

Encoding-decoding models of luminance contrast processing

by

Stuart Jackson

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
CENTER FOR NEURAL SCIENCE
NEW YORK UNIVERSITY

JANUARY, 2016

Abstract

There is a long tradition in sensory neuroscience of fitting precise neural computational models to experimental data. In a series of combined empirical and computational investigations, I illustrate important constraints on the encoding, retention, and read-out of information relating to luminance contrast in the visual world, a fundamental building block of vision. Using two-interval, forced-choice discrimination tasks, I first demonstrate that the efficiency of luminance contrast encoding-decoding is greatly impeded when high-contrast distractors appear in the opposite visual hemifield to a target stimulus; this behavior contrasts with relatively more efficient performance observed on an orthogonal task (orientation discrimination). I then explore a neural computational model of these results based on Fisher Information, and find that, given a particular tuning parameterization, neither of two common models of sensory interaction satisfactorily explain both datasets simultaneously. In a later delayed-estimation experiment, I directly measure the precision with which single estimates of luminance contrast are encoded, maintained, and read-out from memory. The shape of observers' estimate distributions are adequately replicated by a probabilistic model of performance based on neurally-inspired components. In sum, the present thesis highlights key factors governing the precision of luminance contrast encoding and decoding, using complementary empirical and computational approaches. The thesis findings are also relevant to the broader literatures on attentional selection and the short-term retention of sensory information.

Contents

ABSTRACT	I
1 INTRODUCTION	5
1.1 Overview	5
1.2 Behavioral measures of stimulus encoding and decoding	7
1.3 Neural basis of stimulus encoding and decoding	10
1.4 Encoding-decoding performance under natural conditions	15
1.5 Thesis synopsis	20
2 EFFICIENT AND INEFFICIENT SELECTION FROM THE SAME SENSORY NEURAL RE- SPONSE: PSYCHOPHYSICS	22
2.1 Introduction	23
2.2 Materials and Methods	26
2.3 Results	38
2.4 Conclusion	43
3 EFFICIENT AND INEFFICIENT SELECTION FROM THE SAME SENSORY NEURAL RE- SPONSE: COMPUTATIONAL MODEL	46
3.1 Introduction	47
3.2 Materials and Methods	50
3.3 Results	61
3.4 Discussion	68
4 DELAYED ESTIMATION OF LUMINANCE CONTRAST	72
4.1 Introduction	72
4.2 Materials and Methods	75
4.3 Results	83
4.4 Discussion	88

5	CONCLUSION	91
5.1	Encoding-decoding models of luminance contrast processing	91
5.2	Implications for research on attentional selection	92
5.3	Implications for the study of VSTM	94
5.4	Neural noise and encoding-decoding	96
5.5	Final comments	97
	REFERENCES	106

1

Introduction

1.1 OVERVIEW

Progress in sensory neuroscience can greatly benefit from a tight interplay between experiment and model-fitting. While some experimental work is exploratory or descriptive in nature, there is a long tradition of experiments explicitly designed in order to quantitatively distinguish theories. Successful quantitative model-fitting to data allows for stronger con-

clusions than conventional, model-free analyses, while deviations from model predictions or otherwise unexpected results provide strong clues as to how a theory must be modified.

The present thesis focuses on understanding the brain's ability to represent, maintain, and read out information about the luminance contrast of stimuli in the world. As the fundamental building block for all of vision, information about light intensity passes through a variety of important neural processing stages: from retinal responses that signal individual spots of light, through sub-cortical and cortical processing stages, where a transformation towards representing mean luminance contrast in localized parts of the visual image is completed. Luminance contrast is arguably the most basic visual feature for pattern vision, and from it most other behaviorally-useful visual representations are derived (e.g., feature orientations, object boundaries, etc.). In this thesis, we present two sets of behavioral experiments with human observers focused on luminance contrast processing, as well as accompanying models based on contemporary theories of neural encoding and decoding.

Below, we first briefly review standard empirical approaches to understanding the nature of encoding and decoding for single visual stimuli, with a focus on luminance contrast processing. We then review key anatomical and functional characteristics of the visual system, focusing on how the visual system represents and transforms raw sensory input into the meaningful building blocks of vision, such as the amount of local luminance contrast or dominating orientation in small patches of the visual image. We then extend the discussion to more naturalistic conditions: emphasis is placed on the constraints governing behaviors that require encoding and retention of multiple stimuli over short intervals i.e., so called visual short-term memory (VSTM). The precision of encoding-decoding performance with topographically-structured neural representations (e.g., orientation-tuned neural responses)

is contrasted with our current understanding of VSTM for luminance contrast. This debate, and the development of encoding-decoding models of luminance contrast processing, forms the backbone of the thesis. On a more general note, we touch on the links between research on sensory encoding-decoding, VSTM, and the broader literature on attentional selection of sensory information. We conclude the introduction with a brief thesis synopsis.

1.2 BEHAVIORAL MEASURES OF STIMULUS ENCODING AND DECODING

Arguably the guiding pillars of modern computational approaches to understanding brain function are the inter-twined problems of neural encoding and decoding (Dayan & Abbott, 2001; Pouget et al., 2003). To maintain successful and productive behavioral repertoires, animals must adequately encode and make use of various sources of degraded information about the world. The development of realistic models of such behavior requires that researchers first characterize the everyday limits on encoding and decoding performance, using simplified experimental methods.

1.2.1 DEFINITION OF STIMULUS CONTRAST

We begin by reviewing the standard behavioral paradigms used to study the processing of luminance contrast and other basic visual features. Stimuli in such tasks are commonly simplified spatial patterns such as sinusoidal gratings or circular discs, which allow for substantial experimental flexibility. Stimulus luminance contrast is typically given as a percentage of maximum contrast, and can be defined according to more than one convention, depending on the particular stimulus set-up and the experimenter's choice. For present purposes,

when we refer to stimulus contrast, we define it according to Michelson contrast,

$$c_{\text{michelson}} = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}} \quad (1.1)$$

where L_{max} and L_{min} represent the maximum and minimum luminances of the stimulus respectively (e.g., the peak and trough of a sinusoidal grating). In Chapter 4, we make reference to the an alternative definition (Weber contrast), which is described there.

1.2.2 DETECTION AND DISCRIMINATION OF LUMINANCE CONTRAST

Experimentalists have traditionally investigated the encoding-decoding of luminance contrast and other basic visual features using behavioral paradigms such as discrimination or detection (Blake & Holopigian, 1985; Boynton et al., 1999; Legge & Foley, 1980; Skottun et al., 1987). In a two-alternative, forced-choice discrimination task (2-AFC), for example, an observer is presented with a baseline or pedestal stimulus value across each of two intervals, and must correctly distinguish in which of the two intervals an additional increment (i.e., a contrast change) is added to the pedestal. Threshold performance is typically defined as the increment magnitude necessary to achieve some fixed performance criterion (e.g., 75% correct). A contrast-detection task is a limited form of discrimination, in which the baseline contrast is set to background luminance (i.e., 0% contrast), and absolute detection thresholds have commonly been collected alongside discrimination thresholds in the same experimental runs (Bradley & Ohzawa, 1986; Nachmias & Sansbury, 1974).

A classic finding in contrast-discrimination tasks is the improvement in discrimination performance for very low contrast pedestals relative to detection performance (Bradley & Ohzawa, 1986; Legge & Foley, 1980; Nachmias & Sansbury, 1974) (Figure 1.1A). This ef-

fect is thought to arise from an early accelerating non-linearity in contrast encoding, under the assumption that regardless of pedestal contrast level, a fixed change in some internal response (i.e., neural firing) is required for discrimination of changes to the pedestal (Boyn-ton et al., 1999; Legge & Foley, 1980; Nachmias & Sansbury, 1974). Following this early facilitation effect, a gradual increase in thresholds for much of the pedestal contrast axis is typically observed. The slope of this increase is often found to be around 0.5-0.7 on a log-log axis (the so-called near-miss to Weber's Law) (Legge & Foley, 1980; Nachmias & Sansbury, 1974). Presumably, the gradual increase in threshold with increasing pedestal contrast reflects some trade off between the shape of the internal response to stimuli and internal noise levels, a debate which has continued for some time (Gorea & Sagi, 2001). In passing, however, we note that a number of studies which have measured thresholds for very high-contrast pedestals (e.g., above 50% contrast) have found some late flattening or decrease in threshold, thereby suggesting that the later part of the contrast-discrimination function is not necessarily monotonic throughout (Chirimuuta & Tolhurst, 2005; Kingdom & Whittle, 1996; Pestilli et al., 2011; Zenger-Landolt & Heeger, 2003).

1.2.3 OTHER PARADIGMS AND STIMULUS FEATURES

A smaller number of studies have also utilized a matching or adjustment protocol to study luminance contrast processing (Georgeson & Sullivan, 1975; Prinzmetal et al., 1997). For example, Georgeson & Sullivan (1975) had observers adjust the contrast of one of two sinusoidal gratings presented side-by-side, so as to match the other grating in contrast: for relatively broad differences in stimulus spatial frequency, observers could accurately match the contrasts of the variable and standard stimuli, suggesting a substantial degree of adapt-

ability on the part of the local neural mechanisms feeding into stimulus contrast coding.

The effects of luminance contrast on the encoding of another important visual feature, stimulus orientation, have also been studied using discrimination and matching paradigms (Blake & Holopigian, 1985; Mareschal & Shapley, 2004; Prinzmetal et al., 1998; Skottun et al., 1987). As pedestal contrast increases, a characteristic decrease in thresholds for orientation discrimination is found, with typically no substantial improvement in performance beyond pedestals of about 10-20% contrast (Figure 1.1B). In other words, observers appear to reach ceiling performance levels for orientation discrimination at relatively low-to-moderate contrasts, with the bound on performance likely set by fixed levels of internal noise (Mareschal & Shapley, 2004). Results commensurate with this notion of a fixed bound on performance have also been found in matching experiments, where the orientation of a comparison stimulus must be adjusted to match the orientation of a recently presented test stimulus. For example, Prinzmetal et al. (1998) found that observers matching estimates had a typical standard deviation of about 5° to 10° around the test stimulus orientation, analogous to discrimination thresholds under certain stimulation regimes. The specific role of delay time between test and match stimuli in such paradigms will be addressed more directly later.

1.3 NEURAL BASIS OF STIMULUS ENCODING AND DECODING

To better understand the foundation on which behaviors such as contrast discrimination are based, we now briefly describe the architecture and function of the early visual system, and the neural representations that act as the building blocks of cortical vision. We also briefly discuss the known characteristics and effects of neural noise on sensory processing.

1.3.1 ARCHITECTURE AND FUNCTION OF THE EARLY VISUAL SYSTEM

Anatomical and physiological studies in numerous species illustrate common neural processing architectures for visual information (McIlwain, 1996). The primate visual system, for example, is known to progress along two main processing streams which differ in their anatomical and physiological characteristics, the magnocellular and parvocellular pathways (McIlwain, 1996; Shapley, 1990). This pathway segregation begins in the retina, and becomes highly evident in the layering of the major sub-cortical visual relay, the lateral geniculate nucleus (LGN) of the thalamus. The two pathways then converge onto different sub-layers of the primary visual cortex or V_1 , with magnocellular cells terminating in layer $4C\alpha$ and with a large portion of the parvocellular pathway cells terminating in layer $4C\beta$ (McIlwain, 1996; Sincich & Horton, 2005). From here, the organization of the separate processing streams becomes more nuanced in layout, with the parallel processing that predominated at earlier synapses (e.g., retina to thalamus), giving way to cross-talk in cortex. For example, V_1 neurons receiving thalamic inputs may have subsequent synapses onto other layers of V_1 , and individual cortical layers can be reciprocally connected to one another (Callaway, 2003; Sincich & Horton, 2005). Recent advances in neural tracing techniques are aiding in the development of highly detailed wiring diagrams of these circuits in numerous species (Callaway, 2003).

From a functional point of view, broad differences are apparent across these and other parallel processing channels in the early visual system. The transmission of chromatic (i.e., color-related) information from retina to cortex is subserved primarily by parvocellular pathway processing, while the magnocellular pathway is thought to play a more dominant role in processing achromatic, luminance-defined signals and motion (Johnson et al., 2001;

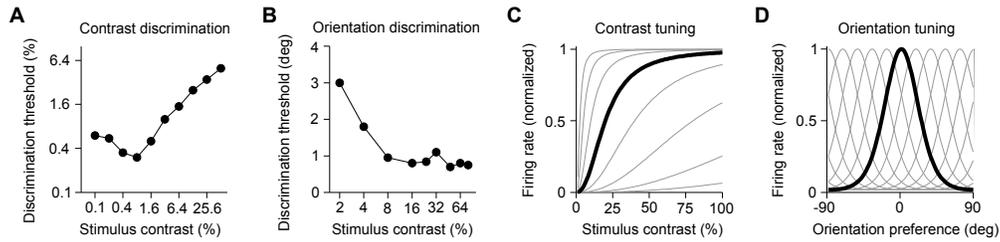


Figure 1.1: Typical discrimination behavior and idealized neural tuning functions for contrast and orientation. A) Typical contrast-discrimination performance for sinusoidal gratings of increasing contrast (adapted by eye from [Legge & Foley \(1980\)](#)). B) Typical orientation-discrimination performance for grating stimuli of increasing contrast. Adapted by eye from [Skottun et al. \(1987\)](#). C) Idealized single-neuron contrast response functions with logarithmically-spaced semi-saturation constant (Naka-Rushton function [Naka & Rushton \(1966\)](#)). D) Idealized (Von Mises) orientation-tuning curves tiling the orientation axis.

[Lennie et al., 1990; Shapley, 1990](#)). In fact, cells in the respective pathways illustrate broadly divergent sensitivity to luminance intensity and contrast, presumably playing qualitatively distinct roles in luminance contrast processing ([Kaplan & Shapley, 1986; Shapley, 1990](#)). In parallel, from the retinal bipolar layer through LGN, sensory responses are carried by cells that become either more or less responsive when stimulated within the center of their receptive field by light increments, the so-called ON/OFF channels in vision ([McIlwain, 1996; Schiller et al., 1986; Wiesel & Hubel, 1966](#)). Presumably, this early separation of responses to increments and decrements in light evolved to allow for maximum system contrast sensitivity at minimal biophysical cost ([Schiller et al., 1986](#)). Of note, numerous findings in recent years have now definitively illustrated asymmetries in the early representation of positive and negative luminance signals ([Chubb et al., 2004; Kremkow et al., 2014; Ratliff et al., 2010; Yeh et al., 2009](#)), a topic we will briefly touch on in the discussion sections of Chapters 2 and 4.

1.3.2 THE BUILDING BLOCKS OF CORTICAL VISION

A series of transformations occur in early vision from representing local, absolute light intensity values in neural responses, to representing luminance contrast in the scene being viewed. The visual system achieves this through a cascade of gain control processes, where the input drive at a given stage of processing (e.g., V_I) is scaled by a broader, suppressive (i.e., inhibitory) signal in the surrounding neural circuit (Carandini & Heeger, 2012; Ohzawa et al., 1985; Shapley & Victor, 1979; Wilson, 1999). Gain control processes serve to optimize the visual system's response under varying input conditions (e.g., for a given time of day or mean luminance level), by ensuring that the system is most sensitive to fluctuations around the mean input level. In general, increases in stimulating contrast typically produce a monotonically increasing contrast-response profile (Figure 1.1C), with lower contrasts producing relatively fewer mean spikes, and higher contrasts eventually saturating the response of the neuron. This firing rate behavior can be captured by relatively simple mathematical expressions, such as the commonly used Naka-Rushton equation (Naka & Rushton, 1966; Albrecht & Hamilton, 1982). By describing the responses of individual neurons using simplified equations with minimal parameters, properties of larger groups of neurons (i.e., local populations) can be summarized in terms of distributions of parameter values (Albrecht & Hamilton, 1982). As stimulus contrast is encoded in a neuron by the strength of its firing, we refer to luminance contrast throughout this thesis as being an intensity-coded feature.

Unlike the intensity-coding used for luminance contrast processing, the detailed architecture of the early visual system provides the ideal substrate for the structured or topographic representation of other basic visual features. In now classic experimental work, Hubel &

Wiesel (1962) demonstrated topographic representation along several visual stimulus feature dimensions e.g., location (retinotopy), orientation (orientation pinwheels), disparity processing (ocular dominance columns). In the intervening decades, neuroscientists have attempted to further refine our understanding of these topographic representational maps, perhaps most successfully for orientation-tuned neural responses (Ferster, 2003). In contrast to the monotonic tuning functions found for contrast, local orientations in an image are mapped onto approximately symmetric neural tuning functions for orientation, with individual neurons typically responding to a limited range of preferred stimulus orientations. An idealized example of symmetric, orientation tuning functions is depicted in Figure 1.1D.

1.3.3 NEURAL NOISE AND THE ENCODING-DECODING PROCESS

Even with a detailed supporting neural architecture, performance in behavioral tasks is never perfect: noise accrues in an observer's representation of the visual world, from the initial photon transduction process in retina, through to noisy neural spiking at sub-cortical and cortical layers of the visual system. Numerous investigators have studied characteristics of neural noise in visual cortex (Tolhurst et al., 1981; Shadlen & Newsome, 1998; Goris et al., 2014). A key finding relates to the link between the mean and variance of a neuron's firing rate: these have often been found to scale together with a ratio of approximately 1, meaning that a neuron's spiking behavior can be reasonably well-approximated as a Poisson process (Tolhurst et al., 1981; Shadlen & Newsome, 1998). However, this relationship is certainly not exact for real neurons, and the ratio of variance-to-mean (i.e., the Fano factor) has been found to range above and below 1 (Shadlen & Newsome, 1998; Goris et al., 2014).

In fact, an appreciation has grown recently for the importance of accounting for trial-to-trial noise fluctuations, and the effects of these fluctuations on stimulus encoding-decoding (Ecker et al., 2015; Goris et al., 2014). For example, fluctuations in the strength of neural response for a given stimulus can presumably arise due to non-sensory sources (e.g., modulations of attention or arousal), thereby acting as a secondary source of noise. Models of encoding-decoding performance have begun to incorporate more realistic gain fluctuations into their computational architecture (May & Solomon, 2015). We will attempt to implement a model of this sort in Chapter 4.

1.4 ENCODING-DECODING PERFORMANCE UNDER NATURAL CONDITIONS

The distinction between research on basic stimulus encoding-decoding, VSTM, and attentional selection are sometimes subtle. While VSTM paradigms inevitably incorporate delay between the time of encoding and read-out, this distinction is not typically considered in standard 2-AFC perception or attention-related discrimination tasks, where stimuli must still be held in memory from one interval to the next. To better link these topics, we first describe how temporal delay affects the classic discrimination behaviors described earlier. We then flesh out our current understanding of how encoding-decoding performance changes under more naturalistic conditions, such as in the presence of target location uncertainty (i.e., for multiple stimulus displays). This discussion is divided into two parts: research that has investigated the effects of varying set-size on performance, and research that has investigated the effects of stimulus interactions on performance. In doing so, we introduce a variety of research on VSTM, with particular emphasis on the delayed-estimation paradigm. We also touch on the broader topic of attentional selection of sensory informa-

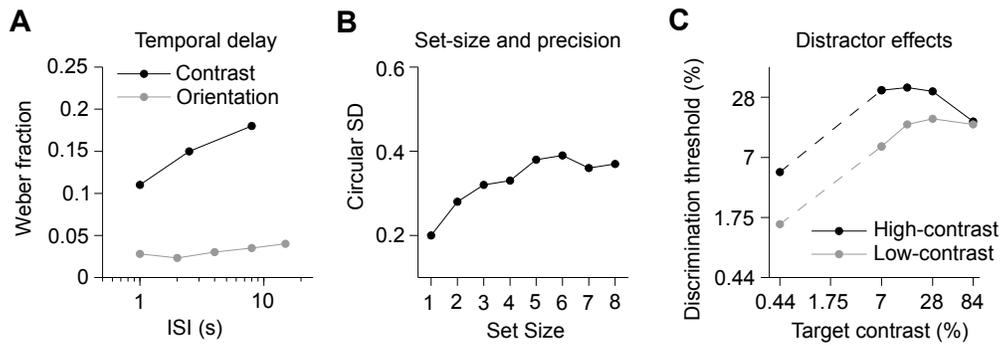


Figure 1.2: Encoding-decoding performance under more natural conditions. A) Effects of temporal delay (ISI duration) on contrast- and orientation-discrimination thresholds (adapted by eye from Magnussen & Greenlee (1999)). B) Effects of set-size on delayed-estimation for orientation. Adapted by eye from van den Berg et al. (2012). C) Effects of stimulus interactions (i.e., distractor contrast) on contrast-discrimination thresholds. Adapted by eye from Pestilli et al. (2011).

tion.

