2022

Computational Mechanisms Underlying Perception Of Visual Motion

Benjamin Ming Chin
University of Pennsylvania

Computational Mechanisms Underlying Perception Of Visual Motion

Abstract

Motion is a fundamental property estimated by human sensory-perception. When visual shapes and patterns change their positions over time, we perceive motion. Relating properties of perceived motion—speed and direction—to properties of visual stimuli is an important endeavor in vision science. Understanding this relationship requires an understanding of the computations performed by the visual system to extract motion information from visual stimuli. The present research sheds light on the nature of these computations. In the first study, human performance in a speed discrimination task with naturalistic stimuli is compared to performance of an ideal observer model. The ideal observer model utilizes computations that have been optimized for discriminating speed among a large training set of naturalistic stimuli. Although human performance falls short of ideal observer performance because of the presence of internal noise, the remarkable finding is that the computations performed minimize, to the maximum possible extent, the performance limits imposed by external stimulus variability. In other words, humans perform computations that are optimal. The second study focuses on how spatial frequency, a basic characteristic of visual patterns, impacts the process by which the visual system integrates motion across time (temporal integration). A continuous target-tracking task demonstrates that longer temporal integration periods are associated with higher spatial frequencies. This predicts a visual depth illusion when the left and right eyes are simultaneously presented stimuli having different spatial frequencies. A second experiment using traditional forced-choice psychophysics confirms this prediction. The third study explores how color impacts estimates of spatial position during motion. We parameterize color in terms of L-cone and S-cone activity modulations in the eye. Using the same continuous target-tracking paradigm from Chapter 2, we demonstrate that position estimates for stimuli comprised of pure S-cone modulations lag behind position estimates for stimuli comprised of pure L-cone modulations. A key finding is that when L-cone and S-cone modulations are combined, processing lag is almost exclusively determined by L-cone modulations.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Psychology

First Advisor

Johannes D. Burge

Keywords

Computation, Perception, Vision

Subject Categories

Psychology

COMPUTATIONAL MECHANISMS UNDERLYING PERCEPTION OF VISUAL MOTION

Benjamin Ming Chin

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

Johannes Burge, Associate Professor of Psychology

Graduate Group Chairperson

Russell Epstein, Professor of Psychology

Dissertation Committee

David Brainard, Professor of Psychology

Joshua Gold, Professor of Neuroscience

Michael Arcaro, Assistant Professor of Psychology

ACKNOWLEDGEMENT

I would like to express my gratitude to a number of people who have been important to me on the road to my PhD.

First, I would like to thank my advisor, Johannes Burge. Johannes has been exceptional at transferring his vast repository of knowledge to me. He has taught me how to do good science, how to communicate good science, and how to navigate the complex world of academia. Most importantly, we get along well.

I would like to thank my thesis committee. David Brainard and Josh Gold have a knack for asking insightful, precise questions that get to the heart of the matter. Their feedback has shaped the direction of my research in important ways. I want to thank Mike Arcaro for agreeing to my last-minute request for him to be on my committee.

I would like to thank other faculty members in the psychology department whom I have learned from: Nicole Rust, Alan Stocker, and Geoff Aguirre. In particular, Geoff and Alan each facilitated busy first-year lab rotations for me, during which I learned a great deal. I also want to thank my undergraduate advisors, Janet Andrews and Ken Livingston, for inspiring me to head down the path I took.

I am thankful to members of the Burge lab who have shared my journey. Arvind Iyer, Seha Kim, and Takahiro Doi were deep sources of knowledge for me throughout my PhD, and great people to be around. I want to thank David White for the funny and sometimes absurd conversations we have had over the years, and for the many times he has had me over at his place. I have often enjoyed being around Victor Rodriguez, Anthony Loprete, and Long Ni, although I have not had the opportunity to get to know them as well, owing to the limited overlap of our respective times in the lab.

There are many people whom I want to thank for their friendship, which has brought me much happiness over the years: Michael Barnett, Lingqi Zhang, Noam Roth,

Catrina Hacker, Kyra Schapiro, Ariana Familiar, Harrison McAdams, Ana Radonjic, Sasha Igdalova, Kara McGaughey, Cheng Qiu, Harrison Mace, Jiang Mao, Melanie White, Rivka Cohen, Anne Park, Simon Bohn, Pedro Ortega, Andy Sironko, Robert Ronan, Nick Olsen, Jay Lancaster, Gianna Constantine, Aaron Bendich, Joel Auerbach, Weizhe Lim, Kenneth Cheong, Rocco Hu.

Lastly, I would like to thank my family: my parents, Gaik Eng Lim and Teng Sam Chin, and my sister, Emma. I am exceptionally grateful to my parents for a happy childhood, and for encouraging my intellectual development as I got older.

ABSTRACT

COMPUTATIONAL MECHANISMS UNDERLYING PERCEPTION OF VISUAL MOTION

Benjamin M. Chin

Johannes Burge

Motion is a fundamental property estimated by human sensory-perception. When visual shapes and patterns change their positions over time, we perceive motion. Relating properties of perceived motion—speed and direction—to properties of visual stimuli is an important endeavor in vision science. Understanding this relationship requires an understanding of the computations performed by the visual system to extract motion information from visual stimuli. The present research sheds light on the nature of these computations. In the first study, human performance in a speed discrimination task with naturalistic stimuli is compared to performance of an ideal observer model. The ideal observer model utilizes computations that have been optimized for discriminating speed among a large training set of naturalistic stimuli. Although human performance falls short of ideal observer performance because of the presence of internal noise, the remarkable finding is that the computations performed minimize, to the maximum possible extent, the performance limits imposed by external stimulus variability. In other words, humans perform computations that are optimal. The second study focuses on how spatial frequency, a basic characteristic of visual patterns, impacts the process by which the visual system integrates motion across time (temporal integration). A continuous target-tracking task demonstrates that longer temporal integration periods are associated with higher spatial frequencies. This predicts a visual depth illusion when the left and right eyes are simultaneously presented stimuli having different spatial frequencies. A second experiment using traditional forced-choice psychophysics confirms this prediction. The third study explores how color impacts estimates of spatial position during motion. We parameterize color in terms of L-cone and S-cone activity modulations in the eye. Using the same continuous target-tracking paradigm from Chapter 2, we demonstrate that position estimates for stimuli comprised of pure S-cone modulations lag behind position estimates for stimuli comprised of pure L-cone modulations. A key finding is that when L-cone and S-cone modulations are combined, processing lag is almost exclusively determined by L-cone modulations.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1: BACKGROUND

1.1 Introduction

Motion is a fundamental physical property of the world. In one sense, the story of our universe is a story of motion. Objects at all manner of scales—atoms, living beings, celestial bodies—change their positions over time. Motion has been treated as a central topic in physics since the very beginning of the field, from Aristotle's *Physics* to Newton's Laws of Motion.

This dissertation concerns the perception of motion, which is the domain of psychology and vision science. The ability to accurately estimate the speed and direction of distal objects in the environment is key to our interactions with the world. Human beings need accurate motion perception to intercept moving objects, or to avoid dangerous objects heading towards us. Accurate perception of self-motion—knowing how fast we are walking, running, or being moved—is necessary to successfully navigate our environment.

The process of perceiving distal motion begins with retinal motion. Light reflected from objects in the environment enters the eye and forms images on the retina. When objects in the environment move, light patterns on the retina move. Thus, in order for the visual system to estimate the motion of distal objects in the environment, accurate estimation of motion on the retina must first be achieved. Understanding the process of estimating retinal motion is the focus of Chapter 2.

1.2 Distinction between Local and Global Motion

In vision science, a distinction is drawn between local motion, which occurs over a small spatial extent, and global motion, which occurs over a large spatial extent. The spatial extent of local motion is not exactly defined but is typically considered to match that of receptive fields belonging to motion-selective neurons in visual cortex. It is widely thought that local motion computations form building blocks that are later integrated to support global motion processing. This belief is backed by psychophysical evidence. Certain global motion percepts can be disrupted by simple stimulus manipulations, such as changing the orientation of small elements in the stimulus (Lorenceau & Alais, 2001), or decreasing the overall contrast of the stimulus (Weiss, Simoncelli, & Adelson, 2002). When global motion percepts are disrupted in this way, subjects typically perceive local motion instead.

Importantly, the work in this chapter focuses on local rather than global motion processing. In the vision science community, the computations underlying local motion processing remain incompletely understood. Since global motion processing is widely thought to rely on local motion processing, it stands to reason that unraveling the nature of these computations will facilitate progress in understanding more complex types of motion.

1.3 Motion Energy

For decades, the motion energy model and its variants (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Watson & Ahumada, 1985) have dominated the study of local motion processing. In its most general form, a motion energy unit consists of a pair of linear spatiotemporal filters oriented in space-time (Fig. 1.1A). The filters are phase-shifted with respect to each other by 90 degrees (also known as a quadrature pair). The output of the motion energy unit is computed by squaring and summing the linear responses of each filter in the quadrature pair. The ‘motion energy’ over a visual stimulus can be computed by convolving the motion energy unit with the space-time representation of the stimulus (Fig. 1.1BC). Motion energy models are explicitly designed to detect orientation in space-time; a property shared by a wide range of visual stimuli that elicit motion percepts.

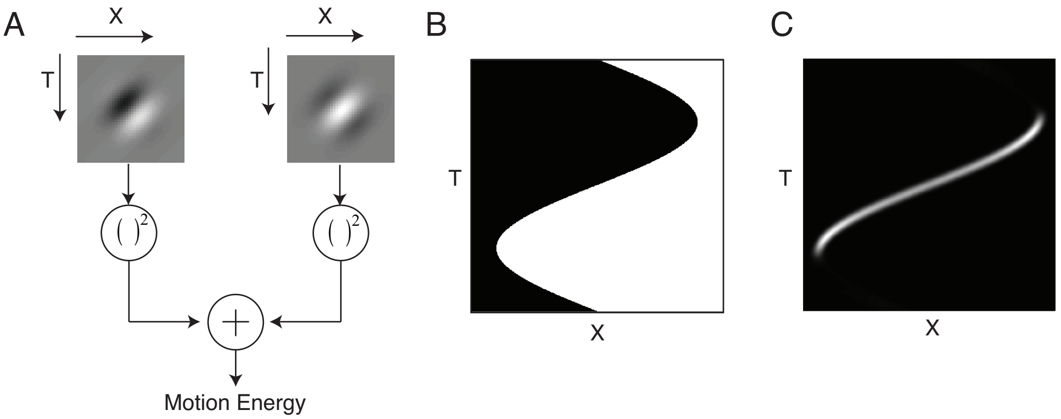

Figure 1.1. The motion energy model. **A** Schematic of the motion energy model. The outputs of two linear spatiotemporal filters are squared and summed. In this example, the filters are oriented in space-time such that they select for leftward motion. The filters are in quadrature phase. **B** An example stimulus that the motion energy model in A could be applied to. The stimulus is represented in space-time, with horizontal position on the x-axis, and time on the y-axis. In this case, the stimulus is a sharp vertical edge that drifts rightwards, and then leftwards, sinusoidally. **C** Output of the motion energy model shown in A when applied to the stimulus shown in B. The motion energy model responds selectively to leftward motion, but not to rightward motion.

Motion energy models have been popular for at least two important reasons. Owing to their quadrature filters, motion energy units have the key property of being phase invariant; they produce a constant response to their preferred motion regardless of the polarity of the stimulus contrast. Such phase invariance is frequently observed in motion-selective MT neurons. Additionally, motion energy models provide a simple, unified explanation for motion percepts observed in several types of visual stimuli that fall outside the ordinary category of continuous, translating motion. These include sampled motion (as seen on cathode-ray tube televisions), the Reverse Phi illusion, and the fluted square-wave illusion.

The motion energy approach, and those of any filter-based models of motion processing, appear counterintuitive at first; absent is the notion of a well-defined object translating through space. Indeed, prior to the introduction of filter-based models, a 'corresponding points' approach to modeling motion processing was popular (Ullman, 1979; Anstis, 1980; Anstis, 1979; Lappin and Bell, 1972). In this approach, the visual system is modeled as extracting motion in several stages: 1) identifying salient features in visual input, 2) locating these features across successive points in time, 3) estimating the time Δt and distance Δx traveled, and 4) computing motion as $\Delta x / \Delta t$. Such an approach is attractive because it often coheres with one's own perceptual experience of motion. However, 'corresponding points' approaches often have difficulty making predictions about how motion is perceived. Such models need to first specify what counts as a salient 'feature'; a challenging endeavor given the staggering variety of possible visual inputs. Additionally, for many visual stimuli in which clear motion is perceived, it is difficult to devise rules specifying which features need to be matched across time. Such stimuli include the Reverse Phi and fluted-square-wave illusions, which are specifically designed to not contain matching features across successive frames. On the other hand, the motion energy model offers an intuitive, easily visualized explanation for these motion perception phenomena. 'Corresponding points' approaches have thus been less popular than filter-based models.

Despite the relative success of the motion energy model compared to other modeling approaches, it is not actually an account of motion perception. The motion energy model alone does not output estimates of motion direction or speed, which are the basic attributes of motion perception. Rather, the motion energy model is a model of motion encoding; a motion energy unit outputs responses to motion information present in its inputs. A decoder is still required to convert these responses into motion estimates.

The motion energy model has other weaknesses. First, it fails to account for a number of motion perception phenomena that have collectively been termed second-order and third-order motion (Badcock & Derrington, 1985; Lu & Sperling, 1995). Second, the form and shape of the spatiotemporal filters (spatial frequency, bandwidth, etc.) comprising a motion energy unit are unconstrained. Thus, it is up to the researcher to choose these parameters when implementing the motion energy model. In many cases, the parameterized forms of the filters are chosen for mathematical convenience rather than based on normative principles. Third, the motion energy model cannot distinguish between changes in speed and contrast; the two properties are confounded. A weak response from a motion energy unit could be due simply due to the input stimulus having low contrast, rather than to a lack of motion in the input stimulus.

1.4 Ideal Observers

Recent breakthroughs in ideal observer modeling have addressed weaknesses of the motion energy model. Accuracy Maximization Analysis (AMA) is a recently developed Bayesian method for discovering the optimal linear filters that, when paired with a matched optimal decoder, maximize performance in a particular task with a particular set of stimuli (Geisler et al., 2009; Burge & Jaini, 2017; Jaini & Burge, 2017). The optimal filters and optimal decoder together form what is known as an ideal observer model. AMA has been successfully applied to the task of retinal speed estimation (Burge & Geisler, 2015). The resulting filters were specifically optimized for the statistics of a large training set of natural image movies. Thus, they constitute a principled starting point for investigating the filters used by the human visual system, rather than an arbitrary choice made by the researcher. Interestingly, the response distributions of the optimal filters can be optimally decoded by motion energy-like computations. This is despite the fact that the optimal filters were not explicitly designed to be used in conjunction with motion energy computations; the filter properties were entirely determined by the training set of stimuli.

AMA is one of the latest developments in a long history of ideal observer modeling. It supports the creation of an important class of ideal observers: image-computable ideal observers. These observers explicitly model the computations performed on the proximal stimulus (Banks et al., 1987; Geisler, 1989). Image-computable ideal observers have been successfully applied to simple visual tasks with simple stimuli, such as detection of a Gabor target in Gaussian white noise. For such cases, the optimal computations can often be derived analytically because the statistical properties of the stimuli are known exactly. In the case of speed estimation with naturalistic stimuli, an analytic solution is not available due to the complex statistics of natural stimuli. These complex statistics make it difficult to determine which features in natural stimuli are relevant to the task of speed estimation (signal), and which features are irrelevant (noise). AMA solves the problem of determining the task-relevant features by numerically searching the space of possible linear filters.

Once an ideal observer has been built for a task, and the optimal computations determined, an obvious question to ask is whether the same optimal computations are also performed by humans. The image-computable ideal observer developed by Burge and Geisler (2015) for speed estimation with naturalistic stimuli yielded predictions that closely approximated the pattern of behavioral data from human observers performing the same task. By fitting a single free parameter, efficiency, ideal observer performance could be quantitatively matched to human performance. The efficiency parameter was necessary because without any free parameters, human performance was significantly lower than that of the ideal observer. This discrepancy does not necessarily mean that humans are performing suboptimal computations; it could be due instead to unsystematic noise in the visual system. The aim of the work described in Chapter 2 was to determine whether the discrepancy was due to suboptimal computations, noise, or a mixture of both.

A defining aspect of the optimal computations emphasized in Chapter 2 is the fact that they were optimized for a set of naturalistic stimuli. These stimuli are naturalistic because they were generated from a large image database of natural scenes. Thus, their statistical properties are far more complex than those typically used in studies of motion perception, such as Gabors. Naturalistic stimuli are typically broadband, containing energy at many spatial frequencies. The ideal observer described in Chapter 2 quantifies the net effect of these naturalistic variations in spatial frequency content on motion perception.

