s} T_x^m \, a T_x^i + \sum_{x \in \Omega_s} T_x^m \; \sigma_n I_x^b \right)}{k} > \delta$$

APPENDIX 4: MORE COMPUTATIONAL RESULTS FOR SIMULTANEOUS TARGET ORIENTATION AND SCALE UNCERTAINTY

The MAX observer family with dynamic criteria for different scales

It is possible to compare the maximum template responses to a criterion based on the target scale (dynamic criterion) because the discriminability of different scales varies substantially. However, the discriminability of different orientations is the same, so there is no value in considering dynamic criteria based on orientation levels. I only simulated three L^p norms (1,2, and infinity) and selected criteria for each scale to optimize the percentage correct (Figure A4.1). For the simple MAX observer (L^{Inf}), the performance is improved but not nearly enough to catch up with other candidate model observers. For the energy-normalized MAX observer (L^2), the performance improvement almost completely covers its initial difference with the ideal observer. With the dynamical criterion, the energy-normalized model observer's performance is almost the same as the ideal observer. For the normalization norm of 1 (L^1), the performance almost stays the same, which reveals that it is not affected by having a single criterion in the first place, as discussed in the main text.

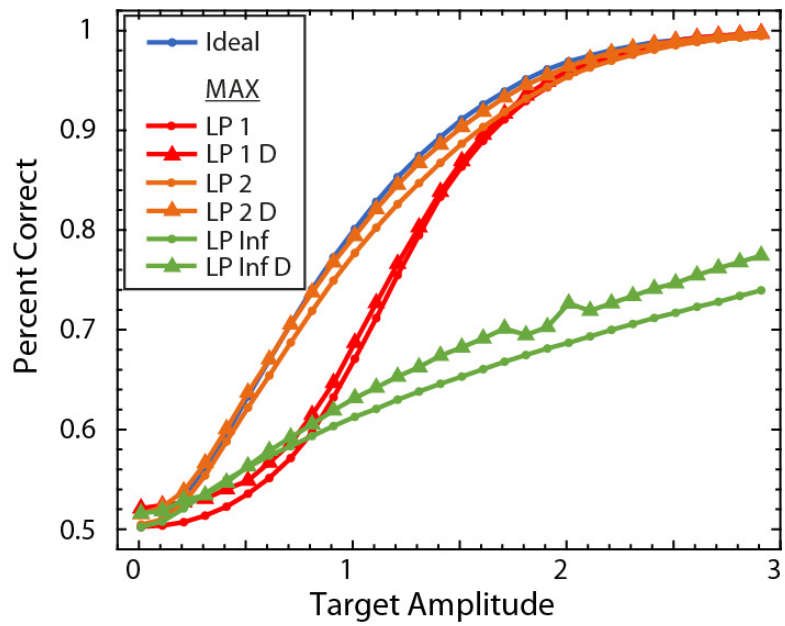


Figure A4.1 Comparison of model observers with dynamic criteria

The overall percentage correct as a function of target amplitude under target scale and orientation uncertainty for the ideal observer and family of MAX observer models. The normalization factor (L^p norm of the target) of each MAX observer is indicated with the number after LP in the legend. For model observers that use dynamic criterion, capital d is added at the end of the legend.

Only Scale Uncertainty

I have simulated model observers when there is only scale uncertainty and averaged the results across orientations since the pattern of results is expected to be the same for every orientation. The overall percentage correct as a function of target amplitude is shown in Figure A4.2A. The relationship between variations of MAX observer is similar (compared to combined uncertainty, no normalization, and normalizing with norms of 3,4 and 5 are closer to the ideal observer). However, the performance of normalizing with the L^1 norm has surpassed the performance of normalizing with L^2 norm in most of the high target amplitudes. The reason is that the need for dynamic criterion is the primary factor in performance in this case. As shown in Figure A4.2B, when the dynamic criterion for different scales is implemented for the L^2 norm, it closely approximates the ideal observer.