1.4.1 ROLE OF TEMPORAL DELAY ON ENCODING-DECODING PERFORMANCE

The similarities and differences between perception and VSTM are immediately apparent when one compares VSTM performance for luminance contrast to performance for features such as stimulus orientation. Contrast-discrimination thresholds increase substantially with response delay periods longer than a few seconds, unlike performance profiles for stimulus properties such as orientation, motion, or spatial frequency, which are much less affected across prolonged delays e.g., 20s or more (Magnussen & Greenlee, 1999) (Figure 1.2A). These findings likely reflect differences in how the relevant stimulus information is neurally represented (i.e., intensity vs. topographic encoding), and how these representations decay over time. In the case of orientation, the topographically-arranged orientation maps provide a natural substrate for precise coding, while no analogous representation facilitates the precise coding of contrast (Xing et al., 2014).

1.4.2 ROLE OF SET-SIZE ON ENCODING-DECODING PERFORMANCE

The effect of stimulus set-size on encoding-decoding precision has been extensively studied, both for attention-related tasks (e.g., visual search) (Palmer, 1990; Mazzyar et al., 2012) and for studies of VSTM using paradigms such as change detection and change localization (Pashler, 1988; Luck & Vogel, 1997; Bays & Husain, 2008). In earlier studies, data were often interpreted as reflecting a bound on performance as a function of set-size, with set-sizes greater than four thought to require more than the available number of encoding slots in memory (Cowan, 2001; Luck & Vogel, 1997). However, until recently such studies typically avoided detailed parametric stimulus manipulation i.e., in terms of the exact set-sizes and the stimulus parameters used. This has changed in recent years, with parametric manipulation across one or more feature dimensions now the norm for research on topics such as VSTM and visual search (Bays & Husain, 2008; Lara & Wallis, 2012; Keshvari et al., 2013). For example, visual search performance has now definitively been shown to systematically decrease as the number of stimuli that make up the search array increases (Mazzyar et al., 2012). Similar effects hold for other paradigms (Bays & Husain, 2008).

Researchers have also taken a more direct approach in recent years to measuring the precision of stimulus encoding and decoding, by utilizing newer paradigms such as delayed estimation. Similar to the matching paradigms briefly discussed above, in a delayed-estimation task an observer is first presented with a brief stimulus display, and after a short delay, must attempt to match the input stimulus precisely (Wilken & Ma, 2004; Bays & Husain, 2008; Fougine et al., 2012; van den Berg et al., 2012; Bays, 2014). Delay time in such tasks is typically controlled, and on the order of 1-2s, thus focusing more on processes of encoding and maintenance into VSTM than perception. For example, in a subset of their

experiments, Wilken & Ma (2004) had observers estimate the orientation, color, or spatial frequency of a stimulus presented 1500ms earlier. As set-size increased, estimate precision decreased in monotonic fashion. Similar results were recently found with highly parametric variation of set-size by; as set-size increased from 1 to 8 in an orientation-estimation task, for example, the error (i.e., circular SD) of estimates increased in a systematic fashion (Figure 1.2B). Counter to the traditional view of VSTM as being made up of slots (Luck & Vogel, 1997; Cowan, 2001), these general results have in recent years come to be understood in terms of a decrease of precision, similar to a view already more accepted for attention (Ma et al., 2014). It should also be noted, that by providing a more continuous measurement of encoding quality, estimation tasks provide richer datasets that lend themselves more naturally to detailed computational model fitting.

1.4.3 ROLE OF STIMULUS INTERACTIONS ON ENCODING-DECODING PERFORMANCE

The effects of distractor stimuli on encoding-decoding precision goes beyond the basic set-size effect. Even when an observer knows in advance that only one stimulus is relevant for a task, there can still be an effect of irrelevant distractors on performance. Systematic investigation of such effects has so far been limited; yet, a number of recent investigations on the topics of attentional selection and VSTM have utilized paradigms where the effects of direct stimulus interactions on behavioral performance can be studied. Of these, a number of studies relevant to the topic of luminance contrast encoding-decoding have emerged (Chen & Seidemann, 2012; Hara & Gardner, 2014; Itthipuripat et al., 2014; Pestilli et al., 2011), as well as theoretical models for studying the general effects of stimulus interactions on decoding (Matthey et al., 2015; Orhan & Ma, 2015).

Results of some of these studies support the view that VSTM for luminance contrast is severely hampered by the presence of irrelevant distractors. This behavior was nicely demonstrated in a series of contrast-discrimination experiments in which observers had to either focus or distribute their attention across multi-item displays (Pestilli et al., 2011; Hara & Gardner, 2014). The presence of a single high-contrast distractor was sufficient to severely disrupt threshold performance at the target location (Pestilli et al., 2011) (Figure 1.2C). The large responses evoked by high-contrast distractors must have dominated in the selection of sensory signals for decision, supporting a model of encoding-decoding using something akin to a max-pooling operation as a decision rule (Pestilli et al., 2011; Palmer et al., 2000; Pelli, 1985).

A number of other recent findings provide a window into the question of stimulus interactions and encoding-decoding precision. For example, using a delayed-estimation paradigm, Bays et al. (2009) found that estimates of the color of a recently presented target item appeared to be accompanied by a size-able proportion of non-target color reports. The authors described a model of this process in which memory for item color and location could interact, thereby systematically affecting the shape of estimate response distributions for target color. This type of effect highlights the difficulty faced by the brain in disentangling the actual sources of sensory responses in the context of multiple stimulus displays. Presumably, as has been found at numerous levels of visual processing, individual neurons might compute weighted sums of constituent inputs, effectively mixing neural responses originating from separate sources (Busse et al., 2009; Recanzone et al., 1997; Zoccolan et al., 2007). This line of thinking has recently been formalized in an encoding-decoding architecture based on the linear-mixing of separate sensory neural responses (Orhan & Ma, 2015).

One key finding of this theoretical work has been to show that mixing of neural responses is substantially more detrimental to decoding performance than gain modifications reminiscent of attentional modulation. In Chapter 3, we explore a model of this sort.

1.5 THESIS SYNOPSIS

The present thesis consists of a complementary pair of investigations that further our understanding of the neural representations and computational rules governing the encoding and decoding of luminance contrast. The thesis builds on the wide variety of studies that relate to the topics of sensory encoding and decoding. Each investigation was comprised of multiple psychophysical experiments with human observers, and involved subsequent fitting of neural computational models to each dataset.

In Chapter 2, we present results from a series of 2-AFC discrimination experiments that directly question the principled nature of VSTM for luminance contrast: in the presence of irrelevant, distractor stimuli, contrast-discrimination performance is found to deteriorate more substantially relative to decoding performance for another low-level, visual feature (orientation). We address this apparent discrepancy in Chapter 3, by simultaneously fitting a probabilistic model of neural responses to each of the datasets described in Chapter 2. Performance of observers across the different discrimination tasks can be adequately accounted for when sufficiently general, but nonetheless neurally-plausible, model components are combined (e.g., divisive normalization and linear mixing of sensory neural responses, tuning function heterogeneity). We then dig deeper in Chapter 4, studying the trial-by-trial precision of VSTM for luminance contrast. Results are presented from a series of delayed-estimation experiments, allowing us to more clearly define the likely neural

constraints governing the encoding and short-term retention of luminance contrast information. Behavior of human observers on these tasks is shown to be highly principled: performance is captured satisfactorily by a probabilistic model of neural responses incorporating biologically-plausible model components (e.g., Poisson spiking).

2

Efficient and inefficient selection from the same sensory neural response: psychophysics

2.1 INTRODUCTION

TO SUCCESSFULLY PERFORM EVEN BASIC VISUAL TASKS, humans must efficiently select and manipulate relevant environmental information. The brain processes this sensory information in various stages, from initial encoding, to the maintenance of signals in visual short-term memory, to the application of appropriate sensory read-out rules. A long tradition of research in sensory neuroscience has studied such processing using very simple stimuli and tasks, allowing for maximal experimental control and model tractability. For example, human visual performance has been commonly characterized by studying how observers discriminate the contrasts (Nachmias & Sansbury, 1974; Legge & Foley, 1980; Boynton et al., 1999) or orientations (Westheimer et al., 1976; Blake & Holopigian, 1985; Skottun et al., 1987) of two successively presented sinusoidal gratings, so-called two-alternative, forced-choice (2-AFC) discrimination tasks. Performance or threshold on such tasks is typically defined as the amount of stimulus change needed to achieve a criterion level of behav-

ioral performance (e.g., 75% correct). When plotted as a function of the pedestal stimulus contrast, contrast-discrimination thresholds typically follow a "dipper" shape, first decreasing with small increases in pedestal contrast above background luminance, and then increasing at higher values of pedestal contrast (Nachmias & Sansbury, 1974; Legge & Foley, 1980). On the other hand, as pedestal contrast increases, the deviations in pattern orientation necessary to reach a threshold level of orientation-discrimination performance typically decrease in a systematic and monotonic fashion (Skottun et al., 1987).

Single-stimulus tasks afford the researcher precise experimental control and simplicity, yet such paradigms are entirely unlike natural vision, in which multiple objects are present at once. In fact, despite substantial progress in understanding how discrimination behavior is linked to local neural computations (Paradiso, 1988; Boynton et al., 1999; Sanborn & Dayan, 2011; Berens et al., 2012), relatively little is known about the selection or decoding strategies implemented by observers when faced with more complex sensory input (i.e., multiple possible target stimuli). Thus, there is a need to develop experimental paradigms where the joint encoding of multiple stimuli is required for successful performance. We focus here on the role of irrelevant distractors in multiple stimulus encoding; specifically, on tasks in which the observer attempts to select from memory information from a single, post-cued stimulus location (Hara & Gardner, 2014; Itthipuripat et al., 2014; Pestilli et al., 2011; Sergent et al., 2011). We do not consider tasks in which all items are actually relevant for the task, such as global target detection (Ma et al., 2011; Palmer et al., 2000).

How do irrelevant distractors influence behavioral performance in discrimination tasks? While direct evidence is so far weak, there are hints that discrimination performance with more complex sensory input varies as a function of the particular task and encoding con-

straints under investigation. For example, unlike the known topographic neural representation for stimulus features such as orientation, the neural representation for contrast is based fundamentally around response intensity (Albrecht & Hamilton, 1982), a characteristic that likely hinders the formation of abstract memory representations for contrast (Xing et al., 2014). Thus, sensory evidence for a particular contrast may be available only briefly after stimulus disappearance, and in imprecise form, encouraging non-selective pooling across stimuli when multiple estimates are made simultaneously (Pestilli et al., 2011; Hara & Gardner, 2014). For example, Pestilli et al. (2011) recently found that when attention was distributed across multiple stimuli that varied in contrast, the presence of a single high-contrast distractor was sufficient to severely disrupt contrast-discrimination performance at a target location. However, sensory evidence for other stimulus properties (e.g., orientation) may be maintained with greater precision and for longer post-stimulus delays. (Magnussen & Greenlee, 1999). Thus when selecting from sensory signals at different spatial locations, performance on tasks such as orientation discrimination may be much less influenced by distractors or prone to sub-optimal decision rules, a conjecture with some experimental support (Anderson et al., 2013; Sergent et al., 2011).

To test this possibility, we ran separate experiments in which observers discriminated changes to either the contrast or orientation of a target stimulus presented in the hemifield opposite to a distractor. Identical stimulus protocols were used across experiments, the only difference being that small contrast increments were added to the target location in the contrast experiment, while small orientation deviations were added to the target in the orientation experiment (Figure 2.1). To measure the effect of distractor strength on performance, we systematically varied the pedestal contrasts assigned to targets and distractors. Results re-

inforced the view that for the two forms of discrimination tested, observers appear to select from identical sensory neural responses in incommensurate ways. In a subsequent chapter, we fit these data using a neural population model, by deriving mathematical expressions based on Fisher information, for the simultaneous estimation of contrast- and orientation-discrimination thresholds from an idealized neural population.

2.2 MATERIALS AND METHODS

2.2.1 PARTICIPANTS

Data from the same eight observers (two authors) were collected in both of the main experiments. Five observers completed the contrast-discrimination experiment prior to the orientation-discrimination experiment, three observers ran in reverse order. Experimental sessions were typically performed over a 2-3 week period, with the different experiments separated by up to several months. Observers were recruited from the general student/staff body at New York University (paid \$10/hr) and amongst lab colleagues, and had varying degrees of experience in psychophysical testing. One additional recruit was not tested beyond the practice session, during which this individual confirmed being diagnosed with an attention-related disorder. Aside from the authors, observers had no knowledge of the specific experimental hypotheses. All observers gave written informed consent, and experiments were carried out with approval of the NYU University Committee on Activities Involving Human Subjects.

2.2.2 CONTRAST DISCRIMINATION

TASK

We tested the effects of distractors of different contrast on contrast-discrimination performance at a post-cued target location (Figure 2.1). Each trial began with the presentation of pre-stimulus arrows pointing left and right of fixation (1 s), cueing the observer to distribute attention equally to two peripheral locations (6° eccentricity), while remaining fixated on a central fixation cross (1° width). After a short delay (100 ms) and auditory tone indicating stimulus onset, a pair of gratings were briefly presented in a first stimulus interval (600 ms), one positioned left and the other right of fixation along the horizontal meridian. The gratings then disappeared, and after a short ISI (200 ms), the gratings reappeared for a second stimulus interval (600 ms). A positive contrast increment was added to one of the gratings (the ‘target’) in one of the two intervals. After the stimuli had left the screen, there was a second short delay (400 ms), followed by presentation of a green arrow indicating the target location. Observers responded during this interval by pressing one of two keys on the keyboard (‘1’ or ‘2’), judging which of the two stimulus intervals contained a higher contrast at the target location. This response interval was of fixed duration (1200 ms), and trials with no response were not replaced. Observers received feedback on each trial (color change of the fixation cross and auditory tone), and were instructed to perform as accurately as possible throughout sessions. Trials were separated by an ITI of pseudo-random duration (800-1200 ms, 100 ms steps). Observers completed five sessions (480 trials per session, \sim 1hr duration). We regarded the first session as a practice session and analyzed only the final four sessions. Observers received a mandatory rest period after every block of 120

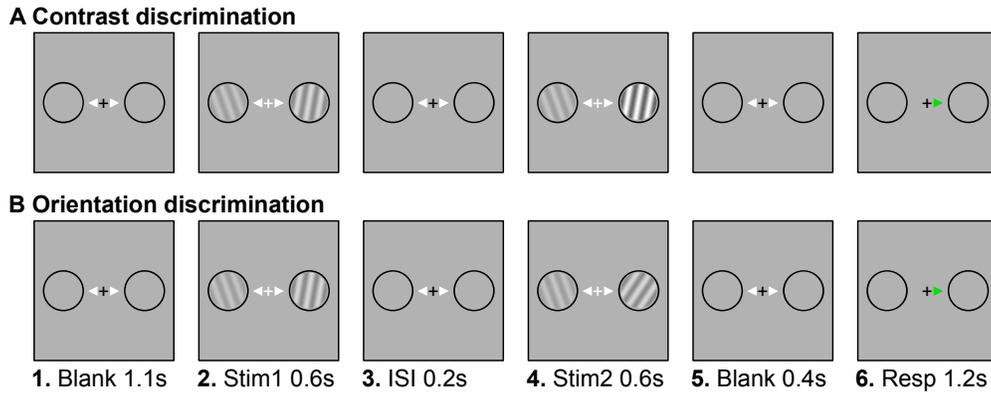


Figure 2.1: Experimental set-up. Contrast- and orientation-discrimination experiments were carried out in separate sessions (i.e., non-interleaved), using a largely identical set-up. Trials began with the presentation of pre-cues pointing left and right of fixation, cueing the observer to distribute attention equally to two peripheral locations (6° eccentricity). A pair of gratings (5° diameter) were then briefly presented in the first stimulus interval, left and right of fixation. The gratings then disappeared, and after a short ISI, reappeared for a second interval. In the contrast-discrimination experiment, a positive contrast increment was added to a target grating in one of the two intervals; in the orientation-discrimination experiment, a clockwise or counter-clockwise orientation increment was added to the target grating. After the stimuli had left the screen, a post-cue (green arrow) indicated the target location. Observers judged in which of the two intervals the target grating had either higher contrast or was rotated more clockwise.

trials, and could also pause presentation at any time by pressing the space bar.

STIMULUS DESIGN AND EXPERIMENTAL CONDITIONS

Stimuli were presented in a darkened room on a gamma-corrected CRT (75 Hz, 1152 x 870 resolution), and were generated using MATLAB (The Mathworks) and MGL (see <http://justingardner.net/mgl>). For the first phase of experiments, we used a gamma correction table calculated some time prior to testing (~ 24 months earlier); a later correction suggested that drift had occurred from original table values. However, mean and maximum luminances always lay in typical ranges for these types of experiments (approximately 35 and 70 cd/m^2 respectively, plus or minus a small amount), and we are not concerned this drift had any meaningful consequences for our stimulus comparisons (we used a rel-

atively coarse sampling of contrasts from the entire luminance contrast range). Gratings were counter-phase flickering sinusoids (5 Hz, 2 cycles/°) measuring 5° in diameter. Gratings were presented inside black circular frames, such that a small gap lay between the frame and the grating edge (raised-cosine, edge width 0.5°). Target and distractor gratings could appear with one of four pedestal contrast values (10, 20, 40 or 80% contrast). Contrast was defined in standard Michelson contrast form: $c = (L_{\max} - L_{\min}) / (L_{\max} + L_{\min})$. An experimental condition is defined as a combination of target and distractor contrast. All combinations of target and distractor contrast were presented, excluding those conditions where target and distractor would appear with identical pedestal contrast. The reason for this exclusion was that in such conditions, an ideal observer could use information from a single interval to perform the task, as positive contrast increments were always added to the target pedestal. Thus, we measured contrast-discrimination thresholds for 12 target-distractor contrast pairs in total.

40 trials per condition were presented in each session, and conditions were randomly interleaved. Of the 40 trials per condition, increments on 32 trials were controlled by an adaptive, 1-up-2-down staircase, with increments on the remaining trials hand-picked on a session-by-session basis. For the hand-picked increment trials, which were randomly interleaved with staircase trials, increments were typically set to low and/or high values so as to improve the quality of fit to baseline/asymptotic performance across session (occasionally some intermediate fixed increments were also presented). In all conditions except 40% and 80% pedestal contrast, we used contrast increments of 0.5, 1, 2, 4, 8, 12, 16, 24, 32, 48 and 64%. When pedestal contrast was 40% or 80%, we modified the increment array such that contrast would not exceed 100% (e.g., for an 80% contrast target, the largest possible incre-

ment was 20%). In a couple of early sessions, we included increments of 6%, 10%, 20% and 40% in the general array; we retained any trials using these increments for analysis, except for a small handful of trials in which target-plus-increment inadvertently equaled distractor contrast (i.e., 20% + 20% = 40% contrast, 8 trials for one observer according to our analysis routine). Staircase endpoints from the practice session were used as staircase starting points in the first test session, and similarly each subsequent test session began with staircase endpoints from the session prior. In anticipation of the orientation discrimination experiment, target and distractor gratings took one of ten pseudo-random orientations (9, 27, 45, 63, 81, 99, 117, 135, 153 or 171°). For each of the 12 conditions present within a 120-trial block, each of the 10 possible orientations was used as target orientation exactly once. The frequency of distractor orientations was not controlled in a similar fashion, but pseudo-randomized such that distractor orientation always differed from target orientation on any given trial. On any trial, orientations were held constant across intervals.