1.5 3D Motion Perception and the Pulfrich Effect

Chapter 3, like Chapter 2, investigates the effect of spatial frequency on motion perception. But rather than investigating the net effect of variations in spatial frequency, Chapter 3 inspects a targeted hypothesis about the relationship between spatial frequency and a key process supporting motion perception: temporal integration. Additionally, Chapter 3 investigates the perception of 3D motion in depth rather than retinal motion.

Central to 3D perception is stereopsis: differences in retinal image positions between the eyes, known as binocular disparities, are strong cues to the depths of objects in space. Percepts of motion in depth result from changes in binocular disparities over time. The computations underlying the estimation of binocular disparity have received significant attention in vision science. Of particular importance is the correspondence problem; the visual system needs to determine what image patterns are to be matched between the two eyes. Only when a match has been established can binocular disparity be computed. With visual stimuli that move, the correspondence problem can be disrupted by differences in temporal processing properties between the eyes. A well-known example of such a disruption is the Pulfrich effect: when the same image oscillates horizontally in both eyes, delaying the processing of the image in one eye leads to an illusory elliptical trajectory in depth. This is because the processing delay induces an effective position shift of the image in one eye relative to the other; an ‘older’ position signal is matched with a ‘newer’ position signal. The processing delay can be achieved by decreasing the luminance (Pulfrich, 1922) or contrast in one eye. If the stimulus is blurred, processing is sped up rather than blurred, leading to an elliptical depth trajectory in the opposite direction (Burge, Rodriguez-Lopez, & Dorronsoro, 2019). This is known as the Reverse Pulfrich effect. Blur speeds up processing by selectively removing higher spatial frequencies, which are known to be processed more slowly than lower spatial frequencies.

1.6 Characteristics of Temporal Processing

Differences in processing delay for different spatial frequencies have been well-established through both psychophysical and neurophysiological methods. Human observers are slower to react to the onset of Gabors with higher spatial frequencies than to lower spatial frequencies, either when pressing a button (Parker, 1980; Mihaylova, Stomonyakov, & Vassilev, 1998; Vassilev, Mihaylova, & Bonnet, 2002) or when pulling a lever (Harweth & Levi, 1978). Visual evoked potentials (VEPs) have longer latencies for higher spatial frequencies (Vassilev, Mihaylova, & Bonnet, 2002). Extracellular recordings of neurons in macaque areas V1 and MT, as well as cat area 17, indicate longer response latencies for higher spatial frequencies (Bair & Movshon, 2004; Frazor et al., 2004).

Here, we explore the effects on 3D motion perception of a comparatively less-studied aspect of temporal processing: the temporal integration period. At various stages of the visual system, motion signals are averaged over time. This has been demonstrated in electrophysiological experiments. Spike-triggered averages (STAs) over velocity in macaque areas V1 and MT typically have full-widths-at-half-height ranging from 25-50ms, meaning that stimulus motion must be sustained for at least 25-50ms to elicit a spike from these neurons (Bair & Movshon, 2004). Notably, STAs are broader in time for higher spatial frequencies, indicating longer temporal integration durations for higher spatial frequencies.

Signatures of temporal integration are also clear in behavior. Smooth pursuit eye movements made in response to dot textures executing a spatially uniform random walk in time contain fewer high temporal frequency components than the stimulus motion itself (Osborne & Lisberger, 2009). Temporal integration can also be measured in target detection experiments by examining how contrast sensitivity increases as a function of stimulus duration (Nachmias, 1967; Burr, 1981; Marx & May, 1983). Short temporal integration periods result in contrast sensitivity curves that saturate quickly as a function of stimulus duration, whereas long temporal integration periods result in contrast sensitivity curves that saturate slowly. As is the case with neurophysiological findings, estimated temporal integration periods from psychophysics are longer for higher spatial frequencies.

Our investigation of temporal integration was driven by two complementary goals. One goal was to characterize the temporal integration periods associated with different spatial frequencies. To do so, we used a recently developed paradigm (Bonnen et al., 2015) requiring observers to track a continuously moving Gabor target with a mouse cursor. The second goal was to demonstrate that these differences in temporal integration periods can measurably impact perception of 3D motion in depth: namely, they cause a previously-reported but poorly-understood anomalous Pulfrich effect.

1.7 The Temporal Binding Problem and Color

The problems investigated in Chapter 3 are manifestations of the Temporal Binding Problem: how should different components of a visual stimulus be bound into a coherent percept, when these components have different temporal processing properties? In the case of binocular disparity, the Binding Problem is solved between the two eyes. From this perspective, the Pulfrich effect and its variants are the consequences of inaccurate solutions to the Temporal Binding Problem; signals originating from different points in time are bound together when they should not be, resulting in a visual illusion.

Chapter 4 focuses on a monocular manifestation of the Temporal Binding Problem. For any given visual stimulus presented to one eye, the Temporal Binding Problem needs to be solved, on some level; even a single pixel of light contains a spectrum of light wavelengths. With this fact in mind, we leveraged the target-tracking task described previously to investigate the binding of signals from different type of photoreceptors. We were particularly interested in the known fact that S-cone modulations have longer processing latencies than L-cone modulations. Given that most visual stimuli modulate both cone types to varying degrees, the Temporal Binding Problem is highly relevant here: when a visual stimulus is comprised of both L-cone and S-cone modulations (as well as M-cone modulations, which we have not yet used in our experiments), what is the processing latency of the combined stimulus, and can this latency be predicted?

CHAPTER 2

ABSTRACT

PREDICTING THE PARTITION OF BEHAVIORAL VARIABILITY IN SPEED PERCEPTION WITH NATURALISTIC STIMULI

Benjamin M. Chin

Johannes Burge

A core goal of visual neuroscience is to predict human perceptual performance from natural signals. Performance in any natural task can be limited by at least three sources of uncertainty: stimulus variability, internal noise, and suboptimal computations. Determining the relative importance of these factors has been a focus of interest for decades, but requires methods for predicting the fundamental limits imposed by stimulus variability on sensory-perceptual precision. Most successes have been limited to simple stimuli and simple tasks. But perception science ultimately aims to understand how vision works with natural stimuli. Successes in this domain have proven elusive. Here, we develop a model of humans based on an image-computable (images in, estimates out) Bayesian ideal observer. Given biological constraints, the ideal optimally uses the statistics relating local intensity patterns in moving images to speed, specifying the fundamental limits imposed by natural stimuli. Next, we propose a theoretical link between two key decision-theoretic quantities that suggests how to experimentally disentangle the impacts of internal noise and deterministic suboptimal computations. In several interlocking discrimination experiments with three male observers, we confirm this link, and determine the quantitative impact of each candidate performance-limiting factor. Human performance is near-exclusively limited by natural stimulus variability and internal noise, and humans use near-optimal computations to estimate speed from naturalistic image movies. The findings indicate that the partition of behavioral variability can be

predicted from a principled analysis of natural images and scenes. The approach should be extendable to studies of neural variability with natural signals.

2.1 Introduction

Human beings are adept at many fundamental sensory-perceptual tasks. A sufficiently difficult task, however, can reveal the limits of human performance. A principal aim of perception science and systems neuroscience is to determine the limits of performance, and then to determine the sources of those limits. Performance limits have been rigorously investigated with simple tasks and stimuli(Burgess et al., 1981; Pelli, 1985; Burgess & Colborne, 1988; Geisler, 1989; Dosher & Lu, 1998; Michel & Geisler, 2011; Abbey & Eckstein, 2014)

Ultimately, perception science aims to achieve a rigorous understanding of how vision works in the real world. In natural viewing, there exist at least three factors that limit performance: natural stimulus variability, suboptimal computations, and internal noise. Testing the relative importance of these sources requires two key ingredients: i) an image-computable (images in, estimates out) ideal observer that specifies optimal performance in the task, and ii) experiments that can distinguish the behavioral signatures of each factor. Here, we develop theoretical and empirical methods that can predict and diagnose the impact of each source in mid-level visual tasks with natural and naturalistic stimuli. We investigate the specific task of retinal speed estimation, a critical ability for estimating the motion of objects and the self through the environment.

When a pattern of light falls on the retina, millions of photoreceptors transmit information to the brain about the visual scene. This information is used to build stable representations of image and scene properties (i.e., latent variables) that are relevant for survival and reproduction, like motion speed, three-dimensional position, and object identity. The visual system successfully extracts these critical latent variables from local areas of natural images despite tremendous stimulus variability; infinitely many unique retinal images (i.e. light patterns) are consistent with each value of a given latent variable.

Some image features that vary across different natural images are particularly informative for extracting the latent variable(s) of interest. These are the features that the visual system should encode. Many other image features carry no relevant information. These features should be ignored. (Stimulus variation unrelated to the latent variable is often referred to as ‘nuisance’ variation.) Variation in both the relevant and irrelevant feature spaces can limit performance. But the impact of stimulus variability on performance is minimized only if all relevant features are encoded. Thus, stimulus variability can differentially impact performance depending on the quality of feature encoding.

Signal detection theory posits that sensory-perceptual performance is based on the value of a decision variable(Green & Swets, 1966). But signal detection theory does not specify how to obtain the decision variable from the stimulus. Image-computable observer models do (Adelson & Bergen, 1985; Simoncelli & Heeger, 1998; Schrater et al., 2000; Ziemba et al., 2016; Schütt & Wichmann, 2017; Fleming & Storrs, 2019). Image-computable *ideal* observer models specify how to optimally encode and process the most useful stimulus features(Burgess et al., 1981; Banks et al., 1987; Geisler, 1989; Burge & Geisler, 2011; 2012; 2014; 2015; Sebastian et al., 2017). Image-computable ideal observer models specify how pixels in the image should be transformed into task-relevant estimates (or categorical decisions) that optimize performance in a particular task.

Ideal observers play an important role in the study of perceptual systems because they allow researchers to precisely ask, given the information available to a particular stage of processing, whether subsequent processing stages use that information as well as possible(Geisler, 1989). The explicit description of optimal processing provided by an image-computable ideal observer specifies how natural stimulus variability should propagate into the decision variable given biological constraints. Optimal processing minimizes stimulus-driven nuisance variation in the decision variable. Thus, stimulus variability and the optimal processing jointly set a fundamental limit on performance.

Human performance often tracks the pattern of ideal observer performance, but rarely achieves the same absolute performance levels. It is common to attribute these discrepancies to noise, but discrepancies can also arise from systematically suboptimal computations. To what extent does each factor contribute?

Using complementary computational and experimental techniques we answer this question for a speed discrimination task with naturalistic stimuli. We show that i) natural stimulus variability equally impacts human and ideal performance, ii) the deterministic computations (encoding, pooling, decoding) performed by the human visual system are very nearly optimal, and iii) the humans underperform the ideal near-exclusively because of stochastic internal sources of variability (e.g. late noise), not a systematic misuse of the available stimulus information. The work demonstrates that with appropriate experimental designs, image-computable ideal observer analysis can identify the reasons for human perceptual limits in visual tasks with natural and naturalistic stimuli.

2.2 Materials & Methods

Experimental design and statistical analyses

Three male human observers participated in the experiment; two were authors, and the third was naïve to purposes of the experiment. All had normal or corrected-to-normal acuity. The research protocol was approved by the Institutional Review Board of the University of Pennsylvania and was in accordance with the Declaration of Helsinki. The study was not preregistered. All experiments were performed in MATLAB 2017a using Psychtoolbox version 3.0.12 (Brainard, 1997). Psychophysical data are presented for each individual human observer. Cumulative Gaussian fits of the psychometric functions were in good agreement with the raw data. Bootstrapped or Monte-Carlo-simulated standard errors or confidence intervals are presented on all data points unless otherwise noted. Data will be made available upon reasonable request.

Equipment

Stimuli were presented on a ViewSonic G220fb 40.2cm x 30.3cm cathode ray tube monitor with 1280x1024pixel resolution, and a refresh rate of 60htz. At the at the 92.5cm viewing distance, the monitor subtended a field of view of 24.5x18.6deg of visual angle. The display was linearized over 8 bits of grey level. The maximum luminance was 74cd/m^2. The mean background grey level was set to 37cd/m^2. The observer's head was stabilized with a chin-and-forehead rest.

Stimuli: Detection experiment

Target stimuli in the detection experiment consisted of static, vertically-oriented Gabor targets in cosine-phase (3cpd and 4.5cpd) with 1.5 octave bandwidths embedded in vertically-oriented (1D) dynamic Gaussian noise that was uncorrelated in space and time. Targets subtended 1.0deg of visual angle for a duration of 250ms (15 frames at 60htz). Stimuli were windowed with a raised-cosine window in space and a flattop-raised-cosine window in time, exactly the same as the image movies in the speed discrimination experiment. The RMS contrast of the target and the noise were varied independently according to the experimental design. To minimize target uncertainty, the target was presented to the subject, without noise every 10 trials.

For the detection experiment, a bit-depth of greater than 8 bits is required to accurately measure contrast detection thresholds. We achieved a bit-depth of more than 10 bits using the LOBES video switcher(Li et al., 2003). The video switcher combines the blue channel and attenuated red channel outputs in the graphics card. Picking the right combination of blue and red channel outputs generates a precise gray-scale luminance signal.

Procedure: Detection experiment

Stimuli in the target detection experiment were presented using a two-interval forced choice (2IFC) procedure. On each trial, one interval contained a target plus noise, and the other interval contained noise only. The task was to select the interval containing the target. Feedback was provided. Psychometric functions were measured for each of four different root-mean-squared (RMS) stimulus noise contrasts (0.00, 0.05, 0.10, 0.20) using the method of constant stimuli, with five different target contrasts per condition. Each observer completed 3200 trials in this experiment (4 noise levels x 5 target contrasts per noise level x 80 trials per target x 2 target frequencies). Each block contained 50 trials. To minimize observer uncertainty, trials were blocked by stimulus and noise contrast. The target stimulus was also presented at the beginning of each block, and then again every 10 trials, throughout the experiment.

In target detection tasks, stimulus (e.g. pixel) noise is under experimental control. Internal noise is not. Both noise types influence target detection thresholds. Target contrast power at threshold is a function of stimulus noise $C_T^2\left(\sigma_{pix}\right) \propto \sigma_{pix}^2 + \sigma_{internal}^2$ and is proportional to the sum of pixel and internal noise variances(Burgess et al., 1981); the constant of proportionality depends on the target. This fact can be leveraged to estimate the internal noise that limits detection performance. For example, when stimulus noise and internal noise have equal variance, the squared detection threshold will be twice what it is when pixel noise is zero: $C_T^2\left(\sigma_{pix} = \sigma_{internal}\right) = 2C_T^2\left(\sigma_{pix} = 0\right)$. The amount of stimulus noise required to double thresholds is known as the equivalent input noise. The amount of internal noise that limits performance in a target detection task can therefore be estimated from the pattern of detection thresholds. The estimate of equivalent input noise from the detection experiment sets an upper bound on the amount of early noise in the human visual system (see Results).

Stimuli: Speed discrimination experiment

Natural image movies were created by texture-mapping randomly selected patches of calibrated natural images onto planar surfaces, and then moving the surfaces behind a stationary 1.0deg aperture. The movies were restricted to one dimension of space by vertically averaging each frame of the movie(Burge & Geisler, 2015). Each movie subtended 1.0deg of visual angle. Movie duration was 250ms (15 frames at 60htz). All stimuli were windowed with a raised-cosine window in space and a flattop-raised-cosine window in time. The transition regions at the beginning and end of the time window each consisted of four frames; the flattop of the window in time consisted of seven frames. Contrast was computed under the space-time window. To prevent aliasing, stimuli were low-pass filtered in space and time before presentation (Gaussian filter in frequency domain with σ_{space}=4cpd, σ_{time}=30htz). No aliasing was visible. Training and test sets of naturalistic stimulus movies were generated. The training set had 10,500 unique stimuli (500 stimuli x 21 speeds); the test set had 61,000 unique stimuli (1000 stimuli x 61 speeds). Training stimuli were used to develop the ideal observer (see below). Test stimuli were used to evaluate the ideal and human observers in the speed discrimination experiment.