The L^1 norm does not benefit from the dynamic criterion considerably because it already has a smaller disadvantage due to having a single criterion.

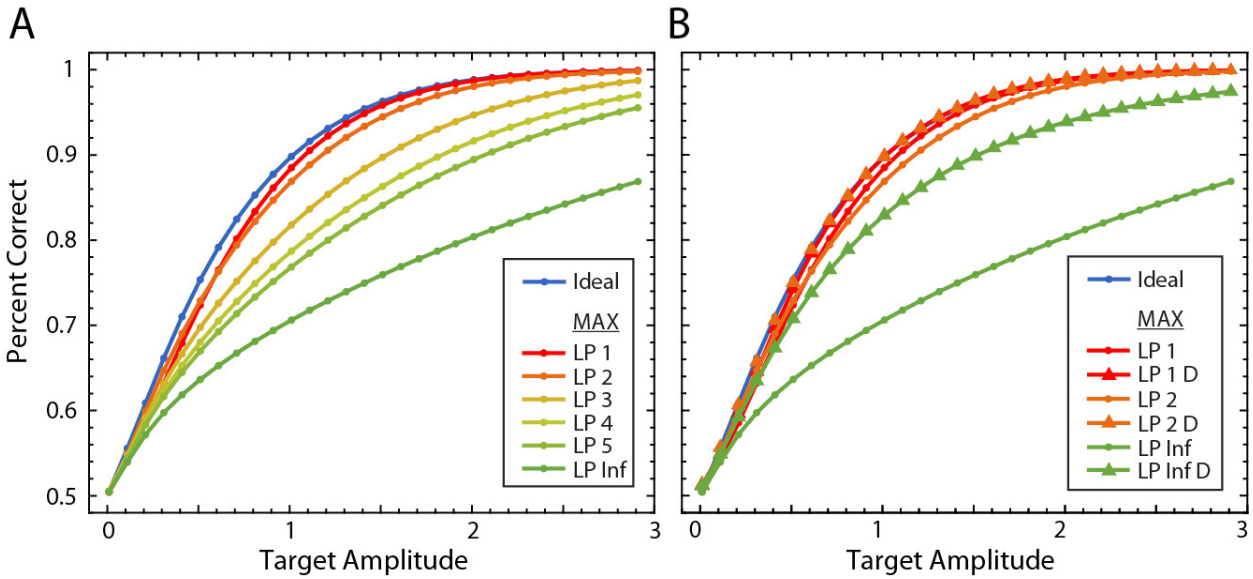


Figure A4.2 Performance under only target scale uncertainty.

The overall percentage correct as a function of target amplitude under target scale and orientation uncertainty for the ideal observer and family of MAX observer models. The normalization factor (L^p norm of the target) of each MAX observer is indicated with the number after LP in the legend. For model observers that use dynamic criterion, capital d is added at the end of the legend.

For the threshold level, the hits and correct rejection rates are shown in Figure A43. The pattern is different the compared to combined scale and orientation uncertainty. The hit rate of two MAX observers is similar, whereas they differ from the ideal observer's hit rates. The energy-normalized MAX observer and ideal observer almost have the same correct rejection rate, whereas simple MAX observers have a lower correct rejection rate. These differences might be used to differentiate between the model observers under only scale uncertainty. However, note that these are the result of optimal criterion placement, and model observers would be able to generate various patterns with non-optimal criteria.

It is common for human participants to place their criterion non-optimally, so the model observer's distinguishability depends on how patterns change with the criterion.