2.2.3 ORIENTATION DISCRIMINATION

TASK, STIMULUS DESIGN AND EXPERIMENTAL CONDITIONS

Aside from a separate training phase (see below) and the increment type (i.e., orientation), all aspects of stimulus presentation and protocol were largely identical to the contrast-discrimination experiment. On each trial, a clockwise or counter-clockwise orientation increment was added to the target grating, and observers judged which interval contained the more clockwise oriented stimulus at the target location, which was post-cued as in the contrast task. In addition, all 16 combinations of target-distractor contrast were presented (all combinations of 10, 20, 40 and 80% contrast). Given the larger number of conditions,

observers completed six 480-trial sessions, with thresholds estimated from the final five sessions. Each session was divided into three 160-trial blocks, and given the slightly longer blocks, observers were encouraged to pause presentation once or twice per block as needed. In each session, thirty trials were presented for each of the sixteen target-distractor contrast conditions. The staircase approach was similar to that used for the contrast-discrimination experiment, with a fixed array of orientation increments provided for each staircase (increments of 1, 2, 4, 6, 8, 12, 16 and 24°), and with a small proportion of trials set to increments which were hand-picked from the fixed array on a session-by-session basis.

TRAINING PROTOCOL

Our training protocol was informed by an initial version of the orientation discrimination experiment. This version was as described above but included 32° orientation increments. Two observers completed several sessions of this experiment (3 and 4 sessions, respectively). We found that their performance in a large number of conditions was near chance even for the 32° increment. This may have resulted from several factors, including the large range of pedestal orientations used (spanning 180°), the time-limited response interval, and the direction ambiguity inherent in circular orientation space (which might be particularly problematic for the 32° increment).

We excluded the data from these early sessions from further analysis and in response to the noted problem, we made several modifications to the experimental design. First, we limited the orientation increments to a maximum of 24° . Second, we introduced a training protocol with fixed large orientation increments (24°) only. During training, observers were verbally encouraged to focus on the global, rotational nature of the discrimination (as

opposed to focusing on the left/right tilt of one end of the grating), and were instructed that the large orientation increments presented in training were the largest possible orientation changes they would experience throughout testing; feedback from several observers confirmed that the large increments were typically well above detection threshold. Trials were presented in blocks of 40 trials, using a pseudo-randomly selected subset of the contrast conditions from the main experiment. Specifically, we ensured in each block that each of the four contrasts appeared once as a target and once as a distractor, and that conditions were matched in opposite pairs so that the target could not be distinguished based on contrast (i.e., if and observers were told their percentage correct after each block. In an effort to step up the difficulty level gradually, we used three different block types: block type 1 used a single grating at fixation only, presented over two intervals; block type 2 used two-grating displays with target pre-cueing; block type 3 used two-grating displays with distributed attention pre-cueing, like in the main experiment. Observers could move from one block type to the next when they scored consistently in the range of 80-90% in a block, as judged by the experimenter. In total, this amounted to observers completing 1 to 3 training sessions of 50 to 60 minutes each. In a final modification, observers were required to complete one or more short warm-up blocks at the beginning of each test session, with the exact number of blocks determined by the experimenter based on an online appraisal of the consistency of the observers performance relative to the pre-test training sessions.

2.2.4 THRESHOLD ESTIMATION AND STATISTICAL TESTS

MAIN ANALYSES

We estimated thresholds separately for each observer in each experiment. For a given experiment, we combined data from the final four (contrast) or five (orientation) sessions, so that we had 160 (contrast) and 150 (orientation) data points per condition from which to estimate threshold. Trials on which no response was made were excluded (less than 1% of trials), meaning that the actual numbers of trials per condition were on average slightly less than the numbers above. We fit a Weibull function to the data from each condition,

$$\text{Proportion correct}(\delta s) = 0.5 + (0.5 - p_{\text{lapse}}) \left(1 - \exp \left(-\left(\frac{\delta s}{a}\right)^b \right) \right), \quad (2.1)$$

where δs is the orientation or contrast increment, $1 - p_{\text{lapse}}$ is the asymptotic proportion correct at very large increments, and the parameters a and b control the midpoint (i.e., bias) and steepness (i.e., slope) of the psychometric function, respectively. To impose the constraint of a fixed lapse rate across conditions, we adopted the following fitting procedure. We varied p_{lapse} in steps of 0.001 between 0.001 to 0.06. At each step, we used maximum-likelihood estimation, implemented using the MATLAB function *fminsearch.m*, to find the best-fitting a and b independently for each condition. The log likelihoods of the resulting fits to the individual conditions were then summed, to give a combined log likelihood value for each lapse rate. The p_{lapse} with the highest log likelihood, and the associated collection of bias and slope value pairs, were taken as the best-fitting parameter values for an individual observer. We then derived thresholds from the fitted psychometric functions by finding the increment necessary to achieve 75% correct performance (Figure 2.2A). A small

number of occurrences of threshold going outside increment boundaries were left as is (e.g., contrast-discrimination threshold exceeding 20% for 80% contrast targets); they played no meaningful role in comparisons of most interest.

To measure the probability of observing mean threshold differences by chance for different distractor contrast levels, the following statistical test was run. Using a randomization analysis, we calculated p-values for each possible distractor pair comparison, separately at each target contrast value. For each comparison, this involved pooling the pair of distractor response distributions for an observer, and re-sampling trials for the two conditions from the randomly shuffled pooled distribution. Psychometric functions were fit to the re-sampled pair, and threshold difference computed (with p_{guess} set to 0.5 and p_{lapse} held fixed at the value calculated in the main analysis). This sampling procedure was repeated 10,000 times to generate a distribution of threshold differences for each pair. The same process was applied to each observer's data separately, such that for each comparison we had 8 observers x 10,000 threshold differences. From these, we calculated a mean distribution for each comparison (averaging across the observers' unsorted distributions at each sample). The resulting mean distribution was sorted from smallest to largest, and the probability of observing by chance the measured mean threshold difference was read off from this sorted distribution (by finding the index with minimum absolute difference from the measured mean threshold difference, ignoring the sign of that minimum difference). Exact p-values are reported in text and tables (rounded to three decimal places, or left as is if $p < 0.001$).

CONTROL ANALYSES

As a precautionary measure, we repeated the main analyses with two modifications. First, we re-ran the randomization analysis after excluding trials containing saccades during the stimulus intervals (see Eye-movement recording). Second, to account for the possibility that some fits might have been over or under-dispersed due to, for example, across-session learning or experimenter bias in handpicking increment values, we re-calculated the p-values after performing a deviance analysis on the psychometric function fits (Wichmann & Hill, 2001). Specifically, the deviance statistic was computed for each psychometric function fit, and then compared to a distribution of simulated deviance statistics (10,000 samples). This distribution was computed using the original fit parameters as generating model, and calculating deviance at each sample between the simulated data and the best-fit psychometric function to the simulated data. Individual fits whose deviance statistic lay outside a relatively narrow confidence interval (84%) were excluded from p-value calculation.

We also carried out two additional control analyses on sub-portions of the data. First, we determined whether the increased exposure observers had from pre-training on the orientation task dampened distractor effects in subsequent test sessions, relative to the size of effects observed in the contrast-discrimination task at least. To test this possibility, we pooled individual observer data from the first test session of the orientation-discrimination task (excluding the practice run), to create a super-observer dataset from which we estimated thresholds. We created a similar pooled dataset from the contrast-discrimination experiment, this time combining data from the final test session, so that our comparison was between sessions in which observers had approximately as much or more prior exposure in

the contrast-discrimination task. A randomization analysis was performed on psychometric function fits to the pooled datasets.

In a final control analysis, we determined whether the difference in orientation between target and distractor stimuli played any role in performance in either experiment (e.g., through grouping of similar orientations, or other spatially broad interactions). To test this possibility, we divided the datasets into two parts - in one part, we placed trials in which the distractor was oriented either 18° or 36° clockwise or counter-clockwise of the target (i.e., closer to parallel); in the other, we placed trials in which the distractor was oriented 54° , 72° or 90° away from the target (i.e., closer to orthogonal). It was our hunch that such effects, if they existed, would likely be small in size, so we pooled data across observers to emphasize mean differences. For a given experiment, we then compared matched conditions across the two portions of data using randomization, to ascertain whether the degree of similarity in stimulus orientations had any obvious effects on performance.

2.2.5 EYE-MOVEMENT RECORDING

In each session, eye position (right eye, 500 Hz) was recorded using an Eyelink 1000 (SR Research) and analyzed offline using custom MATLAB routines. Before each block, a calibration routine (5-pt or 9-pt) was run. Trial onset was controlled in a gaze-contingent manner, beginning only after fixation was maintained within $2\text{-}2.5^\circ$ of fixation for 250 ms. As our experiments involved relatively long duration trials and blocks, most observers systematically blinked during response and inter-trial intervals to limit eye fatigue. To focus our analysis on intervals of interest, we analyzed only position data from the onset of the first stimulus pair until disappearance of the second pair (1.4 s total). The trial-by-trial saccade detection

proceeded as follows, closely following default Eyelink criteria and other well-accepted conventions (Engbert & Kliegl, 2003): from trial onset to first stimulus onset, we calculated the median horizontal and vertical eye position; these values were subtracted from the position data within the analysis window, so as to limit the effect of recording drift across individual blocks. Velocity along the horizontal and vertical axes was calculated by applying a sliding 5-pt window to the position data, and Euclidean velocity was then calculated. Euclidean acceleration was calculated in a similar fashion. To avoid contaminating the saccade detection analysis with blinks, samples that corresponded to blinks (and 100 ms either side) were removed, by searching for intervals where pupil size data was not recorded. Saccades were detected by searching for samples where velocity exceeded $30^\circ/\text{s}$, peak acceleration exceeded $8000^\circ/\text{s}^2$, and amplitude (i.e., Euclidean distance from rising above to falling below $30^\circ/\text{s}$) exceeded 0.5° . We repeated the randomization analyses of the main experiments after excluding trials containing saccades during the analysis window (between $\sim 1\text{--}13\%$ of total trials in different observers, using this relatively small saccadic cut-off). Eye data from a small number of blocks (4 out of ~ 200) was accidentally overwritten during the course of running experiments.

2.2.6 CONTROL EXPERIMENT

In a control experiment, we replicated the main contrast-discrimination experiment, this time providing separate response button pairs for each hand (four-button task). Data were collected from an overlapping group of eight observers (two authors), with one additional recruit withdrawing after completing the practice and first test session of the task. Participants were instructed to discriminate in which interval the target grating had higher con-

trast, this time using the response keys on the side in which the target appeared (indicated by the post-cue) e.g., pressing the nearer of the two buttons ('v' or 'n') for interval 1, or the further of the two buttons ('f' or 'j') for interval 2. This allowed us to estimate thresholds using only trials on which the observer explicitly indicated having responded towards perceived changes at the target location (note that non-target responses were considered incorrect responses, and still modified the staircase position). This procedure also allowed us to estimate the relative frequency of non-target responses across different conditions.

2.3 RESULTS

Performance across experiment differed in terms of its distractor-dependence, as illustrated by the mean thresholds across observer plotted in Figure 2.2B. Mean contrast-discrimination thresholds were estimated to be around 10% contrast for the low and intermediate (10, 20, and 40%) distractor contrast conditions, but were approximately doubled for targets paired with the highest contrast distractor (Figure 2.2B, left). This increase in threshold was not restricted to low-contrast targets, occurring for targets of 10, 20, and 40% contrast. Note that we avoided target contrast values in the very low range, where monotonic increase of threshold is typically most evident, and that the relatively high baseline thresholds observed in this divided attention task are not at odds with previous findings (Pestilli et al., 2011). For orientation discrimination, thresholds were lower overall with higher target contrast (Figure 2.2B, right), albeit with higher baseline threshold than one would obtain under conditions without target location uncertainty (Skottun et al., 1987). Distractor contrast had comparatively weaker effects on orientation discrimination: while mean thresholds increased with each increase in distractor contrast, these effects were graded in fashion, with

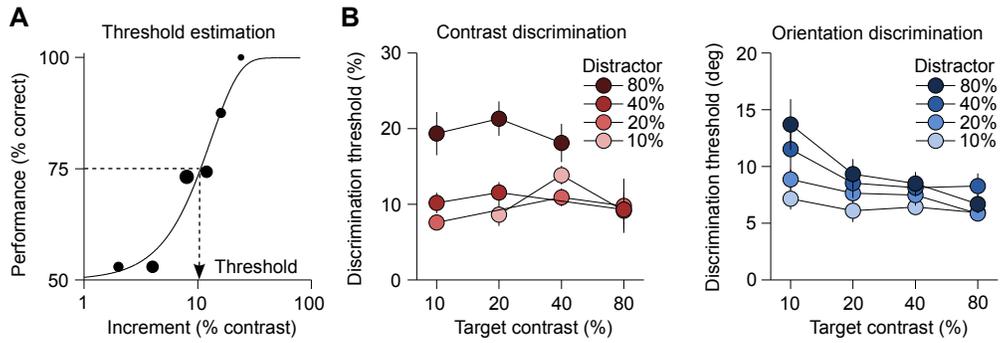


Figure 2.2: Experimental data. A) Example psychometric function and threshold estimation. B) Mean contrast-discrimination (left) and orientation-discrimination thresholds (right) across observers ($n = 8$). Individual curves represent threshold as a function of target contrast, plotted separately for each distractor contrast level. The same eight observers completed both experiments. Bars represent standard error across observers.

no indication of a large threshold jump from 40% to 80% contrast distractor conditions.

These general conclusions were backed up by statistical test. At each target contrast level within each experiment, we performed pair-wise randomization tests via bootstrapping, to estimate the probabilities of observing distractor-mediated differences in threshold by chance (Table 2.1). This pair-wise analysis illustrated comparatively greater distractor influence in the contrast experiment. For 10% contrast targets, for example, contrast-discrimination thresholds were much more affected by 80% compared to 40% ($p = 0.0001$) or 20% contrast distractors ($p = 0.0001$), while the probabilities of observing such differences for orientation-discrimination thresholds were comparatively weaker (80% vs 40%: $p = 0.122$; 80% vs 20%: $p = 0.002$; 80% vs 10%: $p = 0.0001$). Similarly for 20% contrast targets, contrast-discrimination thresholds were again much more strongly affected by 80% compared to 40% ($p = 0.0001$) or 10% contrast distractors ($p = 0.0001$), while the probabilities of observing such differences for orientation-discrimination thresholds were comparatively weaker (80% vs 40%: $p = 0.227$; 80% vs 20%: $p = 0.047$; 80% vs 10%: $p = 0.003$).

	Distractor pair					
	10% vs 20%	10% vs 40%	10% vs 80%	20% vs 40%	20% vs 80%	40% vs 80%
<i>Contrast discrimination</i>						
<i>Target</i>						
10%	-	-	-	.0080	.0001	.0001
20%	-	.0291	.0001	-	-	.0001
40%	.9704	-	.0046	-	.0001	-
80%	.3997	.4595	-	.6349	-	-
<i>Orientation discrimination</i>						
<i>Target</i>						
10%	.0800	.0032	.0001	.0549	.0023	.1217
20%	.0373	.0114	.0030	.1978	.0471	.2274
40%	.1302	.0307	.0119	.2637	.1672	.3524
80%	.5452	.0057	.1422	.0068	.1256	.9654

Table 2.1: Randomization analyses on threshold differences as a function of distractor pair. P-values for each condition were read off from bootstrapped distributions with 10,000 samples each (see Materials and Methods). Smaller p-values indicate larger thresholds for the higher contrast distractor condition.

To rule out a variety of extraneous factors as possible explanations of the data, we ran several control randomization analyses. First, for both datasets we re-calculated thresholds after removing trials containing saccades of 0.5° amplitude or greater (see Materials and Methods and Figure 2.3A). The overall pattern of results across experiment was unchanged (e.g., for 10% contrast targets, probabilities of difference for 80% vs 40% distractors were: $p = 0.0001$ for contrast discrimination and $p = 0.319$ for orientation discrimination; for 20% contrast targets, the probabilities were: $p = 0.0001$ and $p = 0.255$ respectively). Thus, eye movements during the stimulus intervals are unlikely to explain the differences observed across experiments. In addition, we also re-calculated the main randomization p-values after excluding data whose psychometric function fit failed to pass a two-tailed de-

viance analysis measuring goodness-of-fit (see Materials and Methods). The overall pattern of results was again similar (e.g., for 10% contrast targets, probabilities of difference for 80% vs 40% distractors were: $p = 0.002$ for contrast discrimination and $p = 0.03$ for orientation discrimination; for 20% contrast targets, the probabilities were: $p = 0.0001$ and $p = 0.188$ respectively). Thus, factors such as observer fatigue, across-session learning, or experimenter bias in hand-picking increment values are unlikely to have played any causal role in the observed differences across experiment.

Finally, we also carried out two additional analyses on data pooled across observers (see Materials and Methods). First, we verified that the increased exposure observers had on the orientation-discrimination task (due to a pre-experimental training protocol) played no obvious role in our observed experimental differences, by comparing subsets of the data for which observers had as much or more exposure to the contrast-discrimination task relative to the orientation-discrimination task. Effect sizes were comparable to results of our main analyses. In addition, we also carried out an analysis to verify whether distractor orientation played any role in the measured threshold behaviors. To do this, we split the data sets into two parts, trials in which target and distractor had closer to parallel orientations, and trials in which stimuli were closer to orthogonal. We then re-calculated thresholds for each set of trials separately, and computed the threshold differences (thresholds for orthogonal trials minus thresholds for parallel trials). While the resulting threshold differences were generally small for both tasks (on the order of 1 or 2° or 1% contrast for many of the individual conditions), there did appear to be a consistent orientation similarity effect in the orientation-discrimination task, with thresholds for near parallel trials smaller on average than thresholds for orthogonal pair trials, with most benefit at lower contrast targets (Fig-

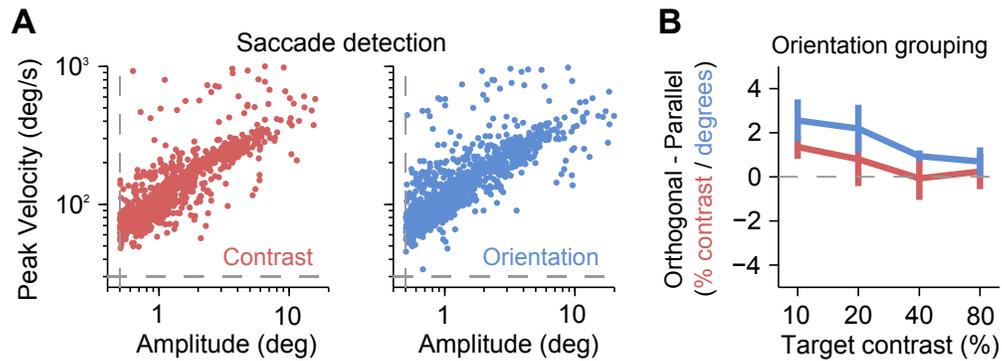


Figure 2.3: Control analyses. A) Main sequence data from the each experiment (pooled data). Thresholds were re-computed after removing saccade trials (see main text for randomization analyses). B) Threshold difference between trials with orthogonal vs. parallel orientation pairs. Thresholds were calculated from pooled observer data (see Materials and Methods). Positive values indicate larger thresholds for trials in which the orientation pairs were closer to orthogonal. SE is calculated across the three (contrast) or four (orientation) levels of distractor contrast.

ure 2.3B, right).