All stimuli were set to have the same mean luminance as the background and had a RMS contrast of 0.14 (equivalent to 0.20 Michelson contrast for sinewave stimuli), the modal contrast of the stimulus ensemble. The RMS contrast is given by

$$C_{RMS} = \sqrt{\frac{\sum_{\mathbf{x}} \mathbf{c}^2(\mathbf{x})\mathbf{w}(\mathbf{x})}{\sum_{\mathbf{x}} \mathbf{w}(\mathbf{x})}} \tag{1}$$

where $\mathbf{c}(\mathbf{x})$ is a Weber contrast image movie, $\mathbf{w}(\mathbf{x})$ is the space-time window, and $\mathbf{x}=\{x,y,t\}$ is a vector of space-time positions. Stimuli were contrast fixed because contrast is known to affect speed percepts and our focus was on how differences in Weber contrast patterns between stimuli impact performance rather than on differences in overall contrast impact performance, which have already been intensively studied(Thompson, 1982; Weiss et al., 2002).

The short (i.e. 250ms) presentation duration was chosen to approximate the typical duration of a human fixation, and to reduce the possibility that large eye movements would occur while the stimulus was onscreen. For stimuli with speeds and contrasts similar to those used in this experiment, the latencies of smooth pursuit eye movements tend to be 140-200ms(Spering et al., 2005). Saccadic latencies tend to be longer than pursuit latencies.

Procedure: Speed discrimination experiment

For the speed discrimination task, data was collected using a 2IFC procedure. On each trial, a standard and a comparison image movie were presented in pseudo-random order (see below). The task was to choose the interval with the movie having the faster speed. Human observers indicated their choice via a key press. The key press also initiated the next trial. Feedback was given. A high tone indicated a correct response; a low tone indicated an incorrect response. Experimental sessions were blocked by absolute standard speed. In the same block, for example, data was collected at the -5 and +5 deg/sec standard speeds. Movies always drifted in the same direction within a trial, but directions were mixed within a block. An equal number of left- and right-drifting movies were presented in the same block to reduce the potential effects of adaptation.

In each pass of the experiment (see below), psychometric data were measured for each of 10 standard speeds (±5, ±4, ±3, ±2, ±1deg/sec) using the method of constant stimuli. Seven comparison speeds were presented for each standard speed, spanning a range centered on each standard speed. Thus, across the entire experiment, observers viewed stimuli with speeds ranging from 0.25 to 8.00deg/sec. Each standard-comparison speed combination was presented 50 times each for a total of 3,500 trials (2 directions x 5 standard speeds x 7 comparison speeds x 50 trials).

The exact same naturalistic movie was never presented twice within a pass of the experiment. Rather, movies were randomly sampled without replacement from a test set of 1,000 naturalistic movies at each speed. For each standard speed, 350 'standard speed movies' were randomly selected. Similarly, for each of the seven comparison speeds corresponding to that standard, 50 'comparison speed movies' were randomly selected. Standard and comparison speed movies were then randomly paired together. This stimulus selection procedure was used to ensure that the stimuli used in the psychophysical experiment had approximately the same statistical variation as the stimuli that were used to train and test the ideal observer model. Assuming the stimulus sets are representative and sufficiently large, the stimuli presented in the experiment are likely to be representative of natural signals.

Ideal observer for speed estimation

As signals proceed through the visual system, neural states become more selective for properties of the environment, and more invariant to irrelevant features of the retinal images. The ideal observer for speed estimation computes the Bayes' optimal speed estimate from the posterior probability distribution over speed $p\left(X \mid \mathbf{R}\right)$ given the responses $\mathbf{R}$ to a stimulus of a small population of optimal space-time receptive fields (Burge & Geisler, 2015). The receptive fields are assumed to be no larger than the stimulus (i.e. 1.0deg) and to have a temporal integration period no longer than the stimulus duration (i.e. 250ms). No restrictions were placed on the smallest size and shortest integration period of the receptive fields. The receptive fields operate on captured retinal images that include the constraints of the early visual system. The optics of the eye, the spatial sampling, wavelength sensitivity, and temporal integration of the photoreceptors, and response normalization all constrain and shape the information available for further processing. Each natural image movie was convolved with a point-spread function consistent with a 2mm pupil—a typical size on a bright sunny day(Wyszecki & Stiles, 1982)—and the chromatic aberrations of the human eye(Thibos et al., 1992). The temporal integration time of the photoreceptors was approximately 30ms, consistent with direct neurophysiological measurements(Schneeweis & Schnapf, 1995). Receptive field responses were normalized consistent with standard practice(Albrecht & Geisler, 1991; Heeger, 1992; Carandini & Heeger, 2012; Burge & Geisler, 2015; Jaini & Burge, 2017; Sebastian et al., 2017; Iyer & Burge, 2019). Given the constraints imposed by natural stimulus variability and the front-end properties of the early visual system, the space-time receptive fields and the subsequent computations for decoding the speed must be optimal in order for the estimates to be considered optimal. The most useful stimulus features and the computations that optimally pool them are jointly dictated by the task and the stimuli. The receptive fields that encode the most optimal stimulus features for the task are determined via a recently developed technique called Accuracy Maximization Analysis(Geisler et al., 2009; Burge & Jaini, 2017; Jaini & Burge, 2017) (AMA). AMA requires a labeled training set, a model of receptive field response, and a cost function, but requires no parametric assumptions about the shape of the receptive fields. When the training set is representative and sufficiently large, as it is here, the learned receptive fields support equivalent performance on test and training stimulus sets.

The joint response of the set of receptive fields to each stimulus is given by $\mathbf{R} = \mathbf{f}^T(\mathbf{c}+\mathbf{n})/\|\mathbf{c}+\mathbf{n}\|$ where $\mathbf{f}$ is the set of filters, $\mathbf{c}$ is the contrast stimulus, and $\mathbf{n}$ is a sample of early noise. The optimal computations for pooling the responses of the receptive fields are specified by how the receptive field responses are distributed. The conditional receptive field responses $p(\mathbf{R}\,|\,X_k) = gauss(\mathbf{R};\mathbf{0},\Sigma_k)$ are jointly Gaussian and mean zero(Burge & Geisler, 2015; Jaini & Burge, 2017) after response normalization. For any observed response $\mathbf{R}$, the computations that specify the likelihood $L(X_u;\mathbf{R}) = p(\mathbf{R}\,|\,X_u)$ that an observed response was elicited by a stimulus moving with speed X_u is obtained by evaluating the response in the response distribution corresponding to that speed. The responses must therefore be pooled in a weighted quadratic sum, with weights $\mathbf{w}_u$ that are given by simple functions of the covariance matrices Σ_u(Burge & Geisler, 2015). A neuron that performs these quadratic computations outputs a response $R_u^L \propto \exp\left[Q_u(\mathbf{R})\right] = L(X_u;\mathbf{R})$ that is proportional to the likelihood that a stimulus moving at speed X_u elicited the response $\mathbf{R}$. After response (e.g. contrast) normalization(Albrecht & Geisler, 1991; Heeger, 1992; Carandini & Heeger, 2012; Sebastian et al., 2017; Iyer & Burge, 2019), these likelihood neurons instantiate an energy-model-like hierarchical LNLN (linear, non-linear, etc.) cascade(Adelson & Bergen, 1985; Jaini & Burge, 2017). Thus, the computations that yield likelihood neurons can be thought of as a recipe, grounded in natural image and scene statistics, for how to construct speed-tuned neurons that are maximally selective for speed and maximally invariant to natural stimulus (i.e. nuisance) variability. Similar computations yield selective invariant tuning for latent variables like defocus blur, binocular disparity, and three-dimensional motion(Burge & Geisler, 2011; 2012; 2014; 2015).

To obtain the posterior probability of each speed, the likelihood must be weighted by the prior $p\left(X_u\right)$ and normalized by the weighted sum of likelihoods $\sum_v L\left(X_v;\mathbf{R}\right)p\left(X_v\right)$. Finally, the optimal estimate must be 'read out' from the posterior probability distribution. In the case of the 0,1 cost function (i.e. L0 norm) the optimal estimate $\hat{X}_{opt} = \underset{X}{\arg\max}\, p\left(X \mid \mathbf{R}\right)$ is the posterior max. If the prior probability distribution is flat, which it is in the training and test sets, the optimal estimate is the latent variable value that corresponds to the maximum of the likelihood function (i.e. the max of the population response over the likelihood neurons).

Ideal, degraded, and human decision variables

The ideal decision variable for the task of speed discrimination is obtained by the subtracting the optimal speed estimates corresponding to the comparison and standard stimuli

$$D_{ideal} = \hat{X}_{ideal}^{cmp} - \hat{X}_{ideal}^{std} \tag{2}$$

where $\hat{X}_{ideal}^{std}$ and $\hat{X}_{ideal}^{cmp}$ are the ideal observer estimates for the standard and comparison stimuli, respectively. The total variance of the ideal observer decision variable is $2\sigma_{ideal}^2$ where σ_{ideal}^2 is the variance of the ideal observer estimates across stimuli at a given speed. If the decision variable is greater than zero, the ideal observer responds that the comparison stimulus was faster. If the decision variable is less than zero, the ideal observer responds that the comparison stimulus was slower. Degraded observer decision variables are similarly obtained, except that the degraded observer estimates are obtained by reading out the responses of suboptimal receptive fields as well as possible.

The human decision variable is a noisy version of the ideal decision variable, under the hypothesis that human inefficiency is due only to internal sources of variability (e.g. noise). Specifically,

$$D_{human} = D_{ideal} + W \tag{3}$$

where $W \sim N\left(0, 2\sigma_I^2\right)$ is a sample of zero mean Gaussian noise, which corresponds to adding noise with variance σ_I^2 to the comparison and standard stimulus speed estimates.

Double pass experiment

A double pass experiment requires that each observer performs all (or a subset) of the unique trials in an experiment twice. In our experiment, each trial was uniquely identified by its standard and comparison movies. An observer completed the first pass by completing each unique trial once over 20 blocks consisting of 175 trials each. The standard speed was always constant within a block. Blocks were counterbalanced. The observer completed the second pass by completing each unique trial again over another 10 blocks. Before collecting data in the main experiment, each human observer completed multiple practice sessions to ensure that perceptual learning had stabilized. Analysis of the practice data showed no significant learning effects. Stimuli presented in practice sessions were not presented in the main experiment.

Estimating decision variable correlation

Human decision variable correlation is estimated via maximum likelihood from the pattern of human response agreement in the double-pass experiment. The log-likelihood of the double-pass response data is given by

$$\hat{\theta} = \underset{\theta}{\arg\max}\, LL \tag{4}$$

where θ is a vector of model parameters describing decision variable distribution and observer criteria across both passes of the double pass experiment. The log-likelihood of the double-pass response data is given by

$$LL = N^{--}\ln p^{--}\left(\theta\right) + N^{-+}\ln p^{-+}\left(\theta\right) + N^{+-}\ln p^{+-}\left(\theta\right) + N^{++}\ln p^{++}\left(\theta\right) \tag{5}$$

where N^{--} and N^{++} are the number of times that the observer chose standard on both passes or the comparison on both passes, respectively, and N^{-+} and N^{+-} are the number of times that the observer chose the standard on first pass and the comparison on the second and vice versa. The likelihoods of observing those samples are given by

$$p^{--} = \int_{-\infty}^{c_1}\int_{-\infty}^{c_2} gauss\left(\mathbf{D};\mathbf{u},\Sigma\right) \tag{6a}$$

$$p^{-+} = \int_{-\infty}^{c_1}\int_{c_2}^{\infty} gauss\left(\mathbf{D};\mathbf{u},\Sigma\right) \tag{6b}$$

$$p^{+-} = \int_{c_1}^{\infty}\int_{-\infty}^{c_2} gauss\left(\mathbf{D};\mathbf{u},\Sigma\right) \tag{6c}$$

$$p^{++} = \int_{c_1}^{\infty}\int_{c_2}^{\infty} gauss\left(\mathbf{D};\mathbf{u},\Sigma\right) \tag{6d}$$

where $\mathbf{D}$ is the joint decision variable across passes with mean $\mathbf{u}$ and covariance Σ and c_1 and c_2 are the observer criteria on passes one and two. The mean decision variable values are set equal to the speed difference $\mu_1 = \mu_2 = X_{cmp} - X_{std}$ between the standard and comparison stimuli in each condition.

In practice, and without loss of generality, we estimate the decision variable correlation using normalized decision variables $\mathbf{Z}$. The parameter vector for maximizing the likelihood of the normalized decision variables is $\theta = \left\{\rho^*, \mu_1^*, \mu_2^*, c_1^*, c_2^*\right\}$ where * indicates that the parameter is associated with the normalized variable, and ρ is the correlation specified by the covariance Σ. The integrals in Eq. 6a-d can be equivalently expressed with limits of integration $c^* = c/\sigma_{human}$ and integrand $gauss\left(\mathbf{Z};\mathbf{Mu},\mathbf{M}\Sigma\mathbf{M}^T\right)$ with normalized mean and normalized covariance

$$\mathbf{Mu} = \begin{bmatrix} \overbrace{\mu_1/\sigma_{human}}^{\mu_1^*} & \overbrace{\mu_2/\sigma_{human}}^{\mu_2^*} \end{bmatrix}^T \tag{7a}$$

$$\mathbf{M\Sigma M}^T = \begin{bmatrix} 1 & \rho^* \\ \rho^* & 1 \end{bmatrix} \tag{7b}$$

where the normalizing matrix is $\mathbf{M} = \begin{bmatrix} 1/\sigma_{human} & 0 \\ 0 & 1/\sigma_{human} \end{bmatrix}$, and where σ_{human} is the standard deviation of the human estimates. Normalizing the variables has the practical advantage that it converts the covariance matrix to a correlation matrix, so that it can be fully characterized with a single parameter: decision variable correlation. It also sets the normalized means equal to sensitivity d'. We fix the normalized means $\mu_1^* = \mu_2^* = d'_{human}$ to the human sensitivity measured in the discrimination experiment. We also fix the normalized criteria to $c_1^* = c_2^* = 0.0$, which is justified both by the data and the experimental design. These choices reduce the number of parameters to be estimated from five to one.

Efficiency and early noise

Efficiency quantifies the degree to which human performance falls short of ideal performance. The exact expression for efficiency is given by

$$\eta = \left(\frac{d'_{human}}{d'_{ideal}} \right)^2 = \frac{\sigma_{ideal}^2}{\sigma_{human}^2} = \frac{\sigma_E^2 + \sigma_{I,early}^2}{\sigma_{human}^2} \tag{8}$$

where σ_{ideal}^2 and σ_{human}^2 are the variances of the ideal and human speed estimates, and σ_E^2 and $\sigma_{I,early}^2$ are the stimulus-driven and early-noise-driven variances in the ideal speed estimates. Note that the early-noise-driven variance in the estimates—and consequently in the decision variable—is distinct from early noise itself, which is defined in the domain of the image pixels instead of the decision variable. This is analogous to how the stimulus-driven variance in the decision variable is distinct from stimulus variability. Stimulus variability, like early noise, is defined in domain of the image pixels and is non-zero in any set of non-identical stimuli having the same value of the latent variable. We computed efficiency using the exact expression in Eq. 8 and the approximate equality presented in the main text, which assumes that the impact of early noise on the ideal decision variable is negligible (see Results). We found that, because the maximum possible amount of early noise in the system is small (i.e. the upper bound on early noise established by the detection experiment is low), both the exact and the approximate expressions yield similar estimates of efficiency.

2.3 Results

The impact of natural stimulus variability, internal noise, and suboptimal computations can only be distinguished by combining an ideal observer with appropriate behavioral experiments. We examine how these factors impact local motion estimation, a sensory-perceptual ability that is critical for appropriate interaction with the environment(Burge et al., 2019). The plan for the manuscript is diagrammed in Fig. 2.1A. First, we develop an image-computable ideal observer model of retinal speed estimation that is constrained by measurements of natural stimulus variability and early noise. Then we compare human to ideal performance with matched stimuli in two main experiments with matched stimuli. The first main experiment shows that humans track the predictions of the ideal but are consistently less sensitive: one free parameter—efficiency—accounts for the gap between human and ideal performance. We hypothesize that human inefficiency is due to stochastic internal sources of variability (e.g. late noise), and not deterministic sub-optimal computations. This hypothesis predicts that natural stimulus variability should equally limit human and ideal observers. The second main experiment tests this hypothesis. Human observers viewed thousands of trials with naturalistic stimuli in which each unique trial was presented twice. In this paradigm, the repeatability of responses reveals the respective roles of stimulus- and noise-driven variability. If our hypothesis about the source of human inefficiency is correct, efficiency should predict response repeatability with zero additional free parameters. These predictions are confirmed by the experimental data.

An image-computable ideal observer for estimating retinal image speed from local regions of natural images is shown in Fig. 2.1B. Given a set of stimuli, it uses the optimal computations (encoding receptive fields, pooling, decoding) for estimating speed from natural image movies(Burge & Geisler, 2015). The ideal observer thus provides a principled benchmark against which to compare human performance. The tradition in ideal observer analysis is to constrain the ideal observer by stimulus and physiological factors that can be well-characterized and are known to limit the information available for subsequent processing(Geisler, 1989). Natural stimulus variability and early measurement noise are two such factors (red text, Fig. 2.1B). The optimal computations govern how these factors propagate into and determine the variance of the ideal decision variable (Fig. 2.1B). The ideal decision variable controls ideal observer performance.