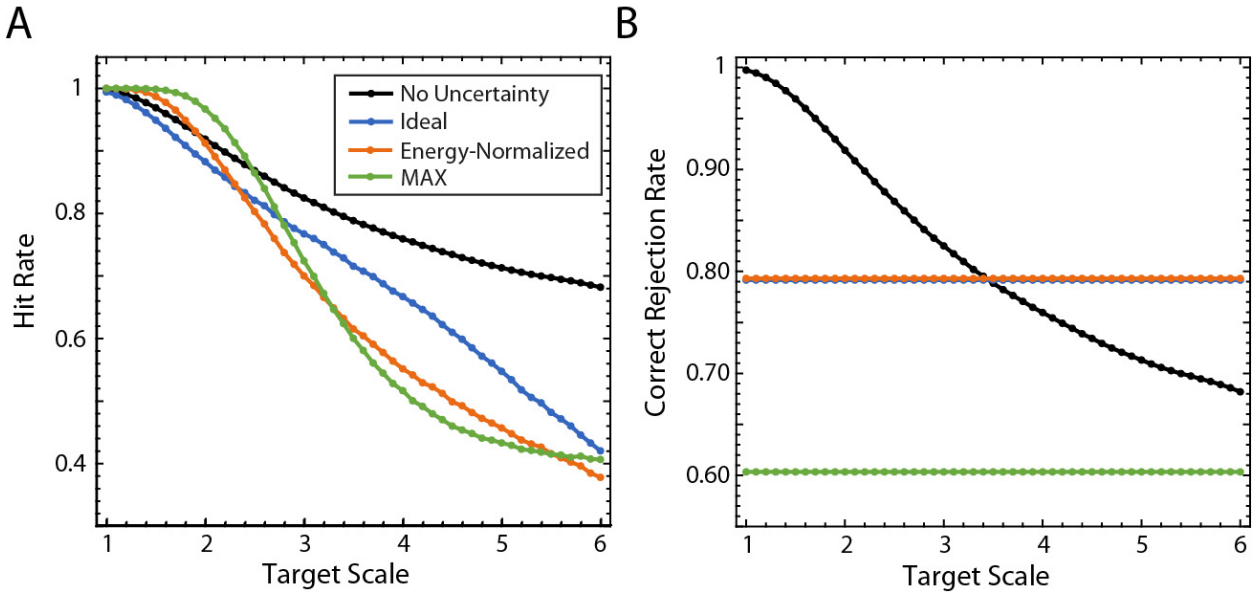


Figure A4.3 Hit and correct rejection rates under scale uncertainty.

A Hit rates as a function of the target scale of three observer models under scale and orientation uncertainty together with no uncertainty. The amplitude is picked such that the overall percentage correct is around 75 percent (threshold level). The expected hit rate patterns for different orientations are the same, so hit rates are averaged over orientation levels. When there is no uncertainty about the target scale and orientation, all models' predictions are the same. **B** Correct rejection rates as a function of target scale for three observer models under scale and orientation uncertainty together with no uncertainty. When there is a combined target scale and orientation uncertainty, target absent trials do not differ systematically, so there is only a single correct rejection rate for a single experiment.

References

Abbey, C. K., & Eckstein, M. P. (2006). Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer. *Journal of Vision*, 6(4), 4-4.

Ahumada, A. J., & Beard, B. L. (1997). Image discrimination models predict detection in fixed but not random noise. *JOSA A*, 14(9), 2471-2476.

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual neuroscience*, 7(6), 531-546.

Angela, J. Y., Giese, M. A., & Poggio, T. A. (2002). Biophysiologically plausible implementations of the maximum operation. *Neural computation*, 14(12), 2857-2881.

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2020). Local features and global shape information in object classification by deep convolutional neural networks. *Vision research*, 172, 46-61.

Beard, B. L., & Ahumada, A. J. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *JOSA A*, 16(3), 755-763.

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, 74(1), 30-39.

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision research*, 37(23), 3327-3338.

Benjamin, A. S., Tullis, J. G., & Lee, J. H. (2013). Criterion noise in ratings-based recognition: evidence from the effects of response scale length on recognition accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1601.

Bochud, F. O., Abbey, C. K., and Eckstein, M. P. (2004). Search for lesions in mammograms: statistical characterization of observer responses. *Medical Physics*, 31(1), 24-36.

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spatial Vision* 10:433-436.

Burge, J., & Geisler, W. S. (2011). Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences*, 108(40), 16849-16854.