CONTROL EXPERIMENT

A number of recent studies have questioned the apparent role of sub-optimal response pooling on behavioral decision in contrast-discrimination tasks like ours (Itthipuripat et al., 2014). For example, it has been suggested that the high-contrast distractor effects observed by Pestilli et al. (2011) could have arisen if observers used a strategy in which they represented the average contrast of all items in a single interval, and took the larger responding interval as their response. In addition, it is possible that observers were on some fraction of trials responding to perceived contrast changes at the distractor location. In an attempt to better understand the strategy employed by observers in the contrast-discrimination task, we replicated our experimental design with a new group of observers, this time providing observers with separate response button pairs for each hand, and instructing them to respond using the target-sided hand only. This allowed us to estimate thresholds using only

trials on which the observer explicitly indicated having responded towards the target location. In addition, it allowed us to estimate the relative frequency of non-target responses across the different conditions.

Data are illustrated in Figure 2.4. Despite having separate response button pairs for target and non-target stimuli, thresholds estimated from the target-sided responses appear again to be highly distractor-dependent (Figure 2.4A). A randomization analysis comparing the mean thresholds across different distractor pairs again illustrated substantial threshold increase with the highest-contrast distractor. Nevertheless, the experiment revealed a tendency for observers to make more non-target responses in higher-contrast distractor conditions (2.4B), suggesting some amount of above-threshold driving of responses by the distractor.

2.4 CONCLUSION

We investigated how observers select sensory information in performing contrast and orientation discriminations, by measuring the effects of high-contrast distractors (i.e., large sensory responses) on behavioral performance in these tasks. Prior work had shown that when selecting from multiple stimuli that vary in contrast, distractors that evoke large sensory responses severely impact contrast-discrimination performance (i.e., lead to larger thresholds), supporting a model of sensory selection in which sensory responses are sub-optimally pooled across space (Chen & Seidemann, 2012; Hara & Gardner, 2014; Pestilli et al., 2011). We observed large increases of contrast-discrimination thresholds when targets appeared in the presence of a single high-contrast distractor placed in the opposite hemifield. For orientation discrimination, however, high-contrast distractors had relatively moderate effects on

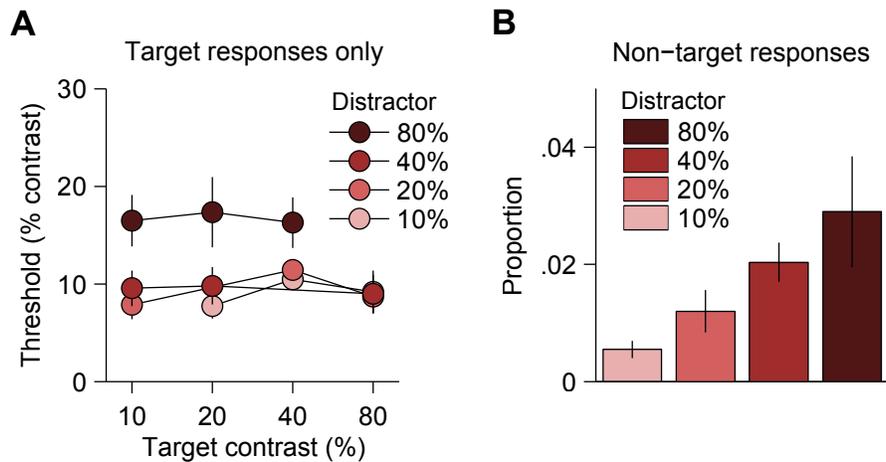


Figure 2.4: Control experiment. A) We replicated the main contrast-discrimination experiment, this time providing observers with two response buttons for each hand, and instructing them to respond with the target-sided hand only. Mean thresholds from target response trials only ($n = 8$). B) Proportion of non-target responses as a function of distractor contrast, averaged over target contrasts and SE then taken across observers.

performance at a target location, disrupting thresholds in a weaker, graded fashion.

What factors underlie the seemingly incommensurate behavior we observed in the separate experiments? In either task, an ideal observer would retain an estimate of the relevant stimulus property from both stimulus locations during interval one, repeat this process for interval two, and compare the difference in estimates across interval. Yet, fundamental differences in how estimates of contrast and orientation are encoded and maintained over time likely give rise to the profiles of threshold behavior we observed. For example, contrast-discrimination performance with single apertures is known to fall off rapidly with response delay periods of only a few seconds, while orientation discrimination is little affected for delays up to 20s or more (Lee & Harris, 1996; Magnussen & Greenlee, 1999). Having to retain estimates of multiple contrasts simultaneously, as in the present task, would presumably burden visual short-term memory to a greater extent, perhaps leading to noisier estimates or poorer separation of individual estimates in memory.

In fact, when we consider the present results alongside several other recent findings (Pestilli et al., 2011; Sergent et al., 2011; Xing et al., 2014), it is difficult to escape the following conclusion: observers appear ill-equipped to represent and store more than a single contrast estimate at a time (Pestilli et al., 2011; Xing et al., 2014); yet, multiple orientation estimates at a time can be easily stored and inspected from memory (Sergent et al., 2011). Such striking differences in short-term memory performance are presumably underpinned by more fundamental encoding or VSTM differences for the two types of sensory information (Xing et al., 2014). In the following chapter, we will fit a neural population model to the data from these experiments, in an effort to better understand the likely source of the differences.

In conclusion, we investigated the task-dependent nature of sensory selection, by testing the effects of large sensory responses on observer performance in two standard visual-discrimination tasks. It is well accepted that individual neurons involved in decision-making likely receive inputs from sensory neurons with widespread retinotopic locations and feature selectivities. In line with recent experimental findings (Chen & Seidemann, 2012; Pestilli et al., 2011), we found evidence that selection during contrast discrimination is severely disrupted by the presence of large sensory responses elsewhere in the visual field. In judging orientation changes, however, observers appeared to encode and maintain the relevant information more precisely in visual short-term memory, with high-contrast distractors having comparatively weaker effects on decision.

3

Efficient and inefficient selection from the same sensory neural response: computational model

3.1 INTRODUCTION

DISENTANGLING THE ROLES of sensory, memory and decision-related factors on psychophysical performance is an inherently convoluted exercise, requiring tightly controlled experiments alongside equally well-formulated models. Key insights on the nature of stimulus encoding and decoding have been gained, for example, by fitting precise neural computational models to data from two-alternative, forced-choice (2-AFC) discrimination tasks (Chirimuuta & Tolhurst, 2005; May & Solomon, 2015; Seriès et al., 2009). As a metric of discrimination behavior, population model approaches have often utilized the concept of Fisher Information (FI), a measure of the best possible decoding precision obtainable by an unbiased estimator (Dayan & Abbott, 2001; Paradiso, 1988; Seung & Sompolinsky, 1993). By providing a precise bound on the decoding precision possible for a given encoding model, FI-based models have shed light on key factors limiting decoding performance in sensory discrimination tasks, such as neural adaptation (Seriès et al., 2009), noise correla-

tions (Averbeck & Lee, 2006; Ecker et al., 2011), and stimulus priors (Ganguli & Simoncelli, 2014).

The computations underpinning sensory discrimination for single, isolated stimuli are now relatively well explored, with the existence of a number of well-developed decoding models for stimulus features such as contrast and orientation (Graf et al., 2011; May & Solomon, 2015; Paradiso, 1988; Sanborn & Dayan, 2011). Decoding performance in the context of target uncertainty, however, is still a very poorly understood problem, with some of the most popular decoding models suffering from notable deficiencies (Ma et al., 2015). In max-pooling models (Palmer et al., 2000; Pelli, 1985; Pestilli et al., 2011), for example, selection of the maximum of a set of neural responses acts as a coarse proxy for decision, neglecting the possibility that more graded stimulus interactions modulate decoding performance in systematic ways e.g., via attentional or sensory gain fluctuations (Ecker et al., 2015; Itthipuripat et al., 2014; Reynolds & Heeger, 2009). These problems are compounded by the fact that, for a given type of stimulus manipulation (e.g., increased set-size, increased distractor salience, etc.), the effects on behavioral performance are likely to be task-dependent. In sum, there is a need for combined empirical and model approaches that more readily explore the interactions between multiple stimuli, across potentially multiple feature dimensions.

Some recent efforts have been made in this regard; yet, these have either been purely theoretical in nature (Orhan & Ma, 2015), or have been focused on the stimulus estimation paradigm (Matthey et al., 2015). For example, numerous investigations have illustrated how neural responses to a given stimulus are often well-approximated by a weighted sum of neural responses to the individual, constituent features of that stimulus (Busse et al., 2009;

Recanzone et al., 1997; Zoccolan et al., 2007). For example, Busse et al. (2009) found that population responses in cat V1 to superimposed gratings were captured by a model that implemented a weighted averaging of the responses to each component grating separately. A recent theoretical investigation suggests that this type of neural response mixing may be substantially more detrimental to decoding performance than sizable decreases in response gain or noise amplitude (Orhan & Ma, 2015). Thus, a linear mixing framework may be particularly appropriate for understanding the strength of distractor effects in different stimulus contexts and tasks, and given the right formal decoding approach (i.e., FI), might be readily studied across multiple decision spaces simultaneously.

In Chapter 2, we systematically tested the effects of distractor contrast on the discrimination of changes to the contrast or orientation of a target stimulus. Contrast-discrimination performance was severely disrupted when high-contrast distractors appeared in the opposite visual hemifield to the target, while disruption of orientation-discrimination performance was more graded in magnitude. In the present chapter, we develop a computational understanding of the nature of these stimulus interactions. To do so, we constructed an FI-based model for the simultaneous estimation of contrast- and orientation-discrimination thresholds from an idealized neural population. We validated the form of our encoding model and the relevance of the FI approach by first fitting the model to single-stimulus discrimination data collected in a separate auxiliary experiment. An encoding model made up of multiple neural subpopulations with heterogeneous parameters was essential to successfully replicate the single-stimulus threshold data. The model was then extended to allow for interactions between two stimuli, using a generalized form of divisive normalization that allowed for linear mixing of separate sensory neural responses (Orhan & Ma, 2015). The

model provided quantitatively good simultaneous fits to the data sets, and for the tuning parameterizations we utilized at least, out-performed (i.e., had smaller mean RMSE) than models with either linear mixing or divisive normalization alone. Formal model comparison will hopefully distinguish between these models further, as well as testing of a broader range of possible tuning parameterizations.

3.2 MATERIALS AND METHODS

3.2.1 BACKGROUND

In Chapter 2, we measured contrast- and orientation-discrimination thresholds for target stimuli presented alongside an irrelevant distractor. The cornerstone of our modeling efforts in this chapter is the computation of FI in an idealized sensory neural population that simultaneously encodes contrast and orientation. Decoding precision is the inverse of the square of the discrimination threshold (Seung & Sompolinsky, 1993). FI is a measure of the best possible precision achievable by an unbiased decoder.

Our approach consisted of two stages. First, we validated the FI approach by fitting single-stimulus contrast- and orientation-discrimination thresholds collected in a separate, auxiliary experiment. This allowed us to understand the basic encoding model components necessary for fitting these types of discrimination data simultaneously. The basic model architecture we define at this stage resembles one other recent approach (May & Solomon, 2015). We then adapted this conceptual framework to our goal of accounting for the distractor effects described in Chapter 2. To do so, we allowed for interactions between spatially separated target and distractor stimuli through a generalized form of divisive normalization with linear mixing of the separate sensory neural responses.

3.2.2 NEURAL MODEL: DISCRIMINATION THRESHOLDS FOR ISOLATED STIMULI

FISHER INFORMATION

We consider the problem of simultaneously encoding contrast c and orientation s in a population of independent Poisson-like neurons. We assume that the mean response of the i -th neuron to (c, s) is given by:

$$f_i(c, s) = g_i(c)b_i(s), \quad (3.1)$$

where $g_i(c)$ controls the contrast response of the neuron and $b_i(s)$ determines the neural tuning to stimulus orientation. For the contrast response function, $g_i(c)$, we select a monotonic equation of Naka-Rushton form (Naka & Rushton, 1966),

$$g_i(c) = \frac{c^{n_i}}{c^{n_i} + \alpha_i^{n_i}}. \quad (3.2)$$

The responsiveness of the neuron to contrast is governed by the exponent, n_i , and the semi-saturation contrast, α_i . For orientation tuning curves, we select a population of homogeneous von Mises functions,

$$b_i(s) = \beta_i \exp(\gamma_i(\cos(s - s_i) - 1)) \quad (3.3)$$

where the parameters β_i and γ_i determine the response gain and the concentration parameter (narrowness) of the tuning curve. We assume independent Poisson noise:

$$p(\mathbf{r}|c, s) = \prod_i p(r_i|c, s) \quad (3.4)$$

$$= \prod_i \frac{1}{r_i!} e^{-f_i(c, s)} f_i(c, s)^{r_i}. \quad (3.5)$$

To compute model thresholds, we first need to compute the Fisher information matrix (FIM):

$$\mathbf{I}(c, s) = \begin{pmatrix} I_{cc} & I_{cs} \\ I_{cs} & I_{ss} \end{pmatrix}. \quad (3.6)$$

This 2x2 matrix has four components: I_{cc} and I_{ss} on the diagonal, and I_{cs} off-diagonal. The components of the FIM are computed in the standard way (Dayan & Abbott, 2001). For example,

$$I_{cc} = - \left\langle \frac{\partial^2 \log p(\mathbf{r}|c, s)}{\partial c^2} \right\rangle \quad (3.7)$$

$$= - \sum_i \left\langle \frac{\partial^2 \log p(r_i|c, s)}{\partial c^2} \right\rangle, \quad (3.8)$$

where $\langle \cdot \rangle$ is the expected value under $p(\mathbf{r}|c, s)$ and we used the independence assumptions.

This can be evaluated as

$$I_{cc} = \sum_i \frac{\left(\frac{\partial f_i(c, s)}{\partial c} \right)^2}{f_i(c, s)}. \quad (3.9)$$

Similarly,

$$I_{ss} = \sum_i \frac{\left(\frac{\partial f_i(c,s)}{\partial s}\right)^2}{f_i(c,s)} \quad (3.10)$$

$$I_{cs} = \sum_i \frac{\frac{\partial f_i(c,s)}{\partial c} \frac{\partial f_i(c,s)}{\partial s}}{f_i(c,s)} \quad (3.11)$$

We can re-write the three FIM components as:

$$I_{cc} = \sum_i h_i(s) \frac{g_i'(c)^2}{g_i(c)} \quad (3.12)$$

$$I_{ss} = \sum_i g_i(c) \frac{h_i'(s)^2}{h_i(s)} \quad (3.13)$$

$$I_{cs} = \sum_i g_i'(c) h_i'(s). \quad (3.14)$$

Per the Cramér-Rao bound (Cover & Thomas, 1991), the covariance matrix of an optimal estimator is the inverse of the FIM:

$$\Sigma(c,s) = \mathbf{I}(c,s)^{-1} = \frac{1}{I_{cc}I_{ss} - I_{cs}^2} \begin{pmatrix} I_{cc} & -I_{cs} \\ -I_{cs} & I_{ss} \end{pmatrix}. \quad (3.15)$$

The predicted thresholds for contrast and orientation are the square roots of the diagonal elements of the covariance matrix:

$$\sigma_c(c,s) = \sqrt{\frac{I_{cc}}{I_{cc}I_{ss} - I_{cs}^2}}; \quad (3.16)$$

$$\sigma_s(c,s) = \sqrt{\frac{I_{ss}}{I_{cc}I_{ss} - I_{cs}^2}}. \quad (3.17)$$

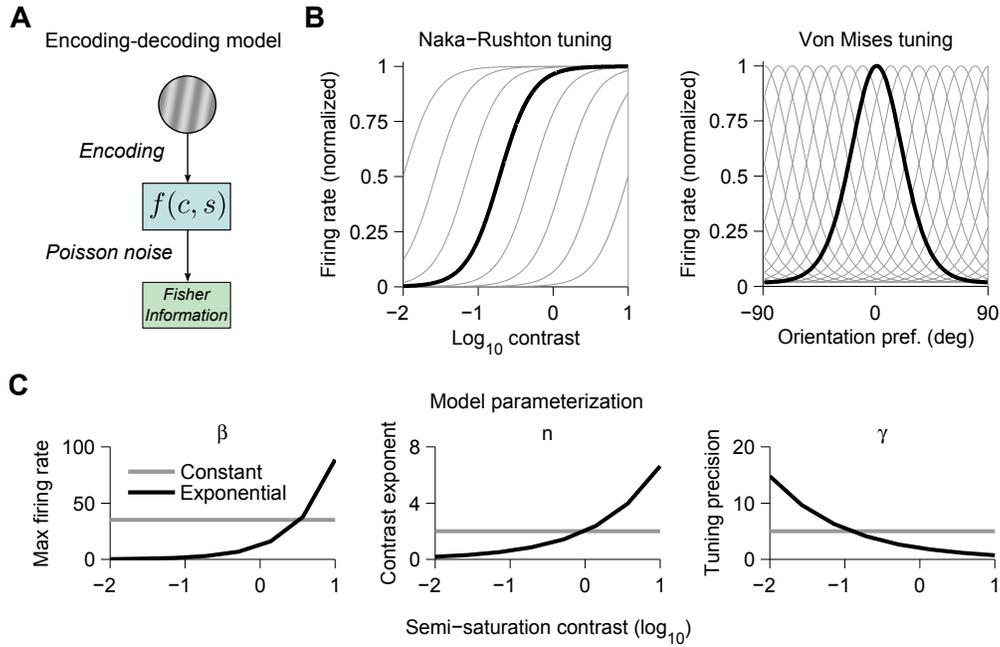


Figure 3.1: Validating the Fisher information approach. A) Encoding-decoding model. B) Illustration of idealized contrast- and orientation-tuning functions, similar to those used in the model. C) Tuning functions were parameterized in one of two ways: either held constant across all neurons, or allowed to vary with semi-saturation contrast according to exponential functions. The gradients illustrated are examples only, and not fits to real data. See [May & Solomon \(2015\)](#) for a similar approach, and Materials and Methods for details.

TUNING PROPERTIES

So far, we have computed optimal thresholds for an idealized population with defined contrast and orientation tuning. However, this tuning has free parameters: n_i , α_i , β_i , γ_i , and s_i . We now specify how we chose these parameters to tile the respective feature dimensions, adapting an approach recently taken by [May & Solomon \(2015\)](#) for similar model-fitting to discrimination thresholds. First, we took a population of 256 neurons, and divided these into 16 subpopulations of 16 neurons each. Within each subpopulation, the preferred orientations of the neurons were equally spaced in orientation space. The remaining parameters were held fixed within each subpopulation, but were allowed to vary across the sub-

populations with certain constraints. Specifically, the semi-saturation contrasts, α , were spaced in equal logarithmic steps between 0.01 and 10 (proportion contrast); values greater than 1 have been observed in neural recordings (Albrecht & Hamilton, 1982). For the remaining parameters we explored two scenarios:

- β , n , and γ held constant for all subpopulations
- β , n , and γ varying by subpopulation index j , according to independent, 2-parameter exponential functions.