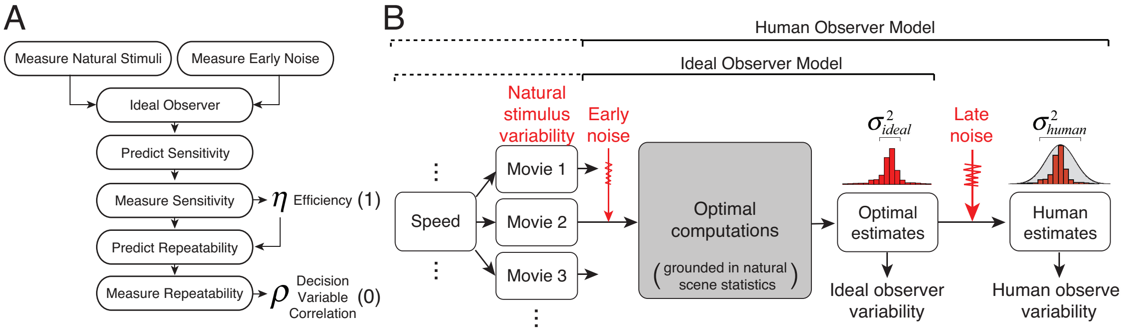

Figure 2.1. Plan for manuscript and ideal observer. **A** Plan for the manuscript. First, we measure natural stimuli and early noise to constrain an ideal observer for speed estimation. Next, we run an experiment and fit the efficiency of each human observer (1 free parameter) by comparing human to ideal sensitivity. Finally, we run a double-pass experiment and show that efficiency predicts human response repeatability and decision variable correlation (0 free parameters). **B** Ideal observer. Speed (i.e. the latent variable) can take on one of many values. Many different image movies share the same speed. The ideal observer is defined by the optimal computations (encoding, pooling, decoding) for estimating speed with natural stimuli. The optimal computations are grounded in natural scene statistics (gray box). For each unique movie, the ideal observer outputs a point estimate of speed. The ideal observer's estimates vary across movies primarily because of natural stimulus variability, variability that is external to the observer. The degraded ideal observer is matched to overall human performance by adding late noise.

Human performance is typically worse than ideal performance. To account for this performance gap, other factors must be considered. We consider suboptimal computations and internal noise, both of which have the potential to increase the variance of the human decision variable relative to the ideal. Suboptimal computations are deterministic, and reflect a systematic misuse of the available stimulus information. Internal noise is random, and is uncorrelated with individual stimuli; although we model it as occurring at the level of the decision variable (see Fig. 2.1B), our methods do not distinguish between different stochastic internal sources of variability (see Discussion). To simultaneously determine the impact of all three factors—natural stimulus variability, suboptimal computations, and internal noise—the ideal observer must be paired with an appropriate psychophysical experiment in which each factor has a distinct behavioral signature. We perform this experiment, and determine the relative importance of each factor. We find that natural stimulus variability and late noise are the primary factors limiting human performance. The impact of suboptimal computations is negligible.

Measuring natural stimuli

A fundamental problem of perception is that multiple proximal stimuli can arise from the same distal cause. This stimulus variability is an important source of uncertainty that limits human and ideal speed discrimination performance. To measure natural stimulus variability, we photographed a large number of natural scenes(Burge & Geisler, 2011; 2015), and then drifted those photographs at known speeds behind a one degree aperture, approximately the size of foveal receptive fields in early visual cortex(Gattass et al., 1981; 1988). This procedure generates motion signals that are equivalent to those obtained by rotating the eye during smooth tracking of a target (Spering et al., 2005; Osborne et al., 2007) (Fig. 2.2A). The sampled set of stimuli approximates, but almost certainly underestimates, the variability present in the natural stimulus ensemble; looming and discontinuous motions, for example, are not represented in our training set(Schrater et al., 2001; Nitzany & Victor, 2014). Thus, the forthcoming estimates of the impact of natural stimulus variability on ideal and human performance are likely to underestimate the impact of stimulus variability on human performance in natural viewing.

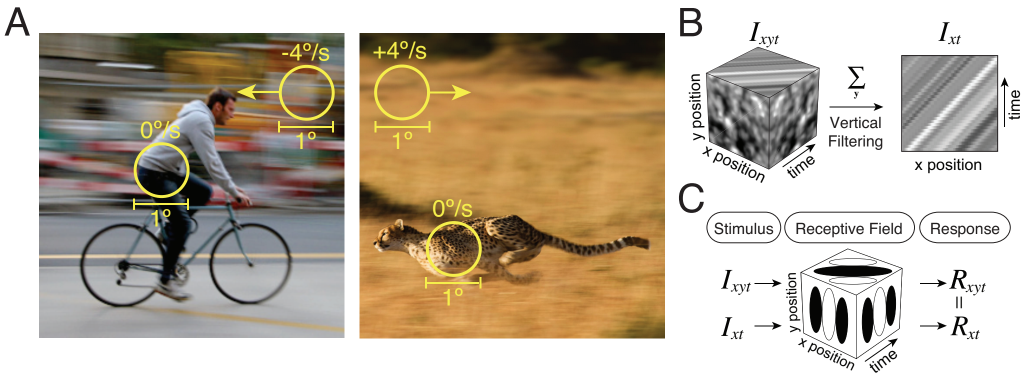

Figure 2.2. Naturalistic image movies and pre-processing. **A** Naturalistic image movies were obtained by drifting photographs of natural scenes at known speeds behind one-degree apertures for 250ms. Rotating the eye in its socket (e.g. tracking an object) creates the same pattern of motion in the stationary background. Optical properties of the eye and the temporal integration of the photoreceptors were also modeled. **B** Full space-time image movies (*Ixyt*) and vertically filtered space-time image movies (*Ixt*). Moving images can be represented as oriented signals in space-time. **C** Vertically oriented receptive fields respond identically to full space-time movies and vertically filtered movies.

Movies drifted leftward or rightward with speeds ranging between 0.25 to 8.0deg/sec. Movies were presented for 250ms, the approximate duration of a typical human fixation. The sampling procedure yielded tens of thousands of unique stimuli (i.e. image movies) at dozens of unique speeds. Image movies were then filtered so that only vertical orientations were present; that is, the stimuli were vertically averaged (i.e. *xt*) versions of full space-time (i.e. *xyt*) movies (Fig. 2.2B). Vertical averaging reduces stimulus complexity, but the resulting stimuli are still substantially more realistic than classic motion stimuli like drifting sinewaves. Furthermore, vertically oriented receptive fields respond identically to vertically averaged and original movies (Fig. 2.2C). Thus, in an individual orientation column, the filtered movies should generate the same response statistics as the full space-time movies(Burge & Geisler, 2015; Jaini & Burge, 2017). Finally, the contrasts of the vertically-averaged stimuli were fixed to the modal contrast in natural scenes (see Discussion). Thus, our stimuli represent a compromise between simple and real-world stimuli, allowing us to run experiments with more natural stimuli without sacrificing quantitative rigor and interpretability. Our analysis should be generalizable to full space-time movies with more realistic forms of motion.

Measuring early noise

All measurement devices are corrupted by measurement noise. The human visual system is no exception. Early measurement noise occurs at the level of the retinal image and places a fundamental limit on how well targets can be detected. Possible sources of early noise include the Poisson variability of light itself and the stochastic nature of the photoreceptor and ganglion cell responses(Hecht et al., 1942). The ideal observer for speed discrimination should be constrained by the same early noise as the human observer if it is to provide an accurate indication of the theoretically achievable human performance limits (see Fig. 2.1A).

Human observers performed a target detection task using the equivalent input noise paradigm(Burgess et al., 1981; Pelli, 1985). The task was to detect a known stationary target embedded in dynamic Gaussian white noise. On each trial, human observers viewed two stimuli in rapid succession, and tried to identify the stimulus containing the target (Fig. 2.3AB). The time-course of stimulus presentation was identical to the forthcoming speed discrimination experiment. Fig. 2.3C shows psychometric functions for target detection in one human observer as a function of target contrast. Each function corresponds to a different noise contrast. Detection thresholds, which are the target contrasts required to identify the target interval 76% of the time (i.e. d-prime of 1.0 in a 2IFC task), are shown for two different targets (3.0 and 4.5 cpd) in Fig. 2.3D. Consistent with previous studies, contrast power at threshold increases linearly with pixel noise(Burgess et al., 1981; Pelli, 1985). Fig. 2.3E shows the same data plotted on logarithmic axes, a common convention in the literature. There are two critical points on this function. The first is its value when pixel noise equals zero, where detection performance is limited only by internal noise. The second is at double the contrast power of the first point—the so-called ‘knee’ of the function—where the pixel noise equals the internal noise. This level of pixel noise is known as the equivalent input noise. Note that the knee of the function, and thus the estimate of equivalent input noise, is robust to whether or not the observer is using a detector (e.g. receptive field) that is optimal for detecting the target.

The equivalent input noise was estimated separately for each target type and human observer. Estimates were consistent across target types and were thus averaged. Noise estimates for the first, second, and third human observers are 2.5%, 2.3% and 2.9%, respectively (Fig. 2.3E). These values are in line with previous reports(Burgess et al., 1981; Pelli, 1985; Williams, 1985).

The estimates of equivalent input noise may reflect the exact amount of early measurement noise alone (Pelli, 1991). The estimates of equivalent input noise may also reflect the combined effect of early measurement noise and noise arising at later processing (e.g. decision) stages. Regardless of which possibility is correct, the target detection experiment provides an upper bound on the amount of early noise in the human visual system. The ideal observer used in the main text is limited by early noise at this upper bound. Because the upper bound is small, early noise only weakly impacts ideal observer performance (see below).

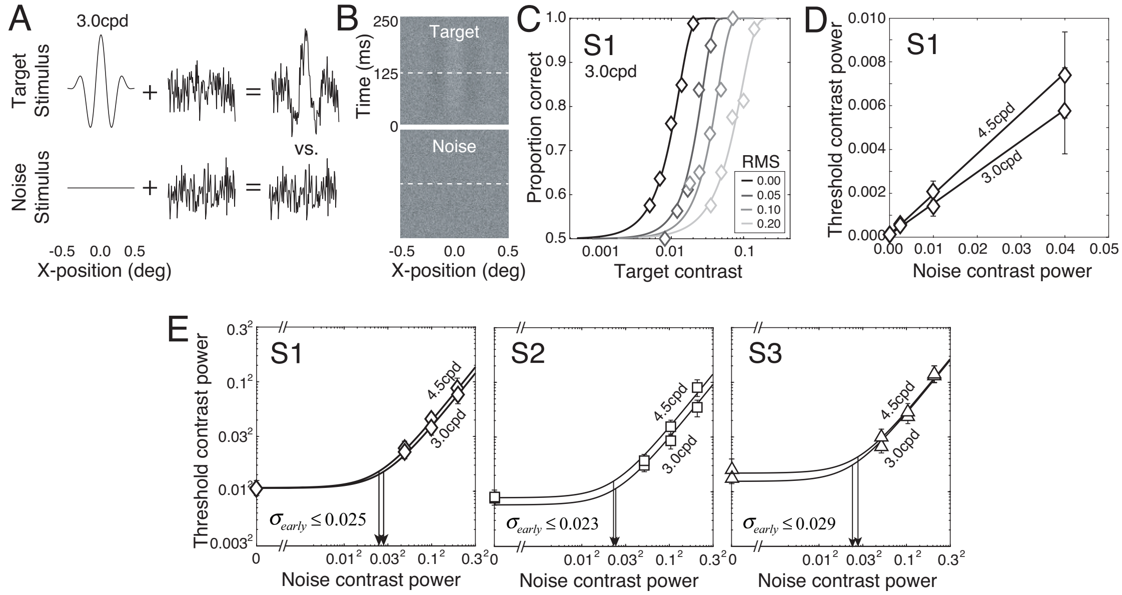

Figure 2.3. Measuring early noise with a target detection experiment**.** **A** Stimulus construction. On each interval, the stimulus was either a stationary target Gabor stimulus or a middle gray field corrupted by dynamic noise. **B** On each trial, the task was to report which of two intervals contained the target stimulus. **C** Psychometric functions from one human observer (S1) for detecting a 3cpd target, in noise having different RMS contrasts (0.00, 0.05, 0.10, 0.20). **D** Threshold target contrast power for the same human observer. Thresholds increase linearly with noise contrast power. Error bars represent 95% bootstrapped confidence intervals; many error bars are smaller than the symbols. **E** Target contrast power at detection threshold plotted on a log-log axis (same data as D) for all three observers. Arrows indicate the estimate of equivalent input noise.

Ideal observer

An ideal observer performs a task optimally, making the best possible use of the available information given stimulus variability and specified biological constraints. In addition to natural stimulus variability and early noise (see Figs. 2.2, 2.3), we model the optics of the eye(Wyszecki & Stiles, 1982; Thibos et al., 1992) , the temporal integration of photoreceptors(Schneeweis & Schnapf, 1995), and the linear filtering(Hubel and Wiesel, 1962) and response normalization(Albrecht & Geisler, 1991; Heeger, 1992; Carandini & Heeger, 2012) of cortical receptive fields. These are all well-established features of early visual processing and determine the information available for subsequent processing.

Assuming the relevant factors have been accurately modeled, ideal observers provide principled benchmarks against which to compare human performance. Given the information available to a particular stage of processing, ideal observers allow the researcher to ask whether subsequent processing stages use that information as well as possible. Humans often track the pattern but fail to achieve the absolute limits of ideal performance. As a consequence, ideal observers often serve as principled starting points for determining additional unknown factors that cause humans to fall short of theoretically achievable performance limits.

Developing an ideal observer with natural stimuli is challenging because it is unclear a priori which stimulus features are most useful for the task. We find the optimal receptive fields for speed estimation using a recently developed Bayesian statistical learning method called Accuracy Maximization Analysis(Geisler et al., 2009; Burge & Jaini, 2017; Jaini & Burge, 2017) (AMA). Given a stimulus set, the method learns the receptive fields that encode the most useful stimulus features for the task (Fig. 2.4A). Once the optimal features are determined, the next step is to determine how to optimally pool and decode the responses $\mathbf{R}=\left[R_1,R_2,\cdots,R_n\right]$ of those receptive fields where n is the total number of receptive fields. Eight receptive fields capture essentially all of the useful stimulus information; additional receptive fields provide negligible improvements in performance(Burge & Geisler, 2015).