Burge, J., & Jaini, P. (2017). Accuracy maximization analysis for sensory-perceptual tasks: Computational improvements, filter robustness, and coding advantages for scaled additive noise. *PLoS computational biology*, 13(2), e1005281.

Burge, J. (2020). Image-Computable Ideal Observers for Tasks with Natural Stimuli. *Annual Review of Vision Science*, 6, 491-517.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, 214(4516), 93-94.

Burgess, A., & Ghandeharian, H. (1984a). Visual signal detection. I. Ability to use phase information. *JOSA A*, 1(8), 900-905.

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

Burgess, A. (1985). Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. *JOSA A*, 2(9), 1498-1507.

Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264(5163), 1333-1336.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21), 8621-8644.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1), 51-62.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of neurophysiology*, 88(5), 2530-2546.

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision research*, 40(10-12), 1203-1215.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research*, 51(13), 1484-1525.

Castella, C., Kinkel, K., Eckstein, M. P., Abbey, C. K., Verdun, F. R., Saunders, R. S., ... & Bochud, F. O. (2008). Mass detection on mammograms: signal variations and performance changes for human and model observers. *In Medical Imaging 2008: Image Perception, Observer Performance, and Technology Assessment* (Vol. 6917, p. 69170K). International Society for Optics and Photonics.

Castella, C., Eckstein, M. P., Abbey, C. K., Kinkel, K., Verdun, F. R., Saunders, R. S., ... & Bochud, F. O. (2009). Mass detection on mammograms: influence of signal shape uncertainty on human and model observers. *JOSA A*, 26(2), 425-436.

Cohn, T. E., and Lasley, D. J. (1974). Detectability of a luminance increment: Effect of spatial uncertainty. *JOSA A*, 64(12), 1715-1719.

Cohn, T. E., and Wardlaw, J. C. (1985). Effect of large spatial uncertainty on foveal luminance increment detectability. *JOSA A*, 2(6), 820-825.

Cohn, T. E., & Lasley, D. J. (1986). Visual sensitivity. *Annual review of psychology*, 37(1), 495-521.

Chen, Y., Bedell, H. E., Frishman, L. J., & Levi, D. M. (1998). Stimulus uncertainty affects velocity discrimination. *Vision research*, 38(9), 1265-IN2.

David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *Journal of Neuroscience*, 24(31), 6991-7006.

Davis, E. T., & Graham, N. (1981). Spatial frequency uncertainty effects in the detection of sinusoidal gratings. *Vision research*, 21(5), 705-712.

Davis, E. T., Kramer, P., and Graham, N. (1983). Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception and Psychophysics*, 33(1), 20-28.

Doehrman, S. (1974). The effect of visual orientation uncertainty in a simultaneous detection-recognition task. *Perception & Psychophysics*, 15(3), 519-523.

Earle, D. C., & Lowe, G. (1971). Channel, temporal, and composite uncertainty in the detection and recognition of auditory and visual signals. *Perception & Psychophysics*, 9(2), 177-181.

Eckstein, M. P., and Whiting, J. S. (1996). Visual signal detection in structured backgrounds I. Effect of number of possible spatial locations and signal contrast. *JOSA A*, 13(9), 1777-1787.

Eckstein, M. P., Whiting, J. S., & Thomas, J. P. (1996). Role of knowledge in human visual temporal integration in spatiotemporal noise. *JOSA A*, 13(10), 1960-1968.

Eckstein, M. P., Ahumada, A. J., & Watson, A. B. (1997). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. *JOSA A*, 14(9), 2406-2419.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). 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.

Eckstein, M. P., & Abbey, C. K. (2001). Model observers for signal-known-statistically tasks (SKS). *In Medical Imaging 2001: Image Perception and Performance* (Vol. 4324, pp. 91-102). International Society for Optics and Photonics.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of vision*, 11(5), 14-14.

Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS biology*, 3(10), e342.