These choices were motivated by a couple of desires. First, having a potentially graded scaling of maximum firing rate for each subpopulation seemed logical, as in the normalized case, the idealized contrast tuning functions naturally asymptote for 100% contrast stimuli at different response amplitudes (note, this may not be obvious when plotted according to log-base10 contrast). Analogous approaches have been used previously (Chirimuuta & Tolhurst, 2005; May & Solomon, 2015). Second, as has been documented a number of times, model fits to contrast-discrimination thresholds seem to require an expansive, population-level contrast response, to account for the flattening or dip in thresholds at very high-contrasts (Chirimuuta & Tolhurst, 2005; Kingdom & Whittle, 1996; May & Solomon, 2015; Zenger-Landolt & Heeger, 2003). From a technical perspective, this could be incorporated in a number of ways; we chose an approach where n can increase in magnitude with subpopulation index j . Finally, the consistent finding of floor effects in orientation-discrimination data suggests that neurons with moderate-to-high semi-saturation contrasts provide no additional performance benefit for orientation discrimination. To promote this type of behavior in our model, we allowed for a grading of sensitivity to orientation

as a function of semi-saturation contrast. The two model variants above had 3 and 6 free parameters, respectively.

3.2.3 NEURAL MODEL: DISTRACTOR EFFECTS ON DISCRIMINATION THRESHOLDS

LINEAR MIXING OF NORMALIZED NEURAL RESPONSES

We tested an FI model incorporating a generalized form of divisive normalization with linear mixing of separate, sensory neural responses (Carandini & Heeger, 2012; Orhan & Ma, 2015). In this model, we assume two neural populations encoding the contrasts and orientations of the target and distractor gratings. For neurons in the first group, the mean response is given by:

$$f_i(c_t, c_d, s_t, s_d) = w_1 \frac{c_t^{n_i}}{c_t^{n_i} + c_d^{n_i} + \alpha_i^{n_i}} h_i(s_t) + w_2 \frac{c_d^{n_i}}{c_t^{n_i} + c_d^{n_i} + \alpha_i^{n_i}} h_i(s_d) \quad (3.18)$$

where c_t and c_d are the target and distractor contrasts, and s_t and s_d are the target and distractor orientations. For neurons in the first group, we assume $w_1 > w_2$, hence they primarily encode the target grating; for neurons in the other group, the weights are switched, so they primarily encode the distractor. We further assume $w_1 + w_2 = 1$. We can think of w_1 and w_2 as roughly capturing receptive field effects, for example, in some higher-level readout stage where spatial receptive fields are broad.

We selected this form of model for a number of reasons. First, divisive normalization is thought to be a canonical computation in sensory and neural systems (Carandini & Heeger, 2012; Heeger, 1992), and may reflect the broader feedforward inhibition typically found in neural circuits. In the present context, it could arise through rapid fluctuations

in attentional gain (i.e., with the individual stimuli acting as competing exogenous attentional cues). In addition, linear mixing has been observed in numerous visual computations (Busse et al., 2009; Recanzone et al., 1997; Zoccolan et al., 2007), and is an appropriate choice of model where graded stimulus interactions are concerned (Orhan & Ma, 2015). Linear mixing might also capture behaviors similar to max-pooling (i.e., for high-contrast distractor conditions).

In addition to the full linear mixing model, we also tested versions of the model in which there was only divisive normalization or only linear mixing. For the divisive normalization model, we set the w term to 1 and allowed a flexible weight term, k , to be applied to the distractor in the equation denominator. For the linear mixing model without divisive normalization, we simply removed the relevant denominator terms from each half of the mixing expression.

As with the earlier model, we assumed Poisson-like noise. We simulated the model performance by computing all sixteen entries of the 4×4 FIM according to:

$$I_{xy}(s_t, s_d, c_t, c_d) = \sum_i \frac{\frac{\partial f_i}{\partial x} \frac{\partial f_i}{\partial y}}{f_i(s_t, s_d, c_t, c_d)} \quad (3.19)$$

where the x and y pair takes on each possible combination of the indices: (c_t, c_d, s_t, s_d) . After computing these terms (summed across sub-population), we then estimated thresholds by inverting the FIM, and taking the square-root of the resulting diagonal variance terms as before.

TUNING PROPERTIES

We adopted the same expansive model architecture as used successfully to fit the single-stimulus discrimination data; specifically, we allowed for 16 neural subpopulations whose semi-saturation contrasts again tiled the contrast axis logarithmically. Maximum firing rate, contrast-response exponent, and orientation tuning width could again vary across subpopulation with exponentially increasing or decreasing gradients. Thus there were a total of seven free parameters for the full linear mixing model (w and six exponential tuning parameters), and also seven each for the linear mixing without divisive normalization model (same parameters) and divisive normalization only model (same tuning parameters, but with k replacing w).

3.2.4 MODEL FITTING

In both model phases, we fit models to contrast- and orientation-discrimination thresholds simultaneously (i.e., constrained by the measured thresholds from both experiments), searching for the set of parameters that gave the smallest combined RMSE relative to observed thresholds. To calculate combined RMSE, observed thresholds were converted to proportions (0-1) and radians respectively, and the squared error relative to model predictions was calculated for each test condition before taking the mean and square-root. For the single-stimulus discrimination data, this included ten conditions. In fitting data for the main experiments, we decided to further constrain the model by including the single-stimulus (i.e., 0% contrast distractor) thresholds. Observer groups were only partially overlapping for the separate study phases; thus, we included the actual measured single-stimulus discrimination thresholds for those observers who ran in both phases of the study ($n = 3$),

and substituted the mean single-stimulus discrimination thresholds for the remaining observers (while this is not ideal, considering the consistency of TvC shape across observers for the single-stimulus experiments, we are not concerned about this approach). In total, we calculated RMSE across thirty-eight conditions in total (10 conditions from the single-stimulus experiments, and 12 contrast and 16 orientation conditions from the main experiments).

In all model-fitting, we used an evolutionary search algorithm known as Covariance Matrix Adaptation (CMA-ES) (Hansen & Ostermeier, 1996). This is a very robust, cutting-edge optimization algorithm that makes few assumptions regarding the nature of the function being optimized. We implemented CMA-ES using a freely available Matlab function version (*cmaes.m*, available at www.lri.fr/~hansen/cmaesintro). As a stochastic algorithm, CMA-ES gives different output depending on the random seed used for model initiation. Thus, for all model fits, we ran multiple separate searches [pending] using the high-performance computing (HPC) cluster at New York University, each with a different set of starting parameter values for the optimization. Best-fit parameters for a particular observer/model were taken from the model run that gave the smallest RMSE out of all runs. Individual parameters were allowed to vary within broad but finite bounds as follows: the scale of the three exponential gradients (or constant in the case of the constant parameterization fits to single-stimulus data) were: β : 1-100; n : 1-10; γ : 1-10. The power of each exponential gradient could vary from -5;5, thus allowing for potentially expansive or decaying behavior in each case. Remaining parameters could range as follows: w (linear mixing): 0.5-1; k (divisive normalization): 0-1. Optimization starting values were randomly drawn from within the 25th-75th percentile ranges for each of these parameter ranges, and the CMA-ES

search width parameter (i.e., σ) was set to one-third of the respective parameter range.

3.2.5 AUXILIARY EXPERIMENT

Data collection for the main experiments is described in detail in Chapter 2. Below, a brief description is given for an auxiliary experiment measuring single-stimulus discrimination thresholds. This was carried out with the aim of appropriately constraining computational models of performance in the main experiments, and is thus described here. The general methods and informed consent procedures were largely identical to those described in Chapter 2.

In the auxiliary experiment, we measured contrast- and orientation-discrimination thresholds for isolated targets, using a largely identical set-up to the main experiments (with distractor contrast now set to 0% contrast). Data were collected from seven observers (in two or three separate sessions each), including three (two authors) who completed the main experiments. Distributed cues (i.e., white arrows) were presented before and during the stimulus intervals, such that observers did not know in advance on which side the stimulus would appear (although the target did appear on the same side for both intervals, so we can assume the second interval location was known). A post-cue again indicated target location. As a primary goal of this experiment was to collect data that might validate our model approach, we included one additional low-contrast target condition (2%), so as to better measure the dynamic range of discrimination thresholds. Attempts to measure performance at much lower pedestal contrast were abandoned, as we felt that orientation discrimination at much lower contrast and with spatial uncertainty amounted to a form of stimulus detection. Thresholds were estimated using procedures similar to those described in Chapter

2.

3.3 RESULTS

FISHER INFORMATION-BASED MODEL OF DISCRIMINATION THRESHOLDS

An important aspect of our experimental design was the use of identical stimulus parameters across experiment, varying only the task performed by the observer. By assuming that identical sensory neural responses are evoked across experiment, characteristics of how the individual sensory neural responses interact and are decoded must underlie the threshold differences we observed in Chapter 2. To formalize this approach to fitting the data, we developed an encoding-decoding model based on Fisher Information (FI), for the simultaneous estimation of contrast- and orientation-discrimination thresholds from an isolated sensory neural response (Figure 3.1). In doing so, we make the simplifying assumption that thresholds are inversely related to the precision of the decoder, and exploit the fact that FI provides a measure of this precision. Specifically, subject to the Cramér-Rao bound, Fisher information sets a lower limit on the accuracy with which the true stimulus value can be decoded by any unbiased estimator (Cover & Thomas, 1991; Seung & Sompolinsky, 1993). We validated the form of this model by first fitting it to single-stimulus contrast- and orientation-discrimination thresholds collected in a separate auxiliary experiment (Figure 3.2). Model fits were doubly constrained by fitting simultaneously on both datasets (i.e., parameters were held fixed across tasks).

For isolated sensory neural responses, a model with tuning sensitivity parameters held fixed across neural subpopulations failed to satisfactorily fit real data: mean RMSE = 0.0230 (0.0022) (Figure 3.2A). This can be easily understood by examining the single-stimulus,

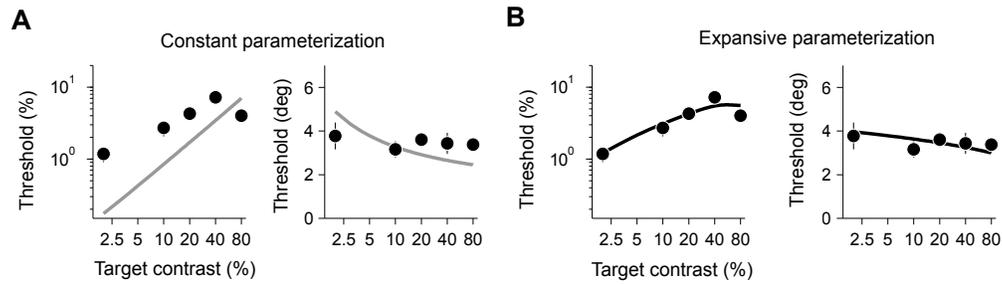


Figure 3.2: Model fits to single-stimulus discrimination data. Contrast- and orientation-discrimination thresholds were collected in an auxiliary experiment. Mean and SE of the observed thresholds ($n = 7$) are plotted in circles. Model fits to the mean thresholds are represented by the lines.

contrast-discrimination thresholds, which exhibit a slope substantially less than 1 and a pointed decrease in thresholds for very high-contrast pedestals. By spacing model neurons such that they had logarithmically-spaced semi-saturation contrasts, while keeping all other parameters constant, idealized model behavior is perfectly Weber-like, with predicted thresholds having a slope of 1 on a log-log axis (May & Solomon, 2015). Thus, some heterogeneity is required in the model architecture to fit real data for even single-stimulus based tasks.

To build greater flexibility into the encoding model, we adapted an approach taken recently by May & Solomon (2015). In this investigation, the authors accounted for the near-miss to Weber's Law by allowing for an exponential parameterization governing the contrast sensitivity of the population (May & Solomon, 2015). We incorporated this feature into our model, thus allowing for an expansiveness to the population-level contrast responses. Preliminary model attempts suggested that flexibility of this sort was essential to fit the relatively size-able late decrease in contrast-discrimination thresholds, a finding which has now been observed a number of times before and fit using qualitatively similar approaches (Chirimuuta & Tolhurst, 2005; Kingdom & Whittle, 1996; Pestilli et al., 2011;

Zenger-Landolt & Heeger, 2003). In addition, we chose to be somewhat agnostic regarding each feature dimension, and thus allowed the relevant sensitivity parameters (n and γ), as well as parameter governing the maximum firing rate of each subpopulation (β), to each vary according to independent 2-parameter exponential functions. In this way, it is the gradients of sensitivity across subpopulation that define important aspects of the overall model behavior.

The flexible, heterogeneous FI model provided much better simultaneous fits to the single-stimulus threshold data: mean RMSE = 0.0116 (0.0011) (Figure 3.2B). Note that while the late dip is not precisely fit here, this can be easily accommodated by allowing for flexibility in the number of subpopulations or on the bounds on the semi-saturation contrasts. For the present purposes, we simply wanted to develop a tuning function parameterization that more closely replicated the single-stimulus data, while not allowing for too much redundancy.

LINEAR MIXING OF NORMALIZED NEURAL RESPONSES

The FI approach tested thus far is inherently local in application. Thus, conceptually at least, it is not be the appropriate level at which to implement models where target and distractor interact across broadly spaced locations and across two feature dimensions. To allow for either suppressive or driving effects across each of the two feature dimensions, we need a single model architecture in which both sets of tuning functions can interact in all possible reasonable ways. To achieve this, we adopted the linear mixing approach described recently (Orhan & Ma, 2015), in which FI is computed from a neural population that receives mixed inputs. In this scheme, the readout neural population might be some higher-order sensory

or decision-related neural population, where broader spatial receptive fields are likely. Such a scheme might allow for more graded and realistic distractor effects to emerge, while still qualitatively allowing for behavior not unlike more extreme models (i.e., max-pooling). In fact, model fits from one recent contrast-discrimination study were qualitatively analogous to a linear mixing of target and distractor responses (Itthipuripat et al., 2014). The present approach, however, is computationally more rigorous, closed-form, and allows for simultaneous decoding (and constraint) across two orthogonal decision spaces.

We developed a linear mixing model of stimulus encoding for the experiments described in Chapter 2, and calculated FI in the simulated population by directly computing the inverse of the population covariance matrix (see Materials and Methods). We compared three different models: a full linear mixing model (with divisive normalization), a linear mixing model without divisive normalization, and a model with divisive normalization only (no linear mixing). We utilized the same model architecture as for the single-stimulus model fits, thus allowing for a potentially expansive population-level contrast response, and potential decay in orientation-tuning sensitivity as a function of subpopulation semi-saturation contrast.

Neither linear mixing nor divisive normalization alone provided quantitatively good fits to data as a whole, or to the high-contrast distractor conditions in the contrast experiment (Table 3.1). The linear mixing model alone, a weighted sum of independent Naka-Rushton (Naka & Rushton, 1966) sensory neural responses, failed to capture the shape of the two-stimulus data in general: mean RMSE = 0.0401(0.0039) (Figure 3.3A). This is likely due to the model being essentially unable to reproduce the observed orientation threshold increase with increasing distractor contrast (i.e., the larger summed gain predicts increased,

RMSE of model fits to the two-stimulus experiments of Chapter 2

	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	Mean(SE)
<i>Model</i>									
LM	.0519	.0404	.0417	.0562	.0344	.0221	.0302	.0247	.0401(.0039)
DN	.0388	.0321	.0398	.0370	.0496	.0353	.0225	.0380	.0366(.0027)
LMDN	.0379	.0238	.0373	.0302	.0484	.0315	.0214	.0284	.0324(.0031)

Table 3.1: RMSE values for each observer from the best simultaneous fits to the data from both experiments. LM = linear mixing model; DN = divisive normalization model; LMDN = both linear mixing and divisive normalization.

not decreased precision). In contrast, a model with divisive normalization alone could replicate the shape of orientation-discrimination data, and was overall better at simultaneously fitting both data sets: mean RMSE = 0.0366(0.0027) (Figure 3.3B).

While neither linear mixing nor divisive normalization alone provided completely satisfactory fits of the data, a combination of both forms of sensory interaction did provide a better fit, with the generalized linear mixing model (with divisive normalization) giving the smallest mean RMSE of the three models: mean RMSE = 0.0324 (0.0031) (Figure 3.3C). Specifically, this model better replicated the large increase in contrast-discrimination thresholds with high-contrast distractors than either of the other two models. However, the magnitude of the improvement relative to the divisive normalization model alone was relatively marginal.

For a number of reasons, however, these model results are still unsatisfactory in many regards. First, we can not rule out the possibility that with the addition of extra flexibility, the divisive normalization model alone might fit the data sets better, and in some preliminary model explorations we have found some tentative evidence of this (specifically, by allowing the upper-bound on the semi-saturation contrast range to be a free-parameter itself).

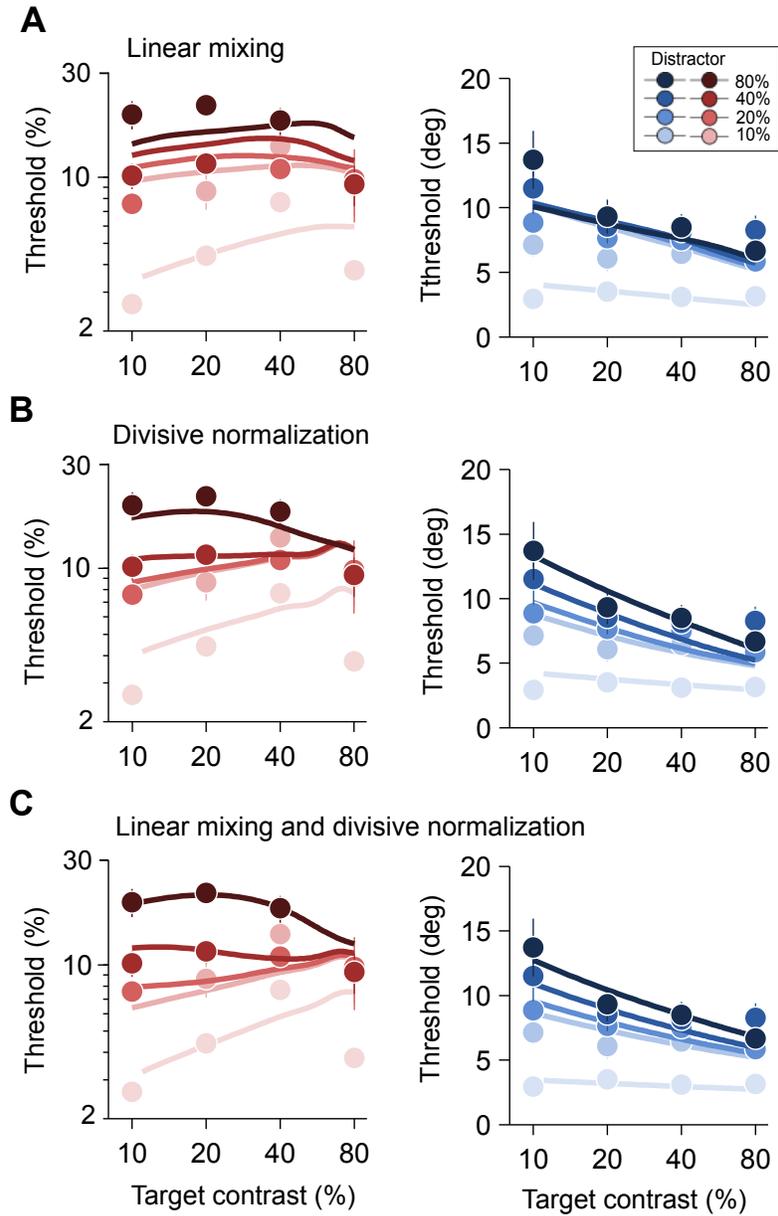


Figure 3.3: Model fits to main experiments. A) Fits of a linear mixing model with Naka-Rushton gain terms for each of the separate neural responses. B) Fits of a divisive normalization model only. C) Fits of a generalized linear mixing model with divisive normalization. See text for details. Note the lowest (2%) target condition for the no distractor trials are not plotted for the sake of visualization, although were included in fitting.