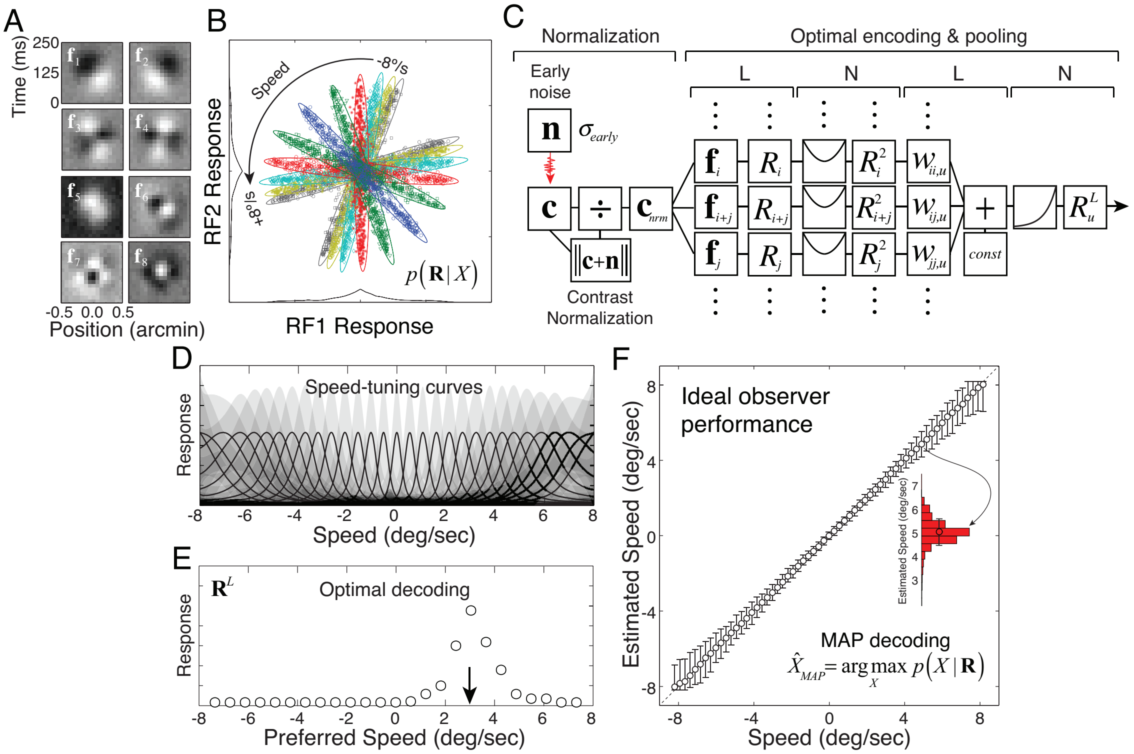

Figure 2.4. Ideal observer receptive fields, response distributions, computations, and estimates. **A** Optimal space-time receptive fields (RFs) for speed estimation given the naturalistic stimulus set and biological constraints. **B** Receptive-field response distributions for RFs 1 and 2, conditioned on the speed of the image movie (colors). Each symbol represents the joint response to an individual movie. The variability of responses for each speed (color) is due to natural stimulus variability; that is, it is the nuisance stimulus variability in the feature space defined by the optimal RFs. **C** Computations of a hypothetical neuron implementing optimal encoding and pooling. Each noisy, contrast-normalized stimulus is processed by the optimal RFs. The responses of these RFs are pooled in a weighted quadratic sum. The weights are determined by the response covariance in B corresponding to the neuron's preferred speed. The response of this hypothetical neuron represents the likelihood that a given stimulus had its preferred speed. The optimal pooling rules thus represent a LNLN (linear, non-linear, etc.) cascade. **D** Speed-tuning curves of hypothetical neurons implementing optimal encoding and pooling, whose responses represent the likelihood of each speed given a stimulus. The speed-tuning curve $\bar{R}^L\left(X_u\right)$ is the average likelihood across stimuli at each of many different speeds. Shaded regions indicate ±1SD confidence intervals on response. This response variability is due to natural stimulus variability. **E** An arbitrary stimulus creates a population response $\mathbf{R}^L$ over hypothetical speed-tuned neurons. Optimal decoding yields the optimal estimate. **F** Ideal observer estimates. The optimal estimate is read out from the population of hypothetical speed-tuned neurons in E, and is equivalent to reading out the posterior probability distribution $p\left(X|\mathbf{R}\right)$ over speed. The variance of ideal observer speed estimates (histogram) is dominated by stimulus-driven variance.

The optimal pooling rules are specified by the joint statistics relating the latent variable and the receptive field responses(Bishop, 2006; Jaini & Burge, 2017). With appropriate response normalization, the responses across stimuli for each speed are conditionally Gaussian(Lyu & Simoncelli, 2009; Burge & Geisler, 2015; Sebastian et al., 2017; Iyer & Burge, 2019) (Fig. 2.4B). To obtain the likelihood of a particular speed, the

Gaussian response statistics require that the receptive field responses to a given stimulus be pooled via weighted quadratic summation (see Fig. 2.4C). The computations for computing the likelihood thus instantiate an enhanced version of the motion-energy model, indicating that energy-model-like computations are the normative computations supporting speed estimation with natural stimuli(Adelson & Bergen, 1985; Jaini & Burge, 2017). The speed tuning curves of hypothetical neurons implementing these computations are approximately log-Gaussian, similar to the approximately log-Gaussian speed tuning curves of neurons in area MT(Nover et al., 2005) (Fig. 2.4D). Finally, an appropriate read out of the population response of these hypothetical neurons is equivalent to decoding the optimal estimate from the posterior probability distribution $p\left(X\,|\,\mathbf{R}\right)$ over speed (Fig. 2.4EF). If a 0,1 cost function is assumed, the latent variable value corresponding to the maximum of the posterior is the optimal estimate. We have previously verified that reasonable changes to the prior and cost function do not appreciably alter the optimal receptive fields, pooling rules, or performance(Burge & Jaini, 2017). This approach provides a recipe for how to construct neurons that are highly invariant to nuisance stimulus variability and tightly tuned to speed. It also provides a normative justification, grounded in natural scene statistics, for descriptive models proposed to account for response properties of neurons in cortex(Adelson & Bergen, 1985; Simoncelli & Heeger, 1998; Perrone & Thiele, 2001; Nover et al., 2005; Rust et al., 2006; Jaini & Burge, 2017).

The factors thus far described in the paper—stimulus variability and early noise, biological constraints, and the optimal computations (encoding, pooling, decoding)—all impact ideal performance in our task. Given a particular stimulus set, the only factor subject to some uncertainty is the precise amount of early noise. However, within the bound set by the detection experiment (see Fig. 2.3), different amounts of early noise have only a minor effect on ideal performance (see below). Thus, estimates of ideal performance are set overwhelmingly by stimulus variability.

Measuring efficiency

The ideal observer benchmarks how well humans use the stimulus information available for the task. Efficiency quantifies how human sensitivity d'_{human} compares to ideal observer sensitivity d'_{ideal} and is given by

$$\eta = \left(\frac{d'_{human}}{d'_{ideal}} \right)^2 = \frac{\sigma^2_{ideal}}{\sigma^2_{human}} \cong \frac{\sigma^2_E}{\sigma^2_{human}} \qquad (9)$$

where σ^2_{human} is the total variance of the human decision variable, σ^2_{ideal} is the total variance of the ideal decision variable, and σ^2_E is the stimulus-driven component of the ideal decision variable. The third approximate equality in Eq. 9 assumes that stimulus-driven variability equals ideal observer variability because the impact of early noise is bounded to be small (c.f. Fig. 2.3).

To measure human sensitivity, we ran a two-interval forced choice (2IFC) speed discrimination experiment. On each trial, human observers viewed two moving stimuli in rapid succession, and indicated which stimulus was moving more quickly (Fig. 2.5A). This design is similar to classic psychophysical experiments with one critical difference. Rather than presenting the same (or very similar) stimuli in each condition hundreds of times, we present hundreds of unique stimuli one time each. This stimulus variability jointly limits human and ideal performance. Human sensitivity is computed using standard expressions from signal detection theory $d'_{human} = \sqrt{2}\Phi^{-1}\left(PC_{human}\right)$ where PC_{human} is the proportion of times that the comparison is chosen in a given condition in a 2IFC experiment and $\Phi^{-1}\left(\cdot\right)$

is the inverse cumulative normal. (This expression is correct assuming the observer uses the optimal criterion, an assumption that is justified by the data.)

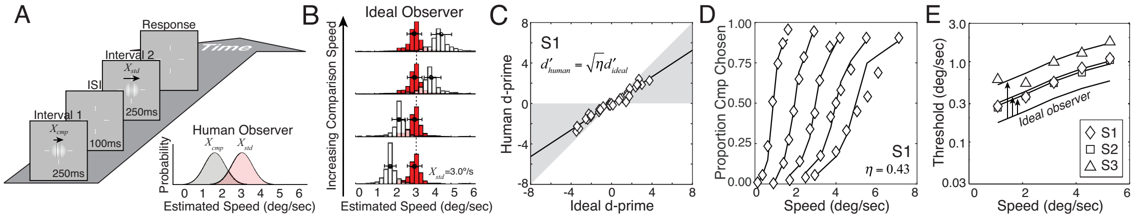

Figure 2.5. Measuring speed discrimination. **A** The task in a two-interval forced choice experiment was to report the interval containing the faster of two naturalistic image movies. Unlike classic psychophysical studies, which present the same stimuli hundreds of times, the current study presents hundreds of unique stimuli one time each. This design injects naturalistic stimulus variability into the experiment. Human responses are assumed to be based on samples from decision variable distributions (inset). **B** Ideal observer estimates across hundreds of standard (red) and comparison movies (white) at one standard speed (3 deg/sec) and four comparison speeds. **C** Human vs. ideal observer sensitivity for all standard and comparison speeds. Shaded regions mark regions of plot where humans are less efficient than ideal but are still performing the task. For all conditions, humans are less sensitive than the ideal observer by a single scale factor: efficiency: $d'_{human} = \sqrt{\eta} d'_{ideal}$. Negative d-primes correspond to conditions in which the comparison was slower than the standard. **D** Psychometric functions for one human observer (symbols) at five standard speeds. The degraded ideal observer (solid curves) matches the efficiency of the human observer (one parameter fit to human data). **E** Human speed discrimination thresholds (d-prime = 1.0) as a function of standard speed for three human observers (symbols) on a semi-log plot. The pattern of human thresholds matches ideal observer thresholds (solid curve). Vertically shifting the ideal observer thresholds by an amount set by each human's efficiency (arrows) shows degraded observer performance (solid curves, one free parameter fit per human).

To measure ideal sensitivity, we ran the ideal observer in a simulated experiment with the same stimuli as the human. (Note that the ideal observer was trained on different stimuli than the human and ideal observers were tested on.) Ideal sensitivity (i.e. d-prime) was computed directly from the distributions of ideal observer speed estimates in each condition (Fig. 2.5B). Human and ideal sensitivities across all speeds are linearly related (Fig. 2.5C). Rearranging Eq. 9 shows that human sensitivity $d'_{human} = \sqrt{\eta} d'_{ideal}$ equals the ideal observer sensitivity degraded (scaled) by the square root of the efficiency. Thus, a single free parameter (efficiency) relates the pattern of human and ideal sensitivities for all conditions. The efficiencies of the first, second, and third human observers are 0.43, 0.41, and 0.17, respectively.

Transforming the sensitivity data back into percent comparison chosen shows that the details of the degraded ideal nicely account for the human psychometric functions (Fig. 2.5D). The psychometric functions can be summarized by the speed discrimination thresholds (d-prime = 1.0; 76% correct in a 2IFC task). The pattern of human and ideal thresholds match; the proportional increases of the human and ideal threshold functions with speed are the same (Fig. 2.5E). These results quantify human uncertainty σ_{human}^{2}, show that an ideal observer analysis of naturalistic stimuli predicts the pattern of human speed discrimination performance, and replicate our own previously published findings(Burge & Geisler, 2015).

Together, the ideal observer and speed discrimination experiment reveal the degree of human inefficiency (i.e. how far human performance falls short of the theoretical ideal). But they cannot determine the sources of this inefficiency. Humans could be inefficient because of late noise (i.e. stochastic internal sources of variability arising after early noise). Humans could also be inefficient because of fixed suboptimal computations. If inefficiency is due exclusively to late noise, stimulus variability must equally limit human and ideal observer performance. If human inefficiency is partly due to suboptimal computations, stimulus variability will cause more stimulus-driven variability in the human than in the ideal. How can human behavioral variability be partitioned to determine the sources of inefficiency in speed perception? To do so, additional experimental tools are required.

Predicting and measuring decision variable correlation

A double pass experiment, when paired with ideal observer analysis, can determine why human performance falls short of the theoretical ideal. In a double pass experiment(Burgess & Colborne, 1988; Gold et al., 2004; Li et al., 2006), each human observer responds to each of a large number of unique trials (the first pass), and then performs the entire experiment again (the second pass). Double pass experiments can 'unpack' each point on the psychometric function (Fig. 2.6AB), providing far more information about the factors driving and limiting human performance than standard single pass experiments. The correlation in the human decision variable across passes—decision variable correlation—is key for identifying the factors that limit performance and determine efficiency(Burgess & Colborne, 1988; Sebastian & Geisler, 2018).

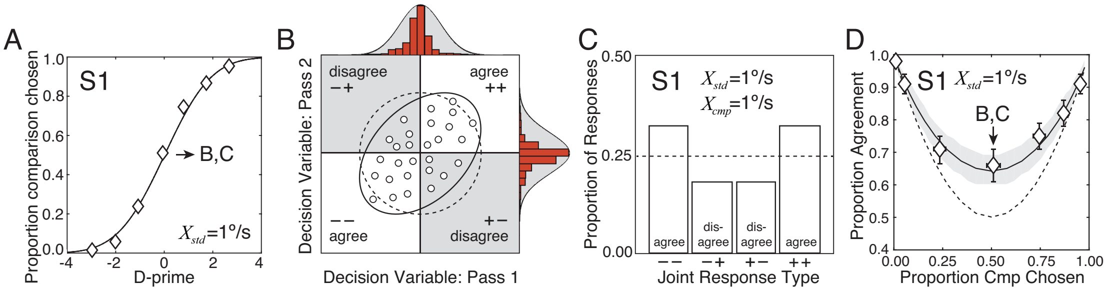

Figure 2.6. Decision variable correlation and response repeatability in a double pass experiment. **A** Psychometric data from the first human observer and cumulative Gaussian fit plotted as proportion comparison chosen vs. d-prime for the standard speed of 1 deg/sec. (Same data as in Fig. 2.5D.) **B** Schematic for visualizing decision variable correlation across passes when standard and comparison speeds are identical (e.g. both equal 1 deg/sec). Samples correspond to individual double pass trials (small circles). The value of each sample represents the difference between the estimated speeds of the comparison and standard stimuli on each trial. Decision variable values corresponding to response agreements and disagreements fall in white and gray quadrants, respectively. Decision variable distributions with the decision variable correlation predicted by efficiency (solid ellipse) and by the null model with a decision variable correlation of zero (dashed ellipse). Decision variable correlation depends on the relative importance of correlated and uncorrelated factors across passes. Stimuli are correlated on each repeated trial of a double pass experiment; internal noise is not. Criteria on each pass (vertical and horizontal lines, respectively) are assumed to be optimal and at zero. **C** Predicted response counts (bars) for each response type (--, -+, +-, ++) across passes (100 trials per condition) given the decision variable correlation shown in B. **D** Proportion of trials on which responses agreed across both passes of the double pass experiment as a function of proportion comparison chosen for one human observer. Agreement data (symbols) and prediction (solid curve) assuming that efficiency predicts decision variable correlation (i.e. that all human inefficiency is due to late noise). The null prediction assumes that the decision variable correlation across passes is zero (dashed curve). The agreement data is predicted directly from the efficiency of the human observer (zero free parameters). Error bars represent 68% bootstrapped confidence intervals on human agreement. Shaded regions represent 68% confidence intervals from 10000 Monte Carlo simulations of the predicted agreement data assuming 100 trials per condition.

The power of this experimental design is that it enables behavioral variability to be partitioned into correlated and uncorrelated factors. Factors that are correlated across passes, like the stimuli, increase the correlation of the decision variable across passes. Factors that are uncorrelated across passes, like internal noise, decrease decision variable correlation. If the variance of the human decision variable is dictated only by stimulus-driven variability, decision variable correlation will equal 1.0. If the variance of the human decision variable is dictated only by internal noise, decision variable correlation will equal 0.0. If both stimulus-driven variability and internal noise play a role, the correlation will have an intermediate value.

Decision variable correlation, like the decision variable itself, cannot be measured directly using standard psychophysical methods. Rather, it must be inferred from the repeatability of responses across passes in each condition. The higher the decision variable correlation, the greater the proportion of times responses agree (i.e. repeat) in a given condition (Fig. 2.6BC).

In each condition, we used the pattern of response agreement to estimate decision variable correlation (Fig. 2.6BC), and then plotted agreement against the proportion of times the human observer (symbols) chose the comparison stimulus as faster (Fig. 2.6D). Human response agreement implies a decision variable correlation that is significantly different from zero. For the seven conditions shown in Fig. 2.6D (i.e. all comparison speeds at the 1 deg/sec standard speed), the maximum likelihood fit of decision variable correlation across the seven comparison levels is 0.43. Thus, 43% of the total variance in the human decision variable is due to factors that are correlated across repeated presentations of the same trials.

How should the estimate of decision variable correlation be interpreted? Human decision variable correlation across passes is given by

$$\rho = \frac{\sigma_E^2}{\sigma_E^2 + \sigma_I^2} = \frac{\sigma_E^2}{\sigma_{human}^2} \tag{10}$$

where σ_E^2 is the variance of the speed estimates due to external (i.e. stimulus) factors, σ_I^2 is the variance due to internal factors (e.g. noise), and σ_{human}^2 is the total variance of the human speed estimates. Decision variable correlation is driven by stimulus variation, because the stimuli are perfectly correlated across passes.