Foley, J. M., & Schwarz, W. (1998). Spatial attention: Effect of position uncertainty and number of distractor patterns on the threshold-versus-contrast function for contrast discrimination. *JOSA A*, 15(5), 1036-1047.

Foley, J. M., Varadharajan, S., Koh, C. C., & Farias, M. C. (2007). Detection of Gabor patterns of different sizes, shapes, phases and eccentricities. *Vision Research*, 47(1), 85-107.

Frazor, R. A., & Geisler, W. S. (2006). Local luminance and contrast in natural images. *Vision research*, 46(10), 1585-1598.

Gould, I. C., Wolfgang, B. J., & Smith, P. L. (2007). Spatial uncertainty explains exogenous and endogenous attentional cuing effects in visual signal detection. *Journal of Vision*, 7(13), 4-4.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (Vol. 1, pp. 1969-12). *New York: Wiley.*

Greenhouse, D. S., & Cohn, T. E. (1978). Effect of chromatic uncertainty on detectability of a visual stimulus. *JOSA A*, 68(2), 266-267.

Greenwald, H. S., & Knill, D. C. (2009). A comparison of visuomotor cue integration strategies for object placement and prehension. *Visual neuroscience*, 26(1), 63-72.

Geisler, W. S., & Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London*. Series B: Biological Sciences, 357(1420), 419-448.

Geisler, W. S. (2003). Ideal observer analysis. *The visual neurosciences*, 10(7), 12-12.

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, 27(3), 379-402.

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

Geisler, W. S., & Ringach, D. (2009). Natural systems analysis. *Visual neuroscience*, 26(1), 1-3.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision research*, 51(7), 771-781.

Geisler, W. S. (2018). Psychometric functions of uncertain template matching observers. *Journal of vision*, 18(2), 1-1.

Georgeson, M. A., Yates, T. A., & Schofield, A. J. (2008). Discriminating depth in corrugated stereo surfaces: Facilitation by a pedestal is explained by removal of uncertainty. *Vision research*, 48(21), 2321-2328.

Graham, N., Kramer, P., & Yager, D. (1987). Signal-detection models for multidimensional stimuli: Probability distributions and combination rules. *Journal of Mathematical Psychology*, 31(4), 366-409.

Guyonneau, R., Kirchner, H., & Thorpe, S. J. (2006). Animals roll around the clock: The rotation invariance of ultrarapid visual processing. *Journal of Vision*, 6(10), 1-1.

Han, M., & Baek, J. (2020). A convolutional neural network-based anthropomorphic model observer for signal-known-statistically and background-known-statistically detection tasks. *Physics in Medicine & Biology*, 65(22), 225025.

Han, Y., Roig, G., Geiger, G., & Poggio, T. (2020). Scale and translation-invariance for novel objects in human vision. *Scientific reports*, 10(1), 1-13.

Hayhoe, M., Gillam, B., Chajka, K., & Vecellio, E. (2009). The role of binocular vision in walking. *Visual neuroscience*, 26(1), 73-80.

Hayhoe, M. M. (2017). Vision and action. *Annual review of vision science*, 3, 389-413.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual neuroscience*, 9(2), 181-197.

Hess, R. F., & Hayes, A. (1994). The coding of spatial position by the human visual system: Effects of spatial scale and retinal eccentricity. *Vision Research*, 34(5), 625-643.

Howarth, C. I., & Lowe, G. (1966). Statistical detection theory of Piper's law. *Nature*, 212(5059), 324-326.

Hübner, R. (1996a). Specific effects of spatial-frequency uncertainty and different cue types on contrast detection: Data and models. *Vision Research*, 36(21), 3429-3439.

Hübner, R. (1996b). The efficiency of different cue types for reducing spatial-frequency uncertainty. *Vision Research*, 36(3), 401-408.

Judy, P. F., Kijewski, M. F., Fu, X., & Swensson, R. G. (1995). Observer detection efficiency with target size uncertainty. *In Medical Imaging 1995: Image Perception* (Vol. 2436, pp. 10-17). International Society for Optics and Photonics.