Second, the issue of allowing for either local (target-centered model neurons, 2×2 FIM) or global read-out (symmetric target- and distractor-centered model neurons, 4×4 FIM) itself may be important: global readout is the natural decoding approach for the linear-mixing model (Orhan & Ma, 2015), however it is not obvious whether this should be the case for a model with only divisive normalization. Third, it is possible that alternative forms of tuning function heterogeneity might provide better grounding for model comparison. However, given the various complexities of model-fitting across two decision spaces, we feel the current approach is likely as good as any other. Regardless of model approach, we still lack the underlying neural response measurements that would be needed for definitive selection of the most realistic encoding model.

Considering again the psychophysical results in Chapter 2, however, there is at least one hint regarding what needs to be added to our model approach. In our secondary analysis of the psychophysical data there, we tested whether the relative differences in orientation played any obvious role in threshold performance. For orientation-discrimination, we found a small, but consistent threshold difference dependent on the orientation difference between target and distractor. In contrast, the FI-based linear mixing model makes the opposite, counter-intuitive prediction (Orhan & Ma, 2015): distractor orientations near to the target orientation disrupt performance more significantly than orientations far from the target. Thus, our orientation-threshold data suggest some sub-optimal decoding on the part of observers (and arguably too, so do the large distractor contrast effects). Further model work will need to investigate whether linear decoders paired with Monte Carlo simulation, for example, can provide more convincing fits to the data sets.

3.4 DISCUSSION

Computational models of observer performance are essential components in any full understanding of task-specific psychophysical behavior. In Chapter 3, we developed a model of the psychophysical performance described in the prior chapter, a model based on linear mixing of divisively-normalized sensory neural responses. While the most flexible model (including both divisive normalization and linear mixing) achieved reasonable fits to the high-contrast distractor effect, overall the fits were not satisfactory; in addition, neither the divisive normalization nor the linear mixing models alone provided good overall fits to the data sets as a whole.

These results speak directly to recent findings on the computations underlying sensory selection and attention-based improvements in behavior (Chen & Seidemann, 2012; Itthipuripat et al., 2014; Pestilli et al., 2011). First, our results highlight the limited scope of several prior findings on attentional selection. Improvements in behavior mediated by attention likely occur via a combination of sensory gain changes and selection strategies, the exact nature of which may change drastically across behavioral task despite identical sensory input. For example, max-pooling rules act essentially as a coarse proxy for decision, leaving unspecified the sensory effects of distracting stimuli (Pestilli et al., 2011). Thus, max-pooling rules likely only perform adequately where distractor conditions are not rigidly defined or where pooling occurs over large numbers of stimuli (Pestilli et al., 2011; Chen & Seidemann, 2012). In contrast, parametric variation of stimulus contrasts using two-stimulus displays, for example, may result in more graded behavioral effects, with complementary roles played by response suppression (i.e., attentional gain fluctuations) and response mixing.

This might explain the better performance of our generalized linear mixing model (with divisive normalization), as well as other recent findings where extreme max-pooling appeared not to be essential (Itthipuripat et al., 2014). However, we failed to find a truly satisfactory fit of both datasets using even this model.

FI-based models of sensory neural responses are an improvement on models that reduce to inherently sub-optimal winner-take-all decoding. In addition, linear mixing seems like a natural way to model the interaction of sensory responses at some higher-order, integrative stage of processing (Orhan & Ma, 2015). Nevertheless, limitations of our model-fitting must also be acknowledged. First, fits were not constrained by simultaneous neural response measurements, unlike in some prior related investigations (Pestilli et al., 2011; Itthipuripat et al., 2014). Second, we found it necessary to allow for an expansive population-level contrast response in model fits, as well as some semi-saturation contrast-dependent orientation tuning. The question of their being biological equivalents to these model components is an awkward one: while there is some evidence of broadly divergent contrast sensitivities across M and P-channels (Kaplan & Shapley, 1986; Shapley, 1990), for example, whether the late contrast threshold decrease we observed relates directly to some distinct neural source (i.e., the response of some subpopulation of neurons with high semi-saturation contrast) is unknown. In addition, our allowance for flexible orientation tuning as a function of semi-saturation contrast was guided purely by threshold data i.e., the typically flat shape of orientation-discrimination threshold data, which might reflect some upper bound on the discrimination precision possible for a given retinal eccentricity (Mareschal & Shapley, 2004; Skottun et al., 1987). Nevertheless, our design choices seemed like reasonable places to start, and allow us to remain somewhat agnostic as to the possible underlying

neural properties involved.

In deriving the Fisher information-based expressions governing threshold behavior, we also made several simplifying assumptions. We chose sub-populations of set size and with homogeneous orientation-tuning functions, and we assumed uniform noise correlation (set to zero) across the neural population. The possibility remains that some variant combination of gain, tuning function and neural noise parameters provides an alternative explanation of our data-sets. For example, our assumptions about the form of neural noise may be limited in their validity, ignoring the large role now thought to be played by modulatory signals in setting the overall amplitude of sensory noise (Goris et al., 2014). While certainly over-simplistic, our assumption of zero correlation across neurons was a reasonable place to begin our model fitting. There is still much debate about how correlations modulate decoding performance, and about the conditions that determine whether increased correlations lead to facilitation or disruption of decoding (Abbott & Dayan, 1999; Ecker et al., 2011). Emphasis on such details here, however, would have obscured the main goal of testing a general model of interaction between separate sensory neural responses.

In conclusion, we studied the behavior of FI-based models of threshold performance, as simultaneously applied to data from separate contrast- and orientation-discrimination experiments described in Chapter 2. We first validated this approach by fitting model expressions to single-stimulus discrimination data collected in auxiliary experiments. We then extended the model to allow for the linear mixing or divisive normalization of neural responses from widely separated target and distractor spatial locations. Neither sensory interaction model alone provided a quantitatively convincing fit to the datasets, although a model incorporating both normalization and mixing provided quantitatively reasonable

simultaneous fits to the high-contrast distractor effects at least. Future model investigations will seek to improve on these fits, although for the present purposes we can at least say that two common models of sensory interaction failed in fitting the data from Chapter 2.

4

Delayed estimation of luminance contrast

4.1 INTRODUCTION

For many years, the contents of VSTM were conceptualized as discrete: an item was either in memory or it was not (Cowan, 2001; Luck & Vogel, 1997). From the point of view of perceptual psychophysics, this view is simplistic to the point of being untenable. In signal detection theory models of threshold detection and discrimination tasks, the internal rep-

resentation of a stimulus is taken to be a noise-corrupted version of that stimulus; thus, an item can be encoded to a greater or lesser degree, depending on the amount of noise. In the study of VSTM, it has taken a long time for the concept of a memory being a noisy version of the stimulus to take hold. Cognitive psychology studies of VSTM typically use coarse stimuli without parametric variation, e.g., change detection among items that were hand-picked with the goal of making them highly discriminable (Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988). In contrast, researchers working in the tradition of threshold psychophysics have long since adopted the idea of noisy memories, a concept implicit, for example, in paradigms that measure the magnitude of stimulus change necessary for some criterion level of discrimination performance (Magnussen & Greenlee, 1999; Palmer, 1990).

Recently the concept of noisy memories has gained ground in VSTM research due to the introduction of a new paradigm of probing VSTM, delayed estimation (Bays, 2014; Fougne et al., 2012; van den Berg et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008). In this paradigm, inspired by earlier work by Prinzmetal and colleagues (Prinzmetal et al., 1997, 1998), the observer reports the identity of a remembered stimulus on a continuum, repeating this process over many trials to create a histogram of estimates. The width of the resulting estimate histogram can then be taken as a measure of the level of noise in the memory. Delayed estimation has been applied most notably to the study of VSTM for stimulus features such as orientation and color (Bays, 2014; Fougne et al., 2012; van den Berg et al., 2012). For example, recent studies have quantified the dependence of noise level on set size (Bays et al., 2009; Fougne et al., 2012; van den Berg et al., 2012), and attempted to determine whether there is an upper limit on the number of remembered items (van den Berg et al., 2012; Zhang & Luck, 2008).

While the vast majority of delayed-estimation papers have studied orientation and color, it cannot be assumed that these two features are representative of all features stored in VSTM. For example, the neural representations of orientation and color rely on very specific neural substrates: topographically-arranged maps in the case of orientation (Ferster, 2003; Hubel & Wiesel, 1962), and specific color-opponent, retino-cortical pathways in the case of color (Brouwer & Heeger, 2009; Gegenfurtner & Kiper, 2003; Johnson et al., 2001; Lennie et al., 1990). If VSTM relies on the same neural networks responsible for initial sensory encoding, as some influential theories posit (Awh & Jonides, 2001; Jonides et al., 2008), then these structured neural representations provide an ideal substrate for relatively precise VSTM encoding and subsequent read-out. It is perhaps therefore not surprising that accurate maintenance of orientation and color information is possible over relatively long delays (Magnussen et al., 1996; Magnussen & Greenlee, 1999; Nilsson & Nelson, 1981).

In Chapter 2, we compared VSTM for two different stimulus features, luminance contrast and orientation, in the presence of a distractor stimulus. Our empirical findings, as well as several prior results, suggest that VSTM for luminance contrast might be fundamentally different from VSTM for color and orientation. First, luminance contrast is an intensity-coded variable (Albrecht & Hamilton, 1982), and thus lacks the precisely-structured neural representations that features like orientation are encoded by. Thus, luminance contrast may likely be encoded into VSTM in a much more abstract way than other features (Xing et al., 2014). Second, evidence exists suggesting that memory for luminance contrast is impoverished: for example, discrimination thresholds increase substantially with inter-stimulus delay (Magnussen et al., 1996; Magnussen & Greenlee, 1999), and in the presence of distractors, as we and others have previously found (Pestilli et al., 2011). Unfortu-

nately, our understanding of both the encoding and retention of luminance contrast information is based primarily on results of coarse, 2-AFC discrimination tasks (Legge & Foley, 1980; Magnussen & Greenlee, 1999; Nachmias & Sansbury, 1974). While one prior study did use delayed estimation for studying luminance contrast encoding, this study involved manipulations of attention and lacked the detailed parametric variation now common in delayed estimation studies (Prinzmetal et al., 1997).

These factors motivated us to examine VSTM for luminance contrast using the delayed-estimation paradigm. In Experiment 1, observers were instructed to hold in memory the perceived luminance contrast of a briefly flashed circular disc, and after a brief delay, to reconstruct the memorized contrast by adjusting the luminance of a subsequently presented match disc. We systematically measured estimate distributions for luminance contrasts spanning from low (6%) to high (76%). We then successfully described the shape of these distributions using a probabilistic model of neural responses, incorporating maximum-likelihood readout. Control experiments investigated how the distribution shapes depended on the onset contrast of the match disc and on the polarity of the stimuli.

4.2 MATERIALS AND METHODS

4.2.1 PARTICIPANTS

Data from eight observers (one author) were collected in Experiment 1. Observers were recruited from the local community and student body at New York University (paid \$10/hr), and amongst lab colleagues. Observers had varying degrees of experience in psychophysical testing. All observers gave written informed consent, and experiments were carried out with approval of the NYU University Committee on Activities Involving Human Subjects.

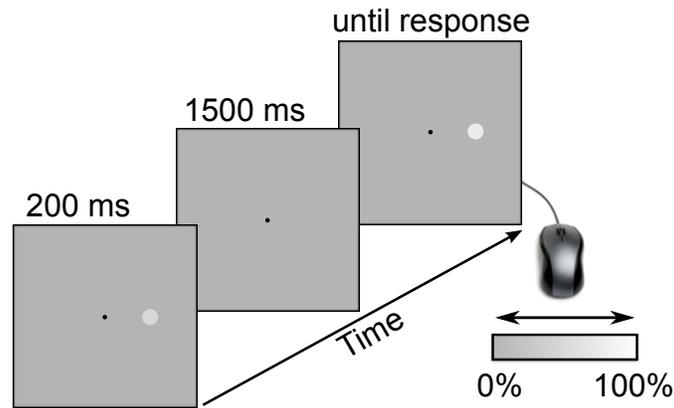


Figure 4.1: Delayed estimation of luminance contrast. Observers matched the luminance contrast of a briefly presented test disc, by adjusting the luminance contrast of a match disc using horizontal motions of a computer mouse.

4.2.2 EXPERIMENT I

TASK

An example trial is illustrated in Figure 4.1. On each trial of the experiment, observers were briefly presented (200 ms) with a small luminance-defined disc (1° diameter) on the computer monitor, either left or right (4° eccentricity) of the black fixation dot. On any trial, the disc appeared with one of eight luminance contrasts (see Test set-up and stimulus design), which was selected from a randomly-shuffled array within each block. After the first disc disappeared and following a brief delay (1500 ms), a second ‘match’ disc appeared at the same location as the first, this time with a luminance contrast chosen randomly from the range of thirty-eight possible estimate luminance contrasts (see Test set-up and stimulus design). Observers were required to estimate the luminance contrast of the first disc by adjusting (with the computer mouse) the luminance contrast of the match disc. Luminance contrast of the match disc was adjusted by making small horizontal motions of the mouse - leftward motions of the mouse made the disc appear dimmer, rightward motions of the

mouse made the disc appear brighter. Observers were instructed that there was no set interval for responding, and to try to perform as accurately as possible. After estimating the luminance contrast of the first disc as accurately as possible, observers pressed the left mouse button to record their estimate.

Observers completed four test sessions each (~ 1 hr per session). Each session consisted of five 80-trial blocks (10 trials per luminance contrast level), preceded by one 40-trial practice block. This gave a total of 1600 test trials per observer (i.e., 200 trials per luminance contrast). During each block, observers rested their chin on a chin-rest, and were instructed to maintain fixation on the central fixation dot throughout each trial (during presentation of the first disc and while adjusting the match disc). Observers also received limited motivational feedback only, after every second block (e.g., 'Well done! You are performing above average.' or 'Good job. Your performance level is around the median of all observers.'). The feedback statements above were alternately selected at random, and feedback was not related to any performance criterion per se.

TEST SET-UP AND STIMULUS DESIGN

Stimuli were presented in a darkened room on an iPad retina display (monitor only), controlled by a Windows-based PC running MATLAB (The Mathworks) and the Psychophysics Toolbox. Resolution of this small monitor was 2048 pixels x 1536 pixels. The display was controlled by an AbuseMarK LCD adapter and fixed in a custom-frame affording three degrees of freedom in monitor positioning. Before beginning each session, the experimenter ensured that the monitor was positioned centrally in front of the observer, with the fixation dot at eye height. The display was positioned in landscape mode (i.e., with the higher

resolution along the horizontal). Viewing distance for Experiment 1 was 28.5 cm.

Monitor brightness was maintained at its default maximum setting (which allows for maximum luminance values up to $\sim 390\text{-}400\text{ cd/m}^2$). To control gray levels appropriately, we first manually gamma-corrected the display across its full range of luminance output (using a Spectrascan PR650 photometer with uniform luminance across the monitor). We then created a reduced luminance range look-up table spanning one-quarter of the full luminance range (i.e., from $0\text{-}93.5\text{ cd/m}^2$), by selecting the first quarter of the gamma-corrected, full range look-up table and interpolating intermediate values to create a vector of 256 RGB intensity values. This reduced luminance range is comparable to ranges typically reported in studies where luminance contrast is manipulated (i.e., in CRT-based experiments). The reduced range, however, meant that nearby RGB indices overlapped somewhat in output luminance. Using the photometer, we manually measured the actual output luminance (with several repeats) for each RGB index in the reduced range. For the positive luminance deflections used in Experiment 1, the 129 RGB levels used (i.e., the background index of 127 and the 128 levels above) mapped onto thirty-eight unique luminance output values, which we used to naturally bin observers' estimates based on the mouse position-RGB index mapping.

We defined the luminance contrast of these thirty-eight unique levels in terms of Weber contrast,

$$c_{\text{weber}} = \frac{I_{\text{stimulus}} - I_{\text{background}}}{I_{\text{background}}} \quad (4.1)$$

where I_{stimulus} was the disc luminance and $I_{\text{background}}$ was the luminance of the gray background. The eight test luminance contrasts were set according to fixed RGB indices using a pseudo-linear increment array; specifically, RGB values of 127 + [8, 16, 24, 32, 48, 64, 80,

96] were used. Using the measured luminance values for these indices, the resulting luminance contrasts spanned from ~ 6 -76% contrast. The circular discs measured 1° in diameter, and were uniform in luminance, except for the very edge of the disc (raised-cosine, edge width 0.1°).

4.2.3 EXPERIMENTS 2 AND 3

We ran two control experiments ($n = 8$ observers each), in an effort to understand the possible roles played by a number of task and stimulus-related factors in Experiment 1. Participant recruitment and informed consent procedures followed similar protocols to those described above.

Experiment 2 was identical to Experiment 1, except that the match disc was set to 0% contrast at onset. Experiment 3 was identical to Experiment 1 except that dark luminance discs were presented on a gray background. We subtracted the array of RGB increments from mid-level gray to calculate each of the eight tested luminance contrasts. We used the same restricted gamma table (0-93.5 cd/m^2). For each RGB index below mid-level gray, we measured the actual output luminance value using the photometer, and confirmed that there were 76 unique luminance values in the 0-127 RGB range.

4.2.4 SUMMARY DATA ANALYSES

We calculated the median and inter-quartile range of each distribution for each observer separately. We then fit these data using a power law to ascertain whether data conformed approximately to Weber's Law. Specifically, we found the least-squares fit that best de-

scribed the data according to:

$$\hat{c}_{75} - \hat{c}_{25} = kc^w \quad (4.2)$$

where k scales the power law, with exponent w , relating luminance contrast of the disc to the width of the estimate distribution. Perfect Weber's Law behavior would give an exponent of one for this relation; in investigations of luminance contrast discrimination, best-fit slope values in the range of 0.5-0.7 have been typically reported (Legge & Foley, 1980; May & Solomon, 2015; Pestilli et al., 2011).

4.2.5 PROBABILISTIC MODEL OF NEURAL RESPONSES

MAXIMUM-LIKELIHOOD ESTIMATION OF CONTRAST

We begin by describing the generative model for our task (Figure 4.3). A stimulus of luminance contrast c is presented to the observer. The stimulus is encoded by a population of noisy sensory neurons, giving rise to a vector of spike counts $\mathbf{r} = \{r_1, r_2, r_3, \dots, r_n\}$. We assume that the observer then decodes this set of spike counts using maximum-likelihood estimation, to arrive at an estimate \hat{c} .