The estimated decision variable correlation is strikingly similar to the efficiency measured for each observer. Although the exact relationship between decision variable correlation and efficiency depends on the source of human inefficiency, the fact that they are similar is no accident. Under the hypothesis that all human inefficiency is due to noise (i.e. stochastic internal factors that are uncorrelated with the stimuli), stimulus variability must impact human and ideal observers identically: the stimulus-driven variance in the human speed estimates (σ_E^2 in Eq. 10) will equal the stimulus-driven variance in the ideal observer speed estimates (σ_E^2 in Eq. 9). Plugging Eq. 9 into Eq. 10 shows that, under the stated hypothesis, human decision variable correlation equals efficiency

$$\rho = \eta \tag{11}$$

This mathematical relationship has important consequences. It means that the estimate of human efficiency from the speed discrimination experiment (Fig. 2.5C) provides a zero-free parameter prediction of human decision variable correlation in the double pass experiment (Fig. 2.6). The behavioral data confirm this prediction. Human efficiency in the discrimination experiment quantitatively predicts human response agreement in the double-pass experiment (Fig. 2.6D; symbols vs. solid curve). The implication of this result is striking. It suggests that natural stimulus variability equally limits human and ideal observers and thus that the source of human inefficiency is due near-exclusively to late noise. Human speed discrimination is therefore optimal except for the impact of late internal noise.

These results generalize across all conditions and human observers. Fig. 2.7A shows measured response agreement vs. proportion comparison chosen for the first human observer in each of the five standard speed conditions. Fig. 2.7B plots measured response agreement against efficiency-predicted agreement, summarizing the agreement data for each human observer across all standard speeds; prediction uncertainty given the number of double-pass trials in each condition is shown as 95% confidence intervals (shaded regions). The decision variable correlations that best account for the response repeatability across all conditions of the first, second, and third human observers are 0.45, 0.43, and 0.18, respectively. For the first two observers, stimulus-driven variance and noise variance have approximately same magnitude. For all observers, the data is consistent with the hypothesis that decision variable correlation equals efficiency (solid curves), and is not consistent with the null model in which decision variable correlation equals zero (dashed curves). Fig. 2.7C plots decision variable correlation against efficiency for each human observer. Efficiency tightly predicts decision variable correlation for all three human observers, with zero additional free parameters.

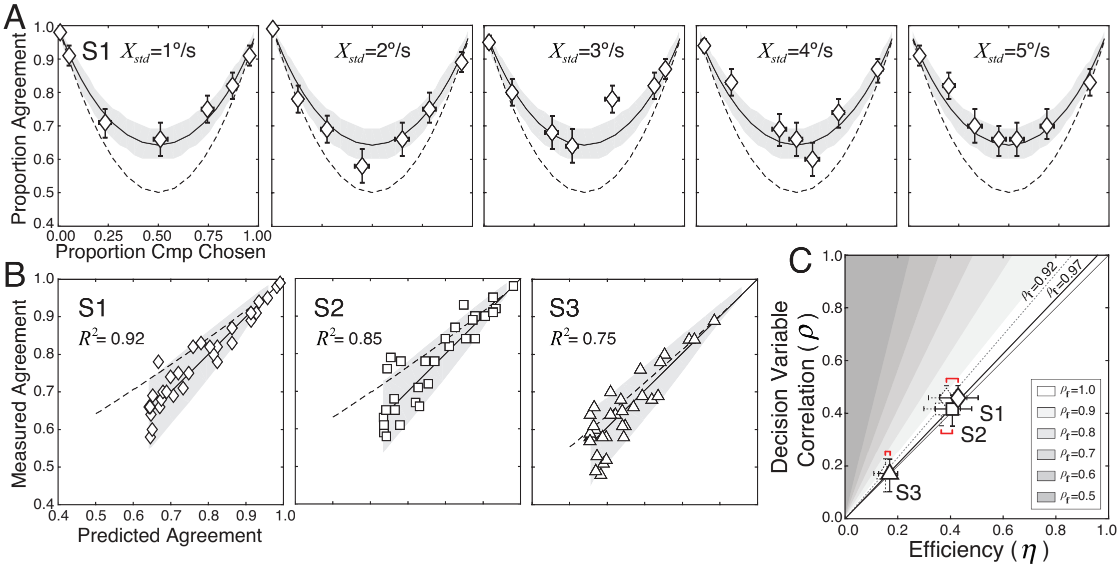

Figure 2.7. Predicted vs. measured response agreement and decision variable correlation. **A** Proportion response agreement vs. proportion comparison chosen for all five standard speeds (1-5deg/sec), for the first human observer. Human data (symbols) and predictions (curves) are shown using the same conventions as Fig. 2.6D. **B** Measured vs. predicted response agreement for all conditions and all human observers (symbols). Human agreement equals efficiency-predicted agreement for all three human observers (solid line); shaded regions indicate 95% confidence intervals on the prediction from 1000 Monte Carlo simulations. Efficiency-predicted agreement for the null model, which assumes decision variable correlation is zero, is also shown (dashed curve). **C** Decision variable correlation vs. efficiency for each human observer (symbols). Human efficiency, measured in first pass of the speed discrimination experiment, tightly predicts human decision variable correlation in the double pass experiment with zero free parameters. Error bars represent 95% bootstrapped confidence intervals on human efficiency and on human decision variable correlation. Shaded regions show the expected relationship between efficiency and decision variable correlation if humans use fixed suboptimal computations (i.e. sub-optimal receptive fields). Red brackets indicate uncertainty about

the precise value of efficiency due to uncertainty about the precise amount of early noise (see Fig. 2.3). Solid and dashed black lines are the best-fit regression lines, corresponding to receptive field correlations of 0.97 and 0.92, respectively.

These results must be interpreted with some caution. Uncertainty about the amount of early measurement noise can cause uncertainty about human efficiency (Eq. 8) and thus about the predicted decision variable correlation (Eq. 11). We simulated ideal observers with different amounts of early noise and computed efficiency for each human observer (Fig. 2.8A). Fortunately, the detection experiment establishes an upper bound on the amount of early noise for each human observer (c.f. Fig. 2.3), thereby constraining the uncertainty about the predicted decision variable correlation (Fig. 2.8B; red brackets). Because the upper bound on early noise is low, the maximum and minimum possible efficiencies differ by approximately 10% depending on whether early noise at the upper or lower bound is assumed (Fig. 2.8AB; red brackets). The measured decision variable correlation values (Fig. 2.8C) are in line with the predictions. Thus, uncertainty about the amount of early noise has only a minor impact on the interpretation of our results.

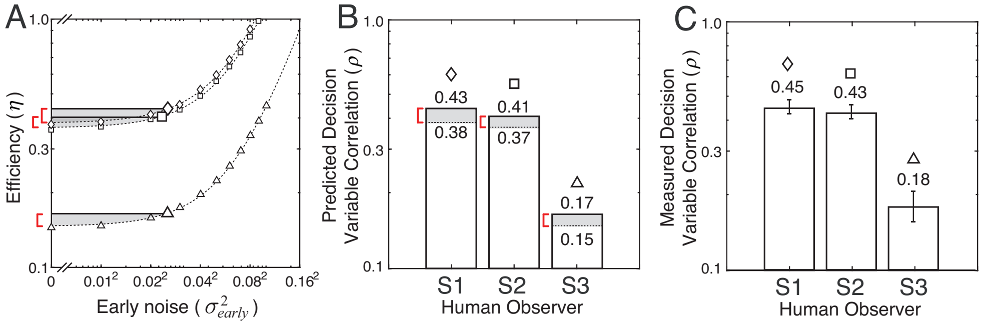

Figure 2.8. Early noise, efficiency, and predicted decision variable correlation. **A** Efficiency in speed discrimination for each human observer (symbols) changes as a function of the amount of early noise modeled in the ideal observer. If early noise is negligible, efficiency is given by $\eta = \sigma^2_E / \sigma^2_{human}$ (Eq. 9). If early noise is non-negligible, efficiency is given by $\eta = \left(\sigma^2_E + \sigma^2_{I,early}\right) / \sigma^2_{human}$ (Eq. 8). The red brackets and shaded regions indicate the minimum and maximum human efficiencies, given the bound on early noise established by the detection experiment (c.f. Fig. 2.3). **B** Predicted decision variable correlation for each human observer given the uncertainty about human efficiency. The maximum (solid line) and minimum (dashed line) predicted decision variable correlations correspond to ideal observers having the maximum and minimum amount of early noise. The predicted decision variable correlations differ by ~10% at maximum. **C** Measured decision variable correlation for each human observer. Error bars are 95% bootstrapped confidence intervals.

In the best performing observers, natural stimulus variability accounts for nearly half of all behavioral variability, despite the fact that the naturalistic stimulus set used to probe speed discrimination performance almost certainly underestimates the importance of stimulus variability in natural viewing (see Discussion). External variability therefore shapes the optimal computations, dictates the pattern of human performance, and predicts the partition of behavioral variability (i.e. the relative importance of external and internal sources of variability). These findings motivate continued efforts to model and characterize how natural stimulus variability impacts neural and perceptual performance in natural tasks.

Suboptimal computations

Human efficiency equals human decision variable correlation (Figs. 2.7C; 2.8BC). To confidently conclude from this result that human inefficiency is almost entirely due to noise (i.e. stochastic internal sources of variability), we must rule out the possibility that suboptimal computations can produce the same result. How do fixed suboptimal computations impact the relationship between efficiency and decision variable correlation? To answer this question, one must determine how suboptimal computations impact the stimulus-driven component of the decision variable. To do so, we analyzed the estimates of a degraded observer that suboptimally encodes stimulus features(Burgess et al., 1981; Dosher & Lu, 1998; Neri & Levi, 2006; Sebastian & Geisler, 2018). If the wrong features are encoded, informative features may be missed, irrelevant features may be processed, and the variance of the stimulus-driven component of the decision variable may be increased relative to the ideal.

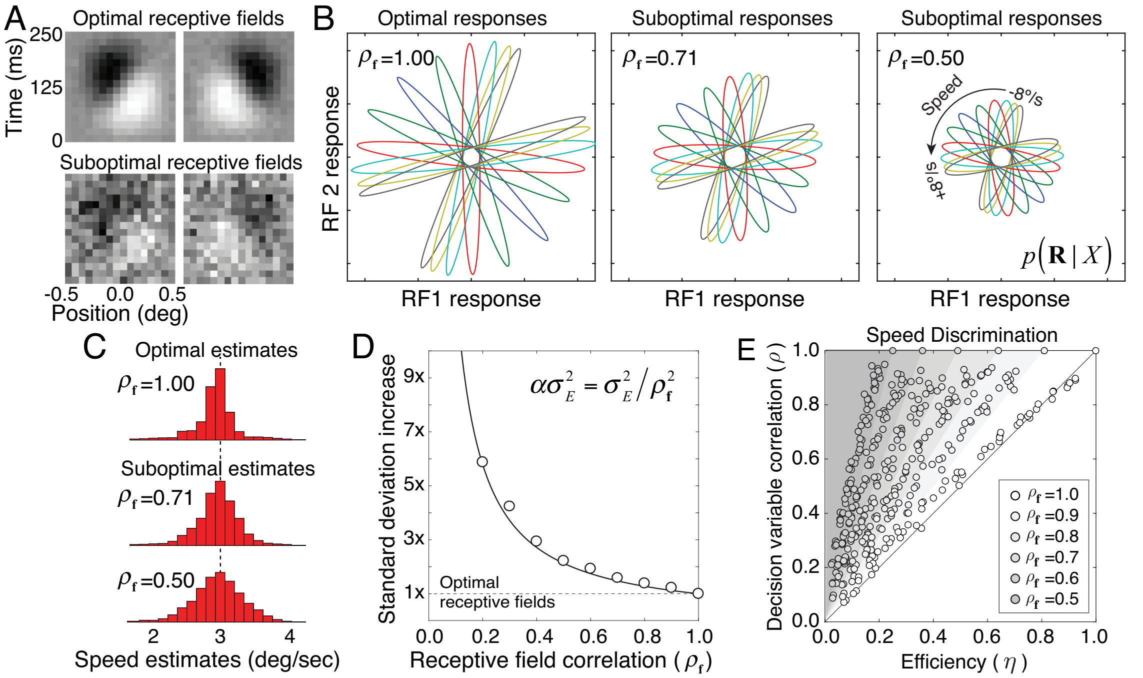

Figure 2.9. Relationship between suboptimal receptive fields and stimulus-driven variability in degraded observers. **A** Optimal receptive fields (top; also see Fig. 2.4A) and suboptimal receptive fields from the degraded observer (bottom); only the first two receptive fields of each observer are shown. To obtain a suboptimal receptive field with a particular receptive field correlation ρ_f, we added fixed samples of Gaussian white noise to the corresponding optimal receptive field. The variance of the corrupting noise is given by $\sigma^2_{corrupt} = \left(\left(1/\rho_f^2\right)-1\right)/N$ where N is the number of pixels defining each receptive field. **B** Impact of suboptimal receptive fields on the conditional response distributions $p(\mathbf{R} \mid X)$. As the receptive fields become more suboptimal, the response distributions (colored ellipses) more poorly distinguish different values of the latent variable (colors). **C** Effect of suboptimal receptive fields on degraded observer speed estimates for movies drifting at one speed (3 deg/sec). As receptive field correlation decreases, the stimulus-driven variance of the estimates increases, because informative stimulus features are not encoded and uninformative features are. **D** The proportional increase of stimulus-driven standard deviation for degraded vs. the ideal observer estimates, assuming that the degraded observer has no late internal noise. Symbols plot the mean result from 100 Monte Carlo simulations. The stimulus-driven variance of the speed estimates increases with the squared inverse of receptive field correlation. **E** Relationship between decision variable correlation and efficiency for degraded observers with different combinations of fixed suboptimal computations (i.e. receptive field correlations; gray levels) and internal noise. Points represent mean decision variable correlation and mean efficiency from 100 Monte Carlo simulations of each degraded observer.

To create suboptimal feature encoders (i.e. suboptimal receptive fields), we corrupted the optimal receptive fields with fixed samples of Gaussian white noise (Fig. 2.9A). Receptive field correlation (i.e. cosine similarity) quantifies the degree of sub-optimality $\rho_f = \mathbf{f}_{opt}^T \mathbf{f}_{subopt} \Big/ \left(\left\| \mathbf{f}_{opt}^T \right\| \left\| \mathbf{f}_{subopt} \right\| \right)$ where $\mathbf{f}_{opt}$ and $\mathbf{f}_{subopt}$ are the optimal and suboptimal receptive fields, respectively. Compared to the responses of the optimal receptive fields, the responses of these suboptimal receptive fields segregate less well as a function of speed (Fig. 2.9B). We generated degraded observers with suboptimal receptive fields having different receptive field correlations and examined estimation performance (Fig. 2.9C). We found that the stimulus-driven variance $\alpha\sigma_E^2$ of the degraded observer estimates is a scaled version of the ideal stimulus-driven variance; the scale factor $\alpha = 1/\rho_{\mathbf{f}}^2$ is equal to the squared inverse of receptive field correlation (Fig. 2.9D). Thus, suboptimal receptive fields systematically increase the variance of the stimulus-driven component of the decision variable.

If humans are well modeled by a degraded observer with both suboptimal receptive fields and late noise, the total variance of the human estimates is given by $\sigma_{human}^2 = \alpha\sigma_E^2 + \sigma_I^2$. Replacing terms in Eqs. 9 and 10 and performing some simple algebra shows that the relationship between efficiency and decision variable correlation is given by

$$\rho = \alpha\eta = \frac{\eta}{\rho_{\mathbf{f}}^2} \qquad (12)$$

Thus, with sub-optimal computations (i.e. receptive fields) decision variable correlation will be systematically larger than efficiency by the inverse square of receptive field correlation. (Note that when receptive field correlation equals 1.0, Eq. 12 reduces to Eq. 11.) For example, if receptive field correlation is 0.5, decision variable correlation is 4x higher than efficiency. We verified the relationship between decision variable correlation and efficiency by simulating degraded observers with different levels of suboptimal computations and late noise (Fig. 2.9E). As predicted by Eq. 12, the more suboptimal the computations (i.e. receptive field correlations), the more decision variable correlation exceeds efficiency. We reanalyzed our results in the context of Eq. 12, comparing the behavioral data to the predictions of various degraded observer models.

For all three observers, decision variable correlation is larger than efficiency by ~5%, corresponding to a receptive field correlation of 0.97 (Fig. 2.7C). (Note that these numbers assume an ideal observer with early noise set to the upper bound established by the detection experiment (see Fig. 2.3). If no early noise is assumed, then decision variable correlation exceeds efficiency by 15%, corresponding to a receptive field correlation of 0.92; Fig. 2.7C). Thus, no more than 15% of human inefficiency can be attributed to fixed suboptimal computations.

Note that the simulations just described only consider the potential impact of fixed suboptimal computations that are linear. We cannot definitively rule out non-linear suboptimal computations that leave stimulus-driven variability unchanged while selectively amplifying the impact of early noise, making amplified early noise indistinguishable from late noise. However, such computations are highly unlikely, given current knowledge of early visual processing. More importantly, suboptimal computations that selectively amplify early noise will not alter the predicted relationships between efficiency and decision variable correlation.