Judy, P. F., Kijewski, M. F., & Swensson, R. G. (1997). Observer detection performance loss: target-size uncertainty. *In Medical Imaging 1997: Image Perception* (Vol. 3036, pp. 39-47). International Society for Optics and Photonics.

Kramer, P., Graham, N., & Yager, D. (1985). Simultaneous measurement of spatial-frequency summation and uncertainty effects. *JOSA A*, 2(9), 1533-1542.

Lampl, I., Ferster, D., Poggio, T., & Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *Journal of neurophysiology*, 92(5), 2704-2713.

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual neuroscience*, 26(1), 51-62.

Lasley, D. J., & Cohn, T. (1981a). Detection of a luminance increment: effect of temporal uncertainty. *JOSA A*, 71(7), 845-850.

Lasley, D. J., & Cohn, T. E. (1981b). Why luminance discrimination may be better than detection. *Vision research*, 21(2), 273-278.

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature neuroscience*, 2(4), 375-381.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual review of neuroscience*, 19(1), 577-621.

Lowe, G. (1967). Interval of time uncertainty in visual detection. *Perception & Psychophysics*, 2(7), 278-280.

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision research*, 38(9), 1183-1198.

Luck, S. J., Hillyard, S. A., Mouloua, M., & Hawkins, H. L. (1996). Mechanisms of visual–spatial attention: Resource allocation or uncertainty reduction?. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3), 725.

Ma, W. J., Navalpakkam, V., Beck, J. M., Berg, R. V. D., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature neuroscience*, 14(6), 783-790.

Ma, W. J., Shen, S., Dziugaite, G., & van den Berg, R. (2015). Requiem for the max rule?. *Vision Research*, 116, 179-193.

Manjeshwar, R. M., & Wilson, D. L. (2001). Effect of inherent location uncertainty on detection of stationary targets in noisy image sequences. *JOSA A*, 18(1), 78-85.

Martens, W., & Blake, R. (1980). Uncertainty impairs grating detection performance in the cat. *Perception & Psychophysics*, 27(3), 229-231.

Martinez-Garcia, M., Bertalmío, M., & Malo, J. (2019). In praise of artifice reloaded: Caution with natural image databases in modeling vision. *Frontiers in neuroscience*, 13, 8.

Meese, T. S., Hess, R. F., & Williams, C. B. (2005). Size matters, but not for everyone: Individual differences for contrast discrimination. *Journal of Vision*, 5(11), 2-2.

Meese, T. S., & Summers, R. J. (2012). Theory and data for area summation of contrast with and without uncertainty: Evidence for a noisy energy model. *Journal of vision*, 12(11), 9-9.

Mertens, J. J. (1956). Influence of knowledge of target location upon the probability of observation of peripherally observable test flashes. *JOSA A*, 46(12), 1069-1070.

Michel, M., & Geisler, W. S. (2011). Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision*, 11(1), 18-18.

Nachmias, J., & Kocher, E. C. (1970). Visual detection and discrimination of luminance increments. *JOSA A*, 60(3), 382-389.

Nachmias, J. (2002). Contrast discrimination with and without spatial uncertainty. *Vision Research*, 42(1), 41-48.

Neri, P. (2010). Visual detection under uncertainty operates via an early static, not late dynamic, non-linearity. *Frontiers in computational neuroscience*, 4, 151.

Nolte, L. W., & Jaarsma, D. (1967). More on the detection of one of M orthogonal signals. *The Journal of the Acoustical Society of America*, 41(2), 497-505.

Ohtani, Y., & Mori, S. (2002). Spatial-frequency uncertainty and cuing effects on psychometric functions for contrast detection. *Psychonomic bulletin & review*, 9(2), 270-277.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), 607-609.

Oruc, I., & Barton, J. J. (2010). Critical frequencies in the perception of letters, faces, and novel shapes: Evidence for limited scale invariance for faces. *Journal of Vision*, 10(12), 20-20.