We assume that spike counts are independent across neurons and governed by Poisson noise. The probability of spike count vector \mathbf{r} , given luminance contrast c , is thus (Dayan & Abbott, 2001),

$$P(\mathbf{r}|c) = \prod_i \frac{1}{r_i!} e^{-g_i(c)} g_i(c)^{r_i}, \quad (4.3)$$

where $g_i(c)$ represents the contrast gain function for neuron i . We assume this gain function

takes the form of a Naka-Rushton equation (Naka & Rushton, 1966),

$$g_i(c) = a_i g(c) = a_i \frac{c^n}{c^n + c_{50}^n} \quad (4.4)$$

where the overall responsiveness to contrast is governed by the exponent, n , and the contrast gain term, c_{50} , and a_i is a vector of maximum firing rate values that might vary across neurons, but that average out in the population during read-out. Assuming the observer is performing maximum-likelihood estimation on the underlying firing rates, then the estimate contrast, \hat{c} , is given by

$$\hat{c} = \arg \max_c \log p(\mathbf{r}|c) \quad (4.5)$$

Some straightforward calculations show that,

$$\hat{c} = \arg \max_c \left(-g(c) \sum_i a_i + \log g(c) \sum_i r_i \right) \quad (4.6)$$

$$= g^{-1} \left(\frac{r}{a} \right), \quad (4.7)$$

where $a = \sum_i a_i$, $r = \sum_i r_i$, and g^{-1} is the inverse function of g . If we approximate the Poisson distribution by a normal distribution, then $r \sim \mathcal{N}(ag(c), ag(c))$ and

$$\frac{r}{a} \sim \mathcal{N} \left(g(c), \frac{g(c)}{a} \right). \quad (4.8)$$

By transforming this probability distribution under the mapping $\frac{r}{a} \mapsto g^{-1} \left(\frac{r}{a} \right)$, we obtain

the conditional probability of estimate contrast, \hat{c} , given c , as

$$P(\hat{c}|c) = \sqrt{\frac{a}{2\pi}} \frac{g'(\hat{c})}{\sqrt{g(c)}} e^{-\frac{a}{2} \frac{(g(\hat{c})-g(c))^2}{g(\hat{c})}} \quad (4.9)$$

MODEL FITTING AND COMPARISON

Using a maximum-likelihood procedure, we found the best-fitting parameter values of the model for each observer individually. In model fitting, we also allowed for a lapse rate parameter, λ . We assume that lapses were uniformly distributed across possible estimate values. Thus, there were four free parameters in total - three gain parameters (a , n , c_{50}) and λ . Model-fitting was done throughout using the MATLAB function *fminsearch.m*, which is a standard optimization algorithm based on the Nelder-Mead simplex algorithm, and suitable considering the relatively small number of free parameters. We allowed each of the parameters to vary within broad but finite bounds (see Table 4.1).

In model-fitting, we also had to make one additional modification to allow for fitting to the hard edge observed in some of the estimate distributions (i.e., the bounding issue at high contrasts notable in some observers data and in the pooled data). Specifically, in model fitting, we included one additional possible estimate value above 100% contrast (110%), calculating the predicted probability for this point in addition to the 38 actual possible response values. We then added this value to the observer's actual final estimate bin (i.e., 100%), in an attempt to more closely approximate the frequency of estimates for that bin. This is probably not the ideal solution, and could be improved upon with a finer-grained simulation of model estimates falling above 100% contrast, or through Monte Carlo firing-rate simulations, where model estimates will naturally fall above 100% on occasion.

4.3 RESULTS

4.3.1 DELAYED ESTIMATION OF LUMINANCE CONTRAST

Observers were presented with small, briefly flashed (200 ms) circular discs in Experiment 1, and had to reconstruct the presented contrast by adjusting the luminance of a match disc (via horizontal movements of a computer mouse). The median and inter-quartile range of estimate distributions for several individual observers, and the pooled data ($n = 8$), are illustrated in Figure 4.2. In general, the position and shape of the distributions changed in a highly consistent fashion across all observers tested. As luminance contrast of the disc increased, so too did the median and width of the estimate distributions. There appeared to be a systematic tendency for the median estimates to be shifted slightly towards the mean presented luminance contrast; this effect could have several possible causes, including the bounded nature of the response range, some form of effort-versus-accuracy trade-off in adjusting the match disc luminance, and a Bayesian prior.

The dependence of the inter-quartile range of the estimate distribution on disc contrast was well fit by a power law, with the mean exponent across observers equal to 0.54 (0.049 SEM). Thus, the exponent, or slope on a log-log axis, stood at a value similar to previous results from contrast-discrimination tasks (Legge & Foley, 1980; May & Solomon, 2015; Pestilli et al., 2011). For example, Legge & Foley (1980) found the slope of the relation between pedestal and contrast-discrimination threshold to lie around 0.6, a relation commonly referred to as the near-miss to Weber's Law (May & Solomon, 2015).

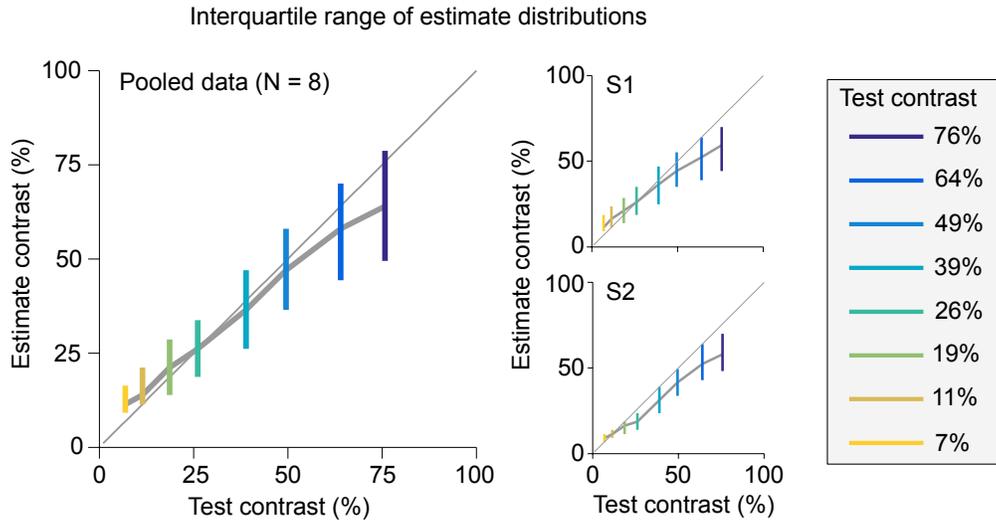


Figure 4.2: Experiment 1 results. Median and inter-quartile range of estimates are illustrated, for pooled data ($n = 8$) and for a number of individual observers. The circular disc could take one of eight luminance contrasts (different colors) across trials.

4.3.2 MODEL

We developed a four-parameter model of contrast estimation based on a hypothesized neural substrate (Figure 4.3). Specifically, we assumed that luminance contrast, c , is encoded by a population of noisy Poisson neurons whose mean spike counts are related to c via a Naka-Rushton gain function. If we assume that the observer performs maximum-likelihood estimation on the responses to obtain a contrast estimate \hat{c} , then a simple, closed-form expression for the conditional probability distribution, $p(\hat{c}|c)$, can be derived (see Materials and Methods).

The model provided quantitatively good fits to the estimate distributions for individual subjects and to the pooled data (Figure 4.4). The widths of the predicted distributions agreed with the summary statistics derived from the data. Parameter estimates from fits to subject data are provided in Table 4.1. The exponent of the Naka-Rushton equation lay

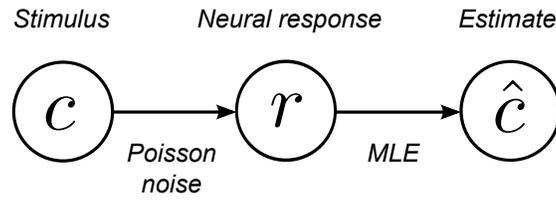


Figure 4.3: Encoding-decoding model. We assume that the presented contrast, c , gives rise to normally-distributed, Poisson firing rates in an array of neurons, r , and that the observer performs maximum-likelihood estimation on these responses, to obtain the estimate \hat{c} .

consistently in the range around 2, with a mean (SEM) across observers of 2.08 (0.2). Values in this range have been consistently observed in luminance- and contrast-response measurements in early visual cortical areas, with smaller values typically reported in fits to retinal data (Wilson, 1999). The preponderance of fitted values near 2 is also interesting given the suggested importance of this range of exponent value for information transmission efficiency in the contrast response (Gottschalk, 2002).

There did appear, however, to be some systematic deviations between the measured and fitted distributions. For low-to-intermediate contrast discs, for example, the model seemed to over-estimate the position of the peak of observer estimates, with the actual estimates weighted towards lower contrasts. This effect was present in numerous individual observers data, and is puzzling considering that the median estimates were biased in the opposite direction. A number of possible factors might be involved. First, our assumption that all neurons have the same $g(c)$ is unrealistic Albrecht & Hamilton (1982). Second, we used maximum-likelihood instead of posterior mean read-out; one could imagine that in an estimation task, observers minimize the expected squared error and therefore choose the posterior mean. Third, the normal approximation to Poisson firing statistics might not be adequate; indeed our assumption of perfectly Poisson noise may be insufficient, with a potential role played by trial-to-trial gain fluctuations (Ecker et al., 2015; Goris et al., 2014; May

Best-fitting parameters									
	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	Mean(SE)
<i>Param.</i>									
<i>a</i>	10.2	16.4	9.3	9.6	9.5	24.9	14.9	11.1	13.3(1.9)
<i>n</i>	1.79	1.87	2.74	2.13	2.75	1.0	2.41	1.95	2.08(.2)
<i>c₅₀</i>	0.91	0.46	1.0	0.44	0.73	0.15	0.98	1.0	0.71(.11)
<i>λ</i>	0.062	0.002	0.054	0.038	0.043	0.006	0.033	0.036	0.034(.007)

Table 4.1: Best-fit parameter values of the probabilistic model. Parameters could vary within broad but finite bounds: *a*, 1-50; *n*, 1-4; *c₅₀*, 0-1; *λ*, 0-1

& Solomon, 2015).

4.3.3 CONTROL EXPERIMENTS

In a series of control experiments, we also tested whether these general results were dependent on the match disc having a non-zero onset contrast, and on the polarity of the disc. First, we had wondered whether systematic deviations of the distribution shapes might have arose due to the specific match onset contrast. For example, as onset contrast was selected randomly in Experiment 1, the match disc was more often than not of higher contrast than a just-presented low-contrast target disc. Thus on average, some local adaptation or memory substitution process could have systematically affected the shape of estimate distributions for low contrast stimuli. By starting the match at 0% contrast, and requiring observers to 'dial up' the memorized target contrast, we hoped to ascertain whether such effects were systematically present in the data. They were not (Figure 4.5). Estimate distributions for trials with a 0% onset contrast were indistinguishable from the original data, with a very similar increase in distribution width with target stimulus contrast. We inter-

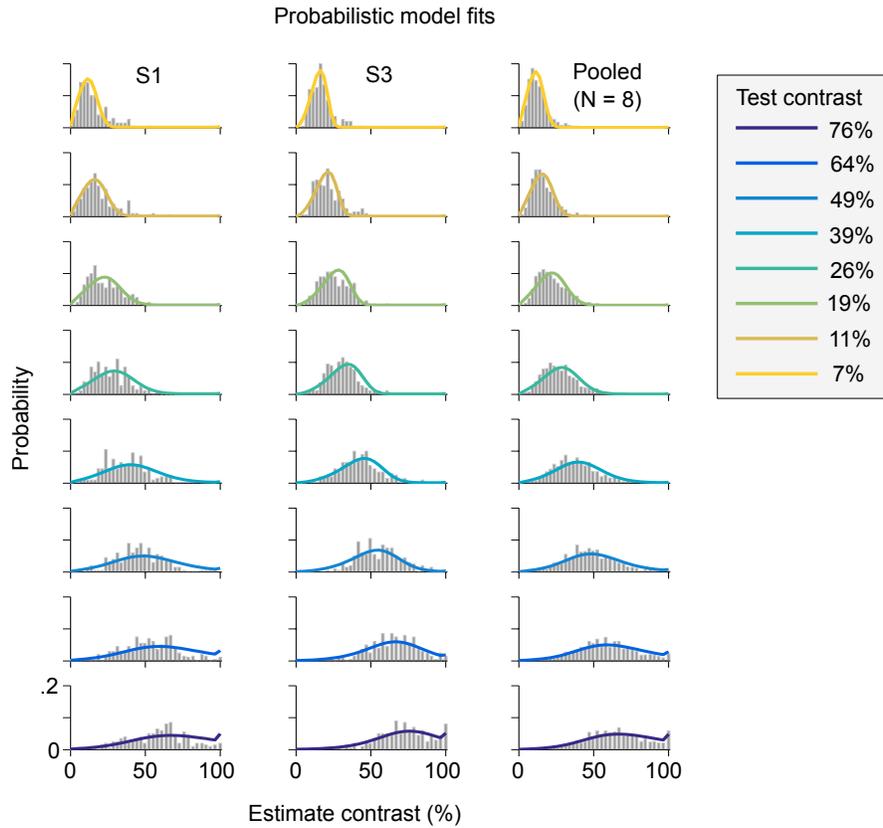


Figure 4.4: Experiment 1 distributions and model fits. Probabilistic model fits to the pooled data and for a number of individual observers. See Materials and methods for details, and Table 4.1 for best-fit parameter values.

pret this to mean that match onset played little systematic role in estimate precision; if it did, our estimation method was likely not sensitive enough to pick up on this.

We also tested the role of disc polarity. In an idealized scenario, we had figured disc polarity would not matter greatly. However, there is some evidence to suggest that dark and light patches play asymmetric roles in luminance and contrast discrimination at high luminance contrasts (Whittle, 1986; Kingdom & Whittle, 1996), as well as evidence for a general asymmetry in the neural representations for darks and lights (Yeh et al., 2009; Kremkow et al., 2014). To test this possibility, we re-ran the basic experiment, this time using negative

luminance increments instead of positive. Of note, many observers in this task now had a substantially broader estimate distribution for low contrast stimuli, and at least a couple exhibited a flattening or dip in distribution width for the highest contrasts. Overall, the distributions appeared to increase in width only negligibly with increasing contrast (Figure 4.5). While we were concerned this more extreme regression to the mean might have arose due to participant factors such as boredom, it is perhaps reasonable to imagine that some asymmetry in luminance processing is at play: for example, positive and negative luminances are known to give rise to either more quickly saturating or more linear neural responses respectively (Kremkow et al., 2014). Thus, the larger deviation in median estimate for weak negative relative to weak positive increments may reflect real asymmetries in the slope of the initial part of the contrast response, effects which are potentially very early in nature (Kremkow et al., 2014). In addition, the late flattening of the median estimate curve for negative increments might also reflect some early luminance response asymmetry. We note also that there is some behavioral evidence to suggest that it is also potentially related to the decrease in contrast-discrimination thresholds for high-contrast sinusoidal gratings (Whittle, 1986; Kingdom & Prins, 2010).

4.4 DISCUSSION

By providing continuous, high-resolution measurements of memory contents, delayed-estimation tasks have elucidated the nature of VSTM, most notably for orientation and color. Such features are relatively stable in memory across time, presumably due to the topographic form of their neural representations. For intensity-coded features such as luminance contrast, however, VSTM might be less stable over time. Unfortunately, VSTM for

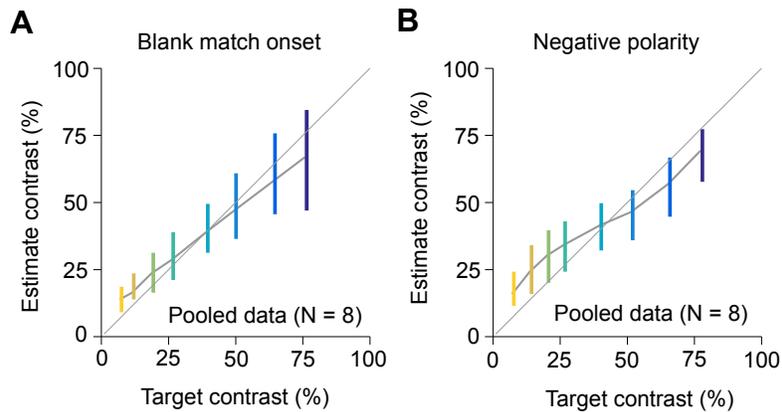


Figure 4.5: Control experiments. Median and interquartile range of the estimate distributions for the control experiments.

luminance contrast has been characterized predominantly using relatively coarse discrimination tasks. Here, we examined memory for luminance contrast using delayed estimation. We systematically measured estimate distributions for luminance contrasts spanning the contrast axis, using small uniform discs as stimuli. Memoranda for specific luminance contrasts were clearly well-defined, with estimate distributions systematically shifting position as a function of stimulus luminance contrast, and showing a small bias towards the mean presented contrast. We also found evidence of a monotonic increase in estimate distribution width with contrast, reminiscent of the near-miss to Weber’s Law often cited in the contrast-discrimination literature (Legge & Foley, 1980; May & Solomon, 2015).

We then fit a low-parameter, neurally-plausible probabilistic model to the distributions. The model assumed Poisson noise and maximum-likelihood readout, and incorporated realistic forms of contrast response function (e.g., Naka-Rushton). Using a mixture model approach that accommodated sources of trial-by-trial noise such as lapses, we successfully fit the general shape of the estimate distributions. The model predicting neurally plausible gain parameter values (e.g., Naka-Rushton exponent of 2), and replicated the monotonic

increase in estimate distribution width with increasing stimulus contrast. Control experiments indicated that match onset contrast played no substantial role in affecting the shape of observers' estimate distributions; however, polarity of the disc appeared to be influential, with a much flatter form to the curve depicting distribution width as a function of contrast. We hypothesize that this difference may be related to asymmetries in the neural representation of darks and lights (Kremkow et al., 2014; Whittle, 1986; Yeh et al., 2009); however, this conclusion is necessarily tentative for now.

Overall, the delayed-estimation protocol we developed is a successful first step in understanding the underlying nature of luminance contrast encoding and VSTM. Luminance contrast encoding has typically been investigated using the discrimination paradigm, where an observer is required to detect the occurrence of a stimulus change (e.g., a contrast increment) between two temporally-separated stimuli. Performance on such tasks is typically summarized by fitting some quantitative model to the data (i.e., a psychometric function), from which a criterion performance level is read-off, yielding a metric of decoding ability (i.e., a threshold). By comparing thresholds across different stimulation conditions (e.g., with or without covert attention, distractors, etc.), investigators have often sought to better understand the underlying neural processes involved in encoding and VSTM. However, whereas 2-AFC discrimination tasks only investigate the underlying probabilistic representations somewhat superficially (Magnussen et al., 1996; Pestilli et al., 2011), our simple task provided a very precise depiction of a fundamental sensory coding ability.

5

Conclusion

5.1 ENCODING-DECODING MODELS OF LUMINANCE CONTRAST PROCESSING

The encoding of local visual stimulus properties such as orientation, color, and luminance contrast has been studied in great detail over the years, using myriad behavioral and neural recording techniques (Brouwer & Heeger, 2009; Hubel & Wiesel, 1962; Legge & Foley, 1980; Lennie et al., 1990). The starting point for such investigations has typically been at

the level of single-stimulus processing, and for performance on simplified tasks such as 2-AFC discrimination. The present thesis focused primarily on the encoding-decoding of luminance contrast, a stimulus property fundamental to all of visual processing. We studied observer behavior using two complementary experimental protocols (discrimination and delayed estimation), and for both single-stimulus and two-stimulus tasks. We also developed neural models of observer performance on these tasks. Below, we recap briefly on the main thesis results, and discuss their implications for our understanding of a number of topics related to sensory processing. Specifically, we first focus on the implications for research on attentional selection and VSTM, along the way discussing possible future extensions of the experimental work described in Chapters 2 and 4. We then describe potential links to research on neural noise statistics and encoding-decoding models, suggesting ways in which the encoding-decoding approaches in Chapters 3 and 4 might be extended in future investigations. We then conclude with a few brief closing statements.

5.2 IMPLICATIONS FOR RESEARCH ON ATTENTIONAL SELECTION

Despite the central role attentional orienting plays in behavior, the neural bases of attentional modulation and selection remain poorly understood. A considerable number of studies now illustrate that a primary neural correlate of attention consists of an additive baseline offset in neural response (Buracas & Boynton, 2007; Chen & Seidemann, 2012; Murray, 2008; Pestilli et al., 2011). For example, by simultaneously measuring the BOLD fMRI response and behavioral performance during a contrast-discrimination task, Pestilli et al. (2011) found that the enhancement in behavioral performance that attention brings could be modeled by combining an additive offset in sensory response with a max pool-

ing rule prior to decision. Related effects have been found elsewhere (Buracas & Boynton, 2007; Chen & Seidemann, 2012; Murray, 2008). Yet, numerous recent findings have also emphasized the apparent multiplicative nature of attentional modulation of neural responses (Herrmann et al., 2010; Itthipuripat et al., 2014; MacAdams & Maunsell, 1999; Reynolds & Heeger, 2009). In one behavioral study, the authors systematically manipulated the spatial extent of an observer's attentional focus, while measuring the contrast-dependence of orientation discrimination for stimuli of different size (Herrmann et al., 2010). Results were in general agreement with a normalization model of attention, in which attention is implemented as a multiplicative weighting of incoming sensory signals (Reynolds & Heeger, 2009).