Thus, our results imply that the deterministic computations performed by the human visual system in speed estimation are very nearly optimal. Although natural stimulus (i.e. nuisance) variability is a major and unavoidable factor that limits performance in natural viewing, its impact is minimized as much as possible by the computations performed by the visual system.

Stimulus variability and behavioral variability

In this paper, we have shown that natural stimulus variability limits behavioral performance and drives response repeatability. Thus, reducing stimulus variability should increase sensitivity (i.e. improve behavioral performance) but decrease response repeatability. To test this prediction, we ran a new speed discrimination experiment using drifting random-phase sinewave gratings (Fig. 2.10). A stimulus set composed of drifting sinewaves has less variability than the set of naturalistic stimuli used in the main experiment. As predicted, with sinewave stimuli human sensitivity improves (Fig. 2.10A), responses become less repeatable (Fig. 2.10B), and decision variable correlation is lower (Fig. 2.10C). Interestingly, reducing stimulus variability affects decision variable correlation in the third human observer less than it does in the first two. This is the expected pattern of results given that the third observer (S3) had low decision variable correlation with naturalistic stimuli and was thus already dominated by internal noise (see Figs. 2.7C, 2.8C). However, not all of the results were quite as expected. We anticipated that decision variable correlation would equal zero for all three human observers with sinewave stimuli. But decision variable correlation exceeded zero for all three observers. What accounts for this discrepancy? We have ruled out commonly considered trial order effects (e.g. feedback-based effects) as the cause (Laming, 1979), but we are unsure of the cause. Whatever the case, with reduced stimulus variability, internal noise—which is uncorrelated across stimulus repeats—becomes the dominant source of variability limiting performance in all human observers.

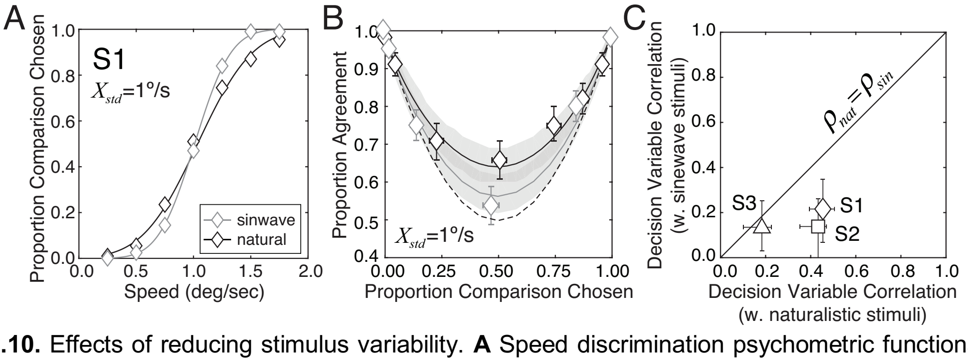

Figure 2.10. Effects of reducing stimulus variability. **A** Speed discrimination psychometric functions for the first human observer with naturalistic stimuli (black curve) and drifting sinewave stimuli (gray curve) for a 1 deg/sec standard speed. Sinewave stimuli can be discriminated more precisely. **B** Proportion response agreement vs. proportion comparison chosen for naturalistic stimuli (black) and artificial stimuli (grey) for the same human observer. **C** Decision variable correlation with artificial stimuli vs. decision variable correlation with naturalistic stimuli for each human observer (symbols). Error bars represent 95% bootstrapped confidence intervals. Decision variable correlation is consistently lower when artificial stimuli are used.

2.4 Discussion

Simple stimuli and/or simple tasks have dominated behavioral neuroscience because of the need for rigor and interpretability in assessing stimulus influences on neural and behavioral responses. The present experiments demonstrate that, with appropriate techniques, the required rigor and interpretability can be obtained with naturalistic stimuli. We have shown that image-computable ideal observers can be fruitfully combined with human behavioral experiments to reveal the factors the limit behavioral performance in mid-level tasks with naturalistic stimuli. In particular, an image-computable ideal observer, constrained by the same factors as the early visual system, predicts the pattern of human speed discrimination performance with naturalistic stimuli(Burge & Geisler, 2015). Perhaps more remarkably, human efficiency in the task predicts human decision variable correlation in a double pass experiment without free parameters, a result that holds only if the deterministic computations performed by humans are very nearly optimal.

Limitations and future directions

One limitation of our approach, which is common to most psychophysical approaches, is that it cannot pinpoint the processing stage or brain area at which the limiting source of internal variability arises. Although we model it as noise occurring at the level of the decision variable, it could also occur at the encoding receptive field responses, the computation of the likelihood, the readout of the posterior into estimates, the placement of the criterion at the decision stage, or some combination of the above. We believe we have ruled out the possibility that the noise limiting speed discrimination is early (Fig. 2.3). But we cannot distinguish amongst other stochastic sources of internal variability. These issues are probably best addressed with neurophysiological methods. Similarly, our approach cannot distinguish between different types of fixed suboptimal computations. We modeled them by degrading each in the set of optimal receptive fields.

But an array of computations that make fixed suboptimal use of the available stimulus information could have similar effects.

Another potential issue is that eye movements were not controlled, raising the concern that human and ideal observers were not on equal footing. If eye movements are stimulus independent, they could manifest like internal noise, and decrease decision variable correlation(Rolfs, 2009; Kowler, 2011). On the other hand, if different eye movements are reliably elicited by different stimuli with the same speed (Turano & Heidenreich, 1999; Rucci & Poletti, 2015), they could manifest like suboptimal computations, and increase decision variable correlation. However, we believe that the steps we took to minimize the possible impact of uncontrolled eye movements are likely to have been largely successful. First, stimuli were presented for 250ms, the approximate duration of a typical fixation, and our stimuli were above half-max contrast for only ~200ms. Under stimulus conditions (i.e. speeds and contrasts) similar to ours, smooth pursuit eye movements have a latency of 140-200ms(Spering et al., 2005). Thus, if large eye movements occurred, it is likely that they would have occurred only in the last fraction of the trial. Second, numerous reports indicate that, when estimating motion, humans and other primates tend to weight stimulus information more heavily at the beginning than at the end of trial(Yates et al., 2017). Thus, the portion of the trial in which the eyes are most likely to have been stable is the portion that is most likely to have contributed to the speed estimate. Finally, fixational eye movements (i.e. drift, microsaccades, tremor) are likely to have contributed to our estimate of early measurement noise, and thus would have equivalently impacted both human and ideal performance. Still, given that eye movements can impact speed percepts under certain conditions(Turano & Heidenreich, 1999; Freeman et al., 2010; Goettker et al., 2018), this issue should be examined rigorously in future experiments.

There are many other possible directions for future work. First, there is a well established tradition of examining how changing overall contrast impacts speed sensitive neurons and speed perception(Thompson, 1982; Schrater et al., 2000; Weiss et al., 2002; Priebe et al., 2003; Priebe & Lisberger, 2004; Jogan & Stocker, 2015; Gekas et al., 2017). All stimuli in the current experiment were fixed to the most common contrast in the natural image movie set. As overall contrast is reduced speed sensitive neurons respond less vigorously, and moving stimuli are perceived to move more slowly(Thompson, 1982; Weiss et al., 2002; Priebe et al., 2003). It is widely believed that these effects occur because the visual system has internalized a prior for slow speed(Weiss et al., 2002). In the current manuscript, rather than covering well-trodden ground, we have focused on quantifying how image structure (i.e. the pattern of contrast) impacts speed estimation and discrimination. Thus, our results likely underestimate the impact of stimulus variability on ideal and human performance in natural viewing. The approach advanced in this manuscript can be generalized to examine how changes in overall contrast impact human and ideal performance. The role of stimulus variability has not been examined in this context, and may make an interesting topic for future work. Experiments should also be performed with full space-time (i.e. *xyt*) movies, with stimuli containing looming and discontinuous motion(Schrater et al., 2001; Nitzany & Victor, 2014). Finally, these same methods could be applied to a host of other tasks in vision and in other sensory modalities. New databases of natural images and natural sounds with corresponding groundtruth information about the distal scenes will significantly aid these efforts(Adams et al., 2016; Burge et al., 2016; Traer & McDermott, 2016).

Sources of performance limits

Efforts to determine the dominant factors that limit performance span research from sensation to cognition. The conclusions that researchers have reached are as diverse as the research areas in which the efforts have been undertaken. Stimulus noise(Hecht et al., 1942), physiological optics(Banks et al., 1987), internal noise(Burgess et al., 1981; Pelli, 1985; Williams, 1985; Pelli, 1991), suboptimal computations(Dosher & Lu, 1998; Beck et al., 2012; Drugowitsch et al., 2016), trial-sequential dependences(Laming, 1979), and various cognitive factors(Tversky & Kahneman, 1971) have all been implicated as the dominant factors that limit performance. What accounts for the diversity of these conclusions? We cannot provide a definitive answer. The relative importance of these factors is likely to depend on several things.

Evolution has pushed sensory-perceptual systems towards the optimal solutions for tasks that are critical for survival and reproduction. Humans are more likely to be assessed as optimal when visual systems are probed with stimuli that they evolved to process in tasks that they evolved to perform. In target detection tasks, for example, humans become progressively more efficient as stimuli become more natural(Banks et al., 1987; Abbey & Eckstein, 2014; Sebastian et al., 2017). Conversely, when stimuli and tasks bear little relation to those that drove the evolution of the system, the computations are less likely to be optimal. A new framework—a sciences of tasks—would be useful to help reconcile these disparate findings.

Image-computable ideal observers

Ideal observer analysis has a long history in vision science and systems neuroscience. In conjunction with behavioral experiments, image-computable ideal observers have shown that human light sensitivity is as sensitive as allowed by the laws of physics(Hecht et al., 1942), that the shape of the human contrast sensitivity function is dictated by the optics of the human eye(Banks et al., 1987), and that the pattern of human performance in a wide variety of basic psychophysical tasks can be predicted from first principles(Geisler, 1989).

To develop an image-computable ideal observer, it is critical to have a characterization of the task-relevant stimulus statistics. Obtaining such a characterization has been out of reach for all but the simplest tasks with the simplest stimuli. The vision and systems neuroscience communities have traditionally focused on understanding how simple forms of stimulus variability (e.g. Poisson or Gaussian white noise) impact performance(Hecht et al., 1942; Burgess et al., 1981; Pelli, 1985; Banks et al., 1987; Frechette et al., 2005). The impact of natural stimulus variability—the variation in light patterns associated with different natural scenes sharing the same latent variable values—has only recently begun to receive significant attention(Geisler & Perry, 2009; Burge & Geisler, 2011; Kane et al., 2011; Burge & Geisler, 2012; 2014; 2015; Sebastian et al., 2015; Schütt & Wichmann, 2017; Sebastian et al., 2017; Kim & Burge, 2018; Sinha et al., 2018).

Many impactful ideal observer models developed in recent years are not image-computable(Landy et al., 1995; Ernst & Banks, 2002; Weiss et al., 2002; Stocker & Simoncelli, 2006; Burge et al., 2010; Wei & Stocker, 2015). The weakness of these models is that they do not explicitly specify the stimulus encoding process, and therefore make assumptions about the information that stimuli provide about the task relevant variable (e.g. the likelihood function in the Bayesian framework). Consequently, these models cannot predict directly from stimuli how nuisance stimulus variability will impact behavioral variability, or explain how information is transformed as it proceeds through the hierarchy of visual processing stages. Image-computable models are thus necessary to achieve the goal of understanding how vision works with real-world stimuli. The current work represents an important step in that direction.

Impact on neuroscience

Behavioral and neural responses both vary from trial to trial even when the value of the latent (e.g. speed) is held constant. In many classic neurophysiological experiments, stimulus variability is eliminated by design, and experimental distinctions are not made between the latent variable of interest (e.g. orientation) and the stimulus (e.g. an oriented Gabor) used to probe neural response. Such experiments are well suited for quantifying how different internal factors impact neural variability. Indeed, it has recently been shown that, under these conditions, neural variability can be partitioned into two internal factors: a Poisson point-process and system-wide gain fluctuations(Goris et al., 2014). This approach provides an elegant account of a widely observed phenomenon (‘super-Poisson variability(Tomko & Crapper, 1974; Tolhurst et al., 1981; 1983)) that had previously resisted rigorous explanation. However, the designs of these classic experiments are unsuitable for estimating the impact of stimulus variability on neural response.

In the real world, behavioral variability is jointly driven by external and internal factors. Our results show that both factors place similar limits on performance. A full account of neural encoding and decoding must include a treatment of all significant sources of response variability. Partitioning the impact of realistic forms of stimulus variability from internal sources of neural variability will be an important next step for the field.

CHAPTER 3

ABSTRACT

PERCEPTUAL CONSEQUENCES OF INTEROCULAR IMBALANCES IN THE DURATION OF TEMPORAL INTEGRATION

Benjamin M. Chin

Johannes Burge

Temporal differences in visual information processing between the eyes can cause dramatic misperceptions of motion and depth. Processing delays between the eyes cause the Pulfrich effect: oscillating targets in the frontal plane are misperceived as moving along near-elliptical motion trajectories in depth (Pulfrich, 1922). Here, we explain a previously reported but poorly understood variant: the anomalous Pulfrich effect. When this variant is perceived, the illusory motion trajectory appears oriented left- or right-side back in depth, rather than aligned with the true direction of motion. Our data indicate that this perceived misalignment is due to interocular differences in neural temporal integration periods, as opposed to interocular differences in delay. For oscillating motion, differences in the duration of temporal integration dampen the effective motion amplitude in one eye relative to the other. In a dynamic analog of the Geometric effect in stereo-surface-orientation perception (Ogle, 1950), the different motion amplitudes cause the perceived misorientation of the motion trajectories. Forced-choice psychophysical experiments, conducted with either different spatial frequencies and/or different onscreen motion damping in the two eyes, show that the perceived misorientation in depth is associated with the eye having greater motion damping. A target-tracking experiment provided more direct evidence that the anomalous Pulfrich effect is caused by interocular differences in temporal integration and delay. These findings highlight the computational hurdles posed to the visual system by temporal

differences in sensory processing. Future work will explore how the visual system overcomes these challenges to achieve accurate perception.

3.1 Introduction

Temporal processing changes with the sensory stimuli being processed. Some sensory signals take longer to process than others. Stimulus-based differences in temporal processing delays—relative latencies—have received significant attention in vision science and neuroscience. Luminance signals are processed more quickly than chromatic signals. High luminance signals are processed more quickly than low luminance signals. High contrast signals are processed more quickly than low contrast signals. And low frequency stimuli are processed more quickly than high frequency stimuli. Despite differences in the speed by which these signals are processed, they are integrated by the brain. The computational rules that govern the integration of complementary signals with different temporal dynamics are not yet well understood. Identifying striking perceptual phenomena that result from combining such signals, and developing high-fidelity tools for measuring and characterizing these phenomena, should aid the discovery of computational principles underlying the combination rules.

Binocular integration of information between the eyes is crucial to depth perception. When a scene is viewed binocularly, the images are different in the two eyes because of their different vantage points on the scene. The spatial differences between the images in the two eyes underlie stereopsis, the perception of depth from binocular information. Estimation of these spatial differences can be impacted by differences in temporal processing between the eyes, especially when the images move.

Simple processing delays between the eyes cause oscillating targets in the frontal plane to be misperceived as moving along near-elliptical motion trajectories in depth (Fig. 3.1AB). Such interocular delays cause effective spatial displacements in one eye relative to the other—a neural disparity—that results in the illusory motion in depth. This illusion is known as the Pulfrich effect (Pulfrich, 1922). Luminance, contrast, and blur differences between the eyes are all known to the effect. Two types of the Pulfrich effect have been reported: the classic Pulfrich effect and the reverse Pulfrich effect. In the classic Pulfrich effect, the eye with lower luminance or contrast is processed more slowly (Lit, 1949; Reynaud & Hess, 2017; Wilson & Anstis, 1969; Fig. 3.1A). In the reverse Pulfrich effect, the eye with lower image quality (due to blur) is processed more quickly (Burge, Rodriguez-Lopez, & Dorronsoro, 2019; Rodriguez-Lopez, Dorronsoro, & Burge, 2020; Fig. 3.1B).