Park, S., Clarkson, E., Kupinski, M. A., & Barrett, H. H. (2005). Efficiency of the human observer detecting random signals in random backgrounds. *JOSA A*, 22(1), 3-16.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *JOSA A*, 2(9), 1508-1532.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision* 10:437-442.

Peterson, W. W. T. G., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE professional group on information theory*, 4(4), 171-212.

Reith, F. H., & Wandell, B. A. (2020). A convolutional neural network reaches optimal sensitivity for detecting some, but not all, patterns. *IEEE Access*, 8, 213522-213530.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11), 1019-1025.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature neuroscience*, 8(12), 1647-1650.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature neuroscience*, 4(8), 819-825.

Sebastian, S., Abrams, J., & Geisler, W. S. (2017). Constrained sampling experiments reveal principles of detection in natural scenes. *Proceedings of the National Academy of Sciences*, 114(28), E5731-E5740.

Sebastian, S., & Geisler, W. S. (2018). Decision-variable correlation. *Journal of Vision*, 18(4), 3-3.

Sekuler, R., & Ball, K. (1977). Mental set alters visibility of moving targets. *Science*, 198(4312), 60-62.

Semizer, Y., & Michel, M. M. (2017). Intrinsic position uncertainty impairs overt search performance. *Journal of vision*, 17(9), 13-13.

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., & Poggio, T. (2005). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. *MASSACHUSETTS INST OF TECH CAMBRIDGE MA CENTER FOR BIOLOGICAL AND COMPUTATIONAL LEARNING.*

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. I*EEE transactions on pattern analysis and machine intelligence*, 29(3), 411-426.

Solomon, J. A., Lavie, N., & Morgan, M. J. (1997). Contrast discrimination function: Spatial cuing effects. *JOSA A*, 14(9), 2443-2448.

Stüttgen, M. C., & Schwarz, C. (2008). Psychophysical and neurometric detection performance under stimulus uncertainty. *Nature neuroscience*, 11(9), 1091-1099.

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

Swets, J. A. (1959). Indices of signal detectability obtained with various psychophysical procedures. *The Journal of the Acoustical Society of America*, 31(4), 511-513.

Swets, J. A., Tanner Jr, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological review*, 68(5), 301.

Tanner Jr, W. P. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*.

Tekin, E., & Roediger, H. L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cognitive Research: Principles and Implications*, 2(1), 1-13.

Tjan, B. S., & Nandy, A. S. (2006). Classification images with uncertainty. *Journal of Vision*, 6(4), 8-8.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, 23(8), 775-785.

Touryan, J., Felsen, G., & Dan, Y. (2005). Spatial structure of complex cell receptive fields measured with natural images. *Neuron*, 45(5), 781-791.

Ukkonen, O., Rovamo, J., and Näsänen, R. I. S. T. O. (1995). Effect of location and orientation uncertainty on rms contrast sensitivity with and without spatial noise in

peripheral and foveal vision. *Optometry and vision science: official publication of the American Academy of Optometry*, 72(6), 387-395.

Walshe, R. C., & Geisler, W. S. (2022). Efficient allocation of attentional sensitivity gain in visual cortex reduces foveal sensitivity in visual search. *Current Biology*, 32(1), 26-36.

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *JOSA A*, 14(9), 2379-2391.

Westheimer, G., & Ley, E. (1996). Temporal uncertainty effects on orientation discrimination and stereoscopic thresholds. *JOSA A*, 13(4), 884-886.

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, 5(1), 102-122.

Wilson, D. L., & Manjeshwar, R. (1999). Role of phase information and eye pursuit in the detection of moving objects in noise. *JOSA A*, 16(3), 669-678.

Wolfe, J. M. (2015). Visual search. *Current Biology*, 20(8), R346.

Yager, D., Kramer, P., Shaw, M., & Graham, N. (1984). Detection and identification of spatial frequency: models and data. *Vision Research*, 24(9), 1021-1035.