The results of Chapters 2 provide some insight into the neural and computational processes governing attentional selection, by highlighting a difference in the efficiency of selection as a function of task: under conditions of target location uncertainty, contrast-discrimination performance at a target location was more substantially hindered by high-contrast distractors than was orientation discrimination. These results indicate that selection of sensory responses in the contrast-discrimination task was spatially coarse in nature, echoing several prior related findings (Chen & Seidemann, 2012; Pestilli et al., 2011). In contrast, orientation-related information appeared to be more precisely selected and decoded. Arguably, the task-dependence of these effects suggests some re-examination of key ideas on the attentional modulation and selection of early sensory neural responses. In Chapter 3, we attempted to develop computational models of the contrast- and orientation-discrimination behaviors measured in Chapter 2, and found that neither of two standard sensory interaction models could convincingly replicate data from the two tasks simultane-

ously. Together, these empirical and computational results suggest the particular task that an observer is engaged in (e.g., contrast vs. orientation-based) is likely also a key factor in the types of behavioral effects researchers observe in attentional selection tasks. Thus, future attempts at discriminating between multiplicative and additive effects of attention on neural response should at least acknowledge the role that feature dimension (i.e., contrast, orientation, color, etc.) might play, designing multi-dimensional sets of experiments and associated models, while keeping other attentional and stimulus manipulations constant. Some efforts in this regard have recently been made in VSTM research and elsewhere (Ecker et al., 2015; Matthey et al., 2015; Orhan & Ma, 2015), where we now turn.

5.3 IMPLICATIONS FOR THE STUDY OF VSTM

Traditionally, studies of VSTM have quantified decoding performance using relatively coarse metrics such as discrimination thresholds or change detection performance (Magnussen & Greenlee, 1999; Ma et al., 2014; Palmer, 1990). In Chapter 4, we added to a growing body of research that attempts to measure more directly the noise properties of memoranda supporting basic visual feature discrimination and comparison over brief delays (Bays, 2014; Fougnie et al., 2012; van den Berg et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008). Using a delayed-estimation protocol, we neatly measured and characterized the shape of observers' estimate distributions for luminance contrast.

How do these results contribute to research on VSTM, and how might they be extended in future? First, we note that the recent surge of interest in using delayed estimation has almost entirely focused on sensory encoding-decoding for stimulus features such as orientation and color (Bays, 2014; Fougnie et al., 2012; van den Berg et al., 2012). Luminance

contrast, on the other hand, has often been treated as a nuisance parameter in such investigations, or has been utilized to create coarsely defined reliability conditions e.g., low vs. high reliability. Thus, we have developed a high-quality, yet simple experimental protocol for studying delayed estimation along an intensity-coded feature dimension. The experimental results, and associated neural model in particular, help to dispel any notion that memory for luminance contrast is not easily characterized (Magnussen & Greenlee, 1999).

Future research might develop along a number of directions, for which our protocol could serve as a basis. For example, the task in Chapter 4 might also be easily extended to study the effects of temporal delay on VSTM for luminance contrast, thereby obtaining a more fine-grained understanding of delay effects than provided by discrimination tasks (Magnussen & Greenlee, 1999). In addition, increasing evidence suggests that individual item representations systematically decrease in precision with increasing set-size, results which add to the view that VSTM relies on a noisy, continuous neural resource (Fougnie et al., 2012; van den Berg et al., 2012). Our protocol could be extended to parametrically vary set-size, thereby providing potentially greater insight into the stimulus interactions reported in Chapter 2 and elsewhere (Pestilli et al., 2011). One obstacle prevented this type of investigation until now: for features such as orientation and color, an implicit assumption is often made that estimate distribution shape does not vary greatly along the axis of the relevant feature dimension (van den Berg et al., 2012), leading to relatively simplified designs for set-size type experiments (e.g., presenting an array of randomly oriented gabors). However, as the data in Chapter 4 illustrate, this assumption would be grossly invalid for the case of luminance contrast; estimate distributions change shape dramatically as a function of stimulus contrast. Future experiments that manipulate set-size would need to account

for this in their design, parametrically varying the array of test contrasts and controlling for direct effects of distractors on target stimulus encoding and decoding. The models of sensory interaction described in Chapter 3 may provide some guidance here.

5.4 NEURAL NOISE AND ENCODING-DECODING

The present results may also be of relevance to recent debate on the nature of neural noise statistics and encoding-decoding. Numerous recent delayed-estimation tasks have consistently found that error distributions are non-Gaussian in form, perhaps reflecting trial-to-trial fluctuations in encoding precision (Bays, 2014; Fougner et al., 2012; van den Berg et al., 2012). While the model presented in Chapter 4 did not incorporate sources of trial-to-trial gain fluctuation, more realistic noise models may be worth investigating in fitting estimation data for luminance contrast. As an intensity-coded feature, luminance contrast is likely encoded into memory in a relatively abstract, albeit firing-rate dependent way (Albrecht & Hamilton, 1982; Xing et al., 2014). Thus, trial-to-trial fluctuations in gain magnitude (i.e., double stochasticity) would presumably directly affect the trial-to-trial variation in estimates made for a given luminance contrast. Overall, estimate distributions for luminance contrast might be impacted more directly by the stochastic properties of early sensory encoding i.e., the shape of observers' estimate distributions might reflect properties of Poisson or super-Poisson noise statistics in some principled fashion, with predictable variation across different luminance contrast levels (Goris et al., 2014; Shadlen & Newsome, 1998). In contrast, recent theoretical work suggests that precision on tasks involving circular, Gaussian-like tuning functions (e.g., orientation-tuning curves) should be little affected by fine-grained characteristics of local noise statistics, such as the doubly-stochastic

nature of gain fluctuations (Ecker et al., 2015; May & Solomon, 2015; Moreno-Bote et al., 2014). Instead, the key limiting factors on precision for such tasks may be network computations that affect response amplitudes for all simultaneously-stored items (e.g., divisive normalization), as well as noise fluctuations that resemble the signal of interest (Bays, 2014; Ecker et al., 2015; Moreno-Bote et al., 2014; Wei et al., 2012). However, it remains to be seen whether a model that incorporates trial-to-trial gain fluctuations will better approximate our contrast estimation data.

5.5 FINAL COMMENTS

Using a combined empirical and computational approach, this thesis explored the nature of luminance contrast encoding and decoding, fundamental operations of visual system processing. By measuring observer performance in a variety of behavioral tasks, and fitting appropriately chosen mathematical models to their data, we highlighted important characteristics of the encoding and decoding of stimulus luminance contrast, such as the large effect of irrelevant distractors on basic discrimination abilities, and the basic profile of observers' internal, noisy estimates of luminance. The thesis findings are relevant to a variety of subfields within the visual and sensory neurosciences, such as research on attention, memory and general models of stimulus encoding-decoding.

References

- Abbott, L. & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, 11, 91–101.
- Albrecht, D. & Hamilton, D. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*, 48, 217–237.
- Anderson, D., Ester, E., Serences, J., & Awh, E. (2013). Attending multiple items decreases the selectivity of populations responses in human primary visual cortex. *The Journal of Neuroscience*, 33, 9273–9282.
- Averbeck, B. & Lee, D. (2006). Effects of noise correlations on information encoding and decoding. *Journal of Neurophysiology*, 95, 3633–3644.
- Awh, E. & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5, 119–126.
- Bays, P. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience*, 34, 3632–3645.
- Bays, P., Catalao, R., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10)(7), 1–11.
- Bays, P. & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321, 851–854.
- Berens, P., AS, E., Cotton, R., Ma, W., Bethge, M., & Tolias, A. (2012). A fast and simple population code for orientation in primate V1. *The Journal of Neuroscience*, 32, 10618–10626.
- Blake, R. & Holopigian, K. (1985). Orientation selectivity in cats and humans assessed by masking. *Vision Research*, 25, 1459–1467.

- Boynton, G., Demb, J., Glover, G., & Heeger, D. (1999). Neuronal basis of contrast discrimination. *Vision Research*, 39, 257–269.
- Bradley, A. & Ohzawa, I. (1986). A comparison of contrast detection and discrimination. *Vision Research*, 26, 991–997.
- Brouwer, G. & Heeger, D. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of Neuroscience*, 29, 13992–14003.
- Buracas, G. & Boynton, G. (2007). The effect of spatial attention on contrast response functions in human visual cortex. *The Journal of Neuroscience*, 27, 93–97.
- Busse, L., Wade, A., & Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, 64, 931–942.
- Callaway, E. (2003). Cell types and local circuits in primary visual cortex of the macaque monkey. In L. Chalupa & J. Werner (Eds.), *The Visual Neurosciences* (pp. 680–694). Cambridge: MIT Press.
- Carandini, M. & Heeger, D. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13, 51–62.
- Chen, Y. & Seidemann, E. (2012). Attentional modulations related to spatial gating but not to allocation of limited resources in primate V1. *Neuron*, 74, 557–566.
- Chirimuuta, M. & Tolhurst, D. (2005). Does a bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research*, 45, 2943–2959.
- Chubb, C., Landy, M., & Econopoulou, J. (2004). A visual mechanism tuned to black. *Vision Research*, 44, 3223–3232.
- Cover, T. & Thomas, J. (1991). *Elements of information theory*. New York: Wiley, 1st edition.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Dayan, P. & Abbott, L. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge: MIT Press, 1st edition.
- Ecker, A., Berens, P., Tolias, A., & Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience*, 31, 14272–14283.

- Ecker, A., Denfield, G., Bethge, M., & Tolias, A. (2015). On the structure of population activity under fluctuations in attentional state. *bioRxiv*, —, —.
- Engbert, R. & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, 43, 1035–1045.
- Ferster, D. (2003). Assembly of receptive fields in primary visual cortex. In L. Chalupa & J. Werner (Eds.), *The Visual Neurosciences* (pp. 695–703). Cambridge: MIT Press.
- Fougnie, D., Suchow, J., & Alvarez, G. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1229.
- Ganguli, D. & Simoncelli, E. (2014). Efficient sensory encoding and bayesian inference with heterogeneous neural populations. *Neural Computation*, 26, 2103–2134.
- Gegenfurtner, K. & Kiper, D. (2003). The processing of color in extrastriate cortex. In L. Chalupa & J. Werner (Eds.), *The Visual Neurosciences* (pp. 1017–1028). Cambridge: MIT Press.
- Georgeson, M. & Sullivan, G. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *Journal of Physiology*, 252, 627–656.
- Gorea, A. & Sagi, D. (2001). Disentangling signal from noise in visual contrast discrimination. *Nature Neuroscience*, 4, 1146–1150.
- Goris, R., Movshon, J., & Simoncelli, E. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17, 858–865.
- Gottschalk, A. (2002). Derivation of the visual contrast response function by maximizing information rate. *Neural Computation*, 14, 527–542.
- Graf, A., Kohn, A., Jazayeri, M., & Movshon, J. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neuroscience*, 14, 239–245.
- Hansen, N. & Ostermeier, A. (1996). Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation*, (pp. 312–317).
- Hara, Y. & Gardner, J. (2014). Encoding of graded changes in spatial specificity of prior cues in human visual cortex. *Journal of Neurophysiology*, 112, 2834–2849.
- Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–197.

- Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. (2010). When size matters: attention affects performance by contrast or response gain. *Nature Neuroscience*, 13, 1554–1559.
- Hubel, D. & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106–154.
- Itthipuripat, S., Garcia, J., Rungratsameetaweemana, N., Sprague, T., & Serences, J. (2014). Changing the spatial scope of attention alters patterns of neural gain in human cortex. *The Journal of Neuroscience*, 34, 112–123.
- Johnson, E., Hawken, M., & Shapley, R. (2001). The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nature Neuroscience*, 4, 409–416.
- Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M., & Moore, K. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 193–224.
- Kaplan, E. & Shapley, R. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences, USA*, 83, 2755–2757.
- Keshvari, S., van den Berg, R., & Ma, W. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology*, 9(2)(e1002927).
- Kingdom, F. & Prins, N. (2010). *Psychophysics: a practical introduction*. London: Academic Press, 1st edition.
- Kingdom, F. & Whittle, P. (1996). Contrast discrimination at high contrasts reveals the influence of local light adaptation on contrast processing. *Vision Research*, 36, 817–829.
- Kremkow, J., Jin, J., Komban, S., Wang, Y., Lashgari, R., Li, X., Jansen, M., Zaidi, Q., & Alonso, J.-M. (2014). Neuronal nonlinearity explains greater visual spatial resolution for darks than lights. *Proceedings of the National Academy of Sciences, USA*, 111, 3170–3175.
- Lara, A. & Wallis, J. (2012). Capacity and precision in an animal model of visual short-term memory. *Journal of Vision*, 12(3)(13), 1–12.
- Lee, B. & Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Research*, 36, 2159–2166.
- Legge, G. & Foley, J. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, 70, 1458–1471.

- Lennie, P., Krauskopf, J., & Sclar, G. (1990). Chromatic mechanisms in striate cortex of macaque. *The Journal of Neuroscience*, 10, 649–669.
- Luck, S. & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- Ma, W., Husain, M., & Bays, P. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17, 347–356.
- Ma, W., Navalpakkam, V., Beck, J., van den Berg, R., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, 14, 783–790.
- Ma, W., Shen, S., Dziugaite, G., & van den Berg, R. (2015). Requiem for the max rule? *Vision Research*, –, –.
- MacAdams, C. & Maunsell, J. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *The Journal of Neuroscience*, 19, 431–441.
- Magnussen, S. & Greenlee, M. (1999). The psychophysics of perceptual memory. *Psychological Research*, 62, 81–92.
- Magnussen, S., Greenlee, M., & Thomas, J. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 202–212.
- Mareschal, I. & Shapley, R. (2004). Effects of contrast and size on orientation discrimination. *Vision Research*, 44, 57–67.
- Matthey, L., Bays, P., & Dayan, P. (2015). A probabilistic palimpsest model of visual short-term memory. *PLoS Computational Biology*, 11(1)(e1004003).
- May, K. & Solomon, J. (2015). Connecting psychophysical performance to neuronal response properties i: discrimination of suprathreshold stimuli. *Journal of Vision*, 15(6)(8), 1–28.
- Mazyar, H., van den Berg, R., & Ma, W. (2012). Does precision decrease with set size? *Journal of Vision*, 12(6)(10), 1–16.
- McIlwain, J. (1996). *An introduction to the biology of vision*. Cambridge: Cambridge University Press.
- Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience*, 17, 1410–1417.

- Murray, S. (2008). The effects of spatial attention in early human visual cortex are stimulus independent. *Journal of Vision*, 8(10)(2), 1–11.
- Nachmias, J. & Sansbury, R. (1974). Grating contrast: discrimination may be better than detection. *Vision Research*, 14, 1039–1042.
- Naka, K. & Rushton, W. (1966). S-potentials from luminosity units in the retina of fish (cyprinidae). *Journal of Physiology*, 185, 587–599.
- Nilsson, T. & Nelson, T. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 141–150.
- Ohzawa, I., Sclar, G., & Freeman, R. (1985). Contrast gain control in the cat's visual system. *Journal of Neurophysiology*, 54, 651–667.
- Orhan, A. & Ma, W. (2015). Neural population coding of multiple stimuli. *The Journal of Neuroscience*, 35, 3825–3841.
- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 332–350.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40, 1227–1268.
- Paradiso, M. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics*, 58, 35–49.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, 44, 369–378.
- Pelli, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America, A*, 2, 1508–1532.
- Pestilli, F., Carrasco, M., Heeger, D., & Gardner, J. (2011). Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. *Neuron*, 72.
- Pouget, A., Dayan, P., & Zemel, R. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, 26, 381–410.
- Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention. I. color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 261–282.

- Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention. 2. brightness and contrast. *Consciousness and Cognition*, 6, 372–412.
- Ratliff, C., Borghuis, B., Kao, Y.-H., Sterling, P., & Balasubramanian, V. (2010). Retina is structured to process an excess of darkness in natural scenes. *Proceedings of the National Academy of Sciences, USA*, 107, 17368–17373.
- Recanzone, G., Wurtz, R., & Schwarz, U. (1997). Responses of mt and mst neurons to one and two moving objects in the receptive field. *Journal of Neurophysiology*, 78, 2904–2915.
- Reynolds, J. & Heeger, D. (2009). The normalization model of attention. *Neuron*, 61, 168–185.
- Sanborn, A. & Dayan, P. (2011). Optimal decisions for contrast discrimination. *Journal of Vision*, 11(14)(9), 1–13.
- Schiller, P., Sandell, J., & Maunsell, J. (1986). Functions of the on and off channels of the visual system. *Nature*, 322, 824–825.
- Sergent, C., Ruff, C., Barbot, A., Driver, J., & Rees, G. (2011). Top-down modulation of human early visual cortex after stimulus offset supports successful postcued report. *Journal of Cognitive Neuroscience*, 23, 1921–1934.
- Seriès, P., Stocker, A., & Simoncelli, E. (2009). Is the homunculus "aware" of sensory adaptation? *Neural Computation*, 21, 3271–3304.
- Seung, H. & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences, USA*, 90, 10749–10753.
- Shadlen, M. & Newsome, W. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of Neuroscience*, 18, 3870–3896.
- Shapley, R. (1990). Visual sensitivity and parallel retinocortical channels. *Annual Reviews of Psychology*, 41, 635–658.
- Shapley, R. & Victor, J. (1979). The contrast gain control of the cat retina. *Vision Research*, 19, 431–434.
- Sincich, L. & Horton, J. (2005). The circuitry of v1 and v2: integration of color, form, and motion. *Annual Reviews of Neuroscience*, 28, 303–326.

- Skottun, B., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *Journal of Neurophysiology*, 57, 773–786.
- Tolhurst, D., Movshon, J., & Thompson, I. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, 41, 414–419.
- van den Berg, R., Shin, H., Chou, W., George, R., & Ma, W. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences USA*, 109, 8780–8785.
- Wei, Z., Wang, X., & Wang, D. (2012). From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *The Journal of Neuroscience*, 32, 11228–11240.
- Westheimer, G., Shimamura, K., & McKee, S. (1976). Interference with line-orientation sensitivity. *Journal of the Optical Society of America*, 66, 332–338.
- Whittle, P. (1986). Increments and decrements: luminance discrimination. *Vision Research*, 26, 1677–1691.
- Wichmann, F. & Hill, H. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63, 1293–1313.
- Wiesel, T. & Hubel, D. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *Journal of Neurophysiology*, 29, 1115–1156.
- Wilken, P. & Ma, W. (2004). A detection theory account of change detection. *Journal of Vision*, 4, 1120–1135.
- Wilson, H. (1999). *Spikes, decisions and actions: dynamical foundations of neuroscience*. Oxford: Oxford University Press.
- Xing, Y., Ledgey, T., McGraw, P., & Schluppeck, D. (2014). The influence of spatial pattern on visual short-term memory for contrast. *Attention, Perception and Psychophysics*, 76, 1925–1932.
- Yeh, C.-I., Xing, D., & Shapley, R. (2009). "Black" responses dominate macaque primary visual cortex v1. *The Journal of Neuroscience*, 29, 11753–11760.
- Zenger-Landolt, B. & Heeger, D. (2003). Response suppression in v1 agrees with psychophysics of surround masking. *The Journal of Neuroscience*, 23, 6884–6893.

Zhang, W. & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235.

Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *The Journal of Neuroscience*, 27, 12292–12307.