The reverse Pulfrich effect is mediated by blur-induced differences in the spatial frequency content between the stimuli in the two eyes (Burge et al., 2019; Rodriguez-Lopez et al., 2020). Blurring an image low-pass filters the image: high spatial frequencies are selectively removed. Because high spatial frequencies are processed more slowly than low spatial frequencies, the sharp image is processed more slowly than the blurry image. Complementarily, high-pass filtering increases the proportion of high frequencies in the image, and causes the high-pass filtered image to be processed more slowly (Burge et al., 2019). Similarly, if the two eyes are stimulated by moving Gabor stimuli with different carrier frequencies, signals from the eye with higher frequencies are processed more slowly (Min, Reynaud, & Hess, 2020). Thus, simple processing delays (i.e. time shifts in neural responses) nicely account for the standard Pulfrich effect: the perception of illusory 3D motion aligned with the true path of motion.

Anomalous Pulfrich percepts have also been reported (Emerson & Pesta, 1992; Harker & O'neal, 1967; Trincker, 1953; Weale, 1954). In such cases, observers report perceiving near-elliptical motion paths with principal axes that are rotated in depth relative to the true direction of motion (Fig. 3.1CD). Simple processing delays cannot explain these percepts. Various explanations have been proposed regarding the cause of anomalous Pulfrich percepts: saccadic suppression, velocity extrapolation, and perceptual distortion of objective visual space (Emerson & Pesta, 1992; Harker & O'neal, 1967; Trincker, 1953; Weale, 1954). But scientific consensus has not coalesced around any of these explanations. The aim of this paper is to explain this previously reported but poorly understood variant of the illusion.

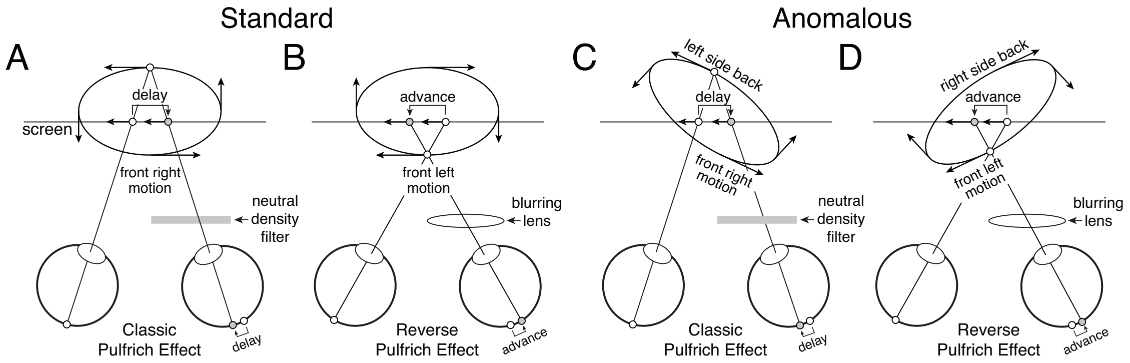

Figure 3.1. Standard and anomalous versions of the classic and reverse Pulfrich effects. **A** Standard version of the classic Pulfrich effect. A neutral density filter delays the signal in one eye relative to the other by decreasing luminance. **B** Standard version of the reverse Pulfrich effect. A blurring lens advances the signal in one eye relative to the other. **C** Anomalous version of the classic Pulfrich effect. **D** Anomalous version of the reverse Pulfrich effect.

We hypothesize that anomalous Pulfrich percepts—illusory motion trajectories that are rotated in depth with respect to the true motion trajectory—are caused by differences in the duration of time over which each eye integrates visual information; that is, different temporal integration periods. To understand this hypothesis, consider the temporal dynamics of sensory processing. The neural response to a sensory stimulus evolves over time. This temporal evolution can be described by an impulse response function. For a moving stimulus, the effective position over time of the neural image is impacted by this impulse response function. If the impulse response function in one eye is delayed relative to the impulse response function in the other eye (i.e. they are time-shifted copies of each other), stereo-geometry predicts the standard Pulfrich effect, when oscillatory motion is presented (Fig. 3.2AB). If, on the other hand, the impulse response function in one eye is both delayed and has a longer temporal integration period than the impulse response function in the other eye, then the amplitude of the effective motion signal will be damped in that eye relative to the other eye (Fig. 3.2CD). In this case, stereo-geometry predicts the anomalous Pulfrich effect: illusory motion-in-depth along a trajectory that is misaligned with the true direction of motion (see Fig. 3.1CD and Discussion).

Informally, we have most often observed anomalous Pulfrich percepts when there is different spatial frequency content in the two eyes. It is well-known that different spatial frequencies are processed both with different delays and with temporal integration periods of different durations. Neurons in early visual cortex (V1) and the middle-temporal area (MT) respond to higher spatial frequencies with more delay and longer temporal integration periods, all else equal (Bair & Movshon, 2004; Frazor, Albrecht, Geisler, & Crane, 2004; Vassilev, Mihaylova, & Bonnet, 2002). Psychophysical experiments have shown that human perceptual responses are similarly affected by spatial frequency (Levi, Harwerth, & Manny, 1979).

Here, with a traditional two-alternative forced choice (2AFC) paradigm previously used to study the Pulfrich effect, we first presented different spatial frequencies to the two eyes and asked observers to report the perceived orientation of the motion trajectory in depth (‘left side back’ vs ‘right-side back’; see Fig. 3.1CD). Anomalous (i.e. non-fronto-parallel) motion trajectories were reported in the expected direction. Next, to confirm that effective motion damping in the eye with the higher spatial frequency was indeed the proximal cause of anomalous Pulfrich percepts, we presented identical stimuli in the two eyes and independently damped the onscreen amplitudes of the left and right eye motion trajectories. Again, anomalous Pulfrich percepts occur as expected directions. Then, we conducted an experiment with multiple levels of onscreen damping and measured psychometric functions. This experiment allowed us to estimate the relative neural damping caused by interocular differences in spatial frequency.

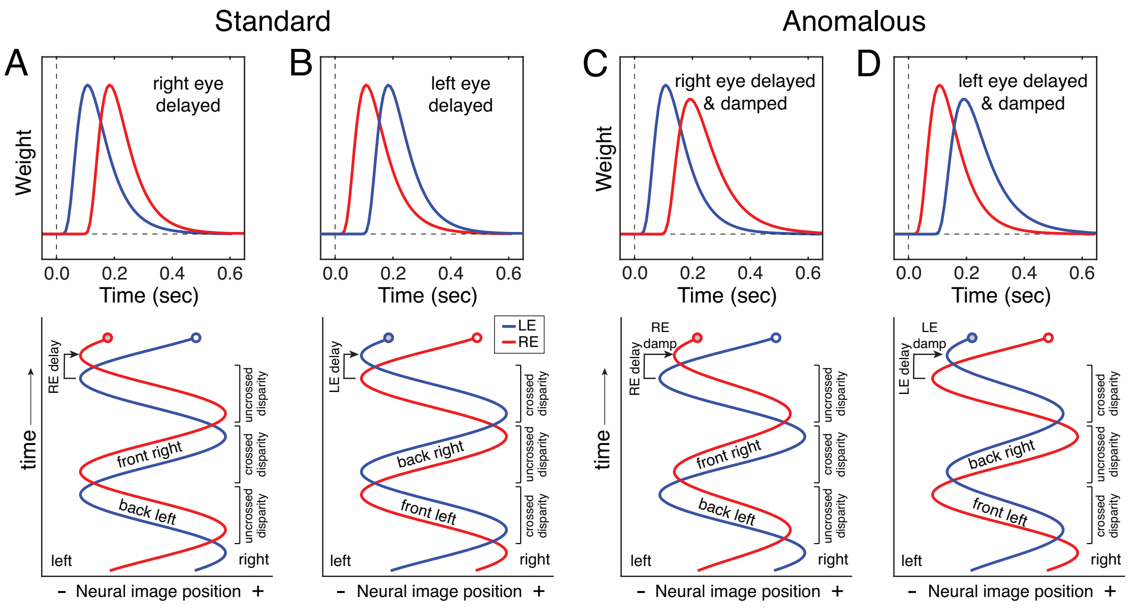

Figure 3.2. Predicting standard and anomalous Pulfrich percepts. Temporal impulse response functions (top) and effective neural image positions over time (bottom) for the left (blue) and right (red) eyes when **A** processing in the right eye is delayed relative to the left, **B** processing in the left eye is delayed relative to the right, **C** processing in the right eye is delayed and damped (due to a longer temporal integration period) relative to the left, and **D** processing in the left eye is delayed and damped relative to the right. Standard Pulfrich percepts result from delays only. Anomalous Pulfrich percepts result when the effective motion trajectory in one eye is both delayed and damped relative to the other (see Discussion).

Finally, with a recently developed target-tracking paradigm for continuous psychophysics that uses hand movement as the measure of response, we demonstrate that the visuomotor system processes high spatial frequencies with more delay and longer temporal integration periods than low spatial frequencies. These results dovetail with those from the traditional forced-choice experiments. Previous studies have shown that delays in sensory processing are faithfully preserved in the movement of the hand (Burge & Cormack, 2020). (Similar findings have been reported for the smooth pursuit eye movements of the oculomotor system; Lee, Joshua, Medina, & Lisberger, 2016). We conclude that differences in temporal integration between the eyes can cause anomalous Pulfrich percepts.

3.2 Results

We conducted four separate experiments to test the hypothesis that mismatched temporal integration periods can cause anomalous Pulfrich percepts. We used a within-subjects design. The first three experiments used a traditional forced choice paradigm. Observers binocularly viewed an oscillating Gabor stimulus and indicated the perceived orientation of its motion trajectory in depth. The fourth experiment was conducted using continuous target-tracking psychophysics (Bonnen, Burge, Yates, Pillow, & Cormack, 2015; Bonnen, Huk, & Cormack, 2017; Burge & Cormack, 2020; Knöll, Pillow, & Huk, 2018; Mulligan, Stevenson, & Cormack, 2013). Observers manually tracked a randomly moving Gabor stimulus with a cursor. The results of all experiments support the conclusion that different temporal integration periods can cause differential motion damping in the two eyes, and that this differential damping is the cause of anomalous Pulfrich percepts.

Experiment 1: Neural damping with mismatched spatial frequencies in the two eyes

Experiment 1 was designed to establish the dependence of the anomalous Pulfrich effect on spatial frequency. On each trial, the observer was dichoptically presented an oscillating Gabor stimulus. The onscreen disparities specified a near-elliptical trajectory in depth that was aligned with the screen.

A different carrier spatial frequency was presented to each eye. Under our working hypothesis, the Gabor with the higher spatial frequency should be processed with more delay and (crucially) with a longer temporal integration period than the lower frequency Gabor in the other eye. The longer temporal integration period should cause damping of the effective motion in that eye. The damping, in turn, should cause the illusory orientation of the motion trajectory in depth. We predict that observers will report more 'left-side-back' orientations when the left eye has the lower spatial frequency, and more 'right-side-back' orientations when the right eye has the lower spatial frequency (Fig. 3.3AB).

Observers were asked to report the apparent orientation of the motion trajectory in depth by indicating with a key press whether the principal axis of the trajectory appeared rotated left-side-back or right-side-back from the plane of the screen. Recall that the onscreen disparities specified that the motion trajectory was aligned with the screen. Absent eye-specific effects of spatial frequency, observers should not perceive left- and right-side-back orientations, such that each response is equally probable. If, on the other hand, if the two spatial frequencies are processed with different temporal integration periods, observers should report more 'right-side-back' orientations when the right eye has the lower spatial frequency, and vice versa.

Observers reported more ‘right-side back’ orientations of the perceived motion trajectory in depth when the right eye contained the lower frequency and more ‘left-side back’ orientations when the left eye contained the lower frequency (Fig. 3.3CD; also see Fig. 3.S1). In one observer, the effect appeared in all four conditions. For three observers, this effect was present in three out of the four conditions. This pattern of results is consistent with the experimental hypothesis. The null hypothesis is that spatial frequency has no effect on perceived orientation. Thus, if the null hypothesis was correct, observers should have reported right-side back orientations 50% of the time regardless of whether the left or right eye was presented the higher spatial frequency. We performed binomial tests on the group data to determine whether the null hypothesis could be rejected (see Methods). The tests rejected the null hypothesis for interocular spatial frequency combinations of 1cpd vs. 3cpd (p<0.01), 2cpd vs. 4cpd (p<0.001), and 3cpd vs. 6cpd (p<0.001), but not 1cpd vs. 2cpd (p=0.25). These results are largely consistent with the hypothesis that higher frequencies are processed by the visual system with longer temporal integration periods.

Conditions in which the left eye was stimulated with the lower spatial frequency (e.g. 1cpd to the left eye, 3cpd to the right eye; Fig. 3.3A) were interleaved with conditions in which the right eye was stimulated the lower spatial frequency (e.g. 3cpd to the left eye, 1cpd to the right eye; Fig. 3.3B). There are two benefits to this design. First, idiosyncratic block-specific response biases that may be present on a given block should be equally distributed amongst both conditions and have little effect on the final results. Second, because humans are poor at utrocular discrimination, it is difficult to determine which of the two eyes are being presented a given stimulus (Blake & Cormack, 1979; Schwarzkopf, Schindler, & Rees, 2010). Intermixing conditions ensures that, on any given trial, observers were unclear about which eye was being presented which stimulus. Hence, it would be quite difficult for observers to deliberately respond in a manner consistent with the experimental hypothesis.

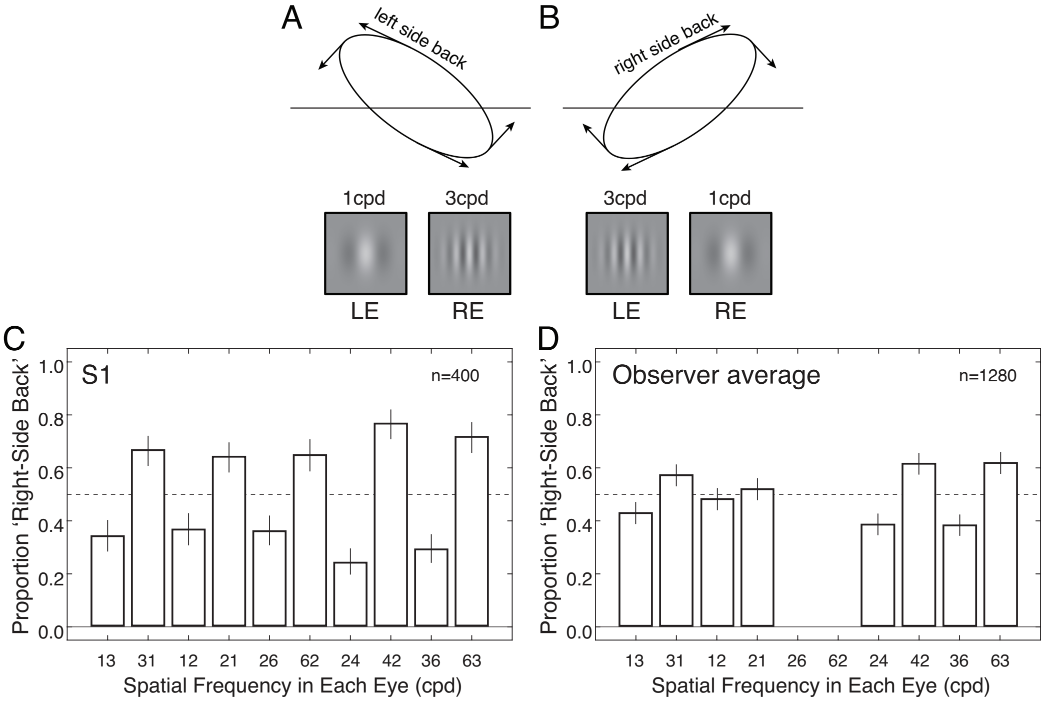

Figure 3.3. Experiment 1 stimuli, conditions, and results. **A** A low frequency Gabor in the left eye and a high frequency Gabor in the right eye predicts that the target will be perceived as moving along a trajectory oriented left-side-back in depth. **B** A high frequency Gabor in the left eye and a low frequency Gabor in the right eye predicts that the target will be perceived as moving along a trajectory oriented right-side-back in depth. Although fusion was imperfect, all observers reported percepts of motion in depth. **C** Mean-centered results from one observer (see Methods). A different spatial frequency was presented to each eye. In all cases, 'right-side back' orientations were reported less often when the low frequency was in the left eye, and more often when the low frequency was in the right eye and. **D** Results combined across observers. Note that the 2cpd vs. 6cpd condition is absent from this plot. In the screening phase (see Methods), no observers other than observer S1 were able to fuse the stimuli in this condition, even when large onscreen disparities were present.

Presenting substantially different images to the two eyes in a task that is supported in part by stereopsis raises concerns about poor binocular fusion. We screened eight observers for their ability to fuse and perform a stereo-motion task with mismatched stimuli in the two eyes (see Methods). Four of the eight screened observers were able to perform the task. That is, four of eight observers reported seeing depth and were able to acceptably fuse the target when large binocular disparities were presented onscreen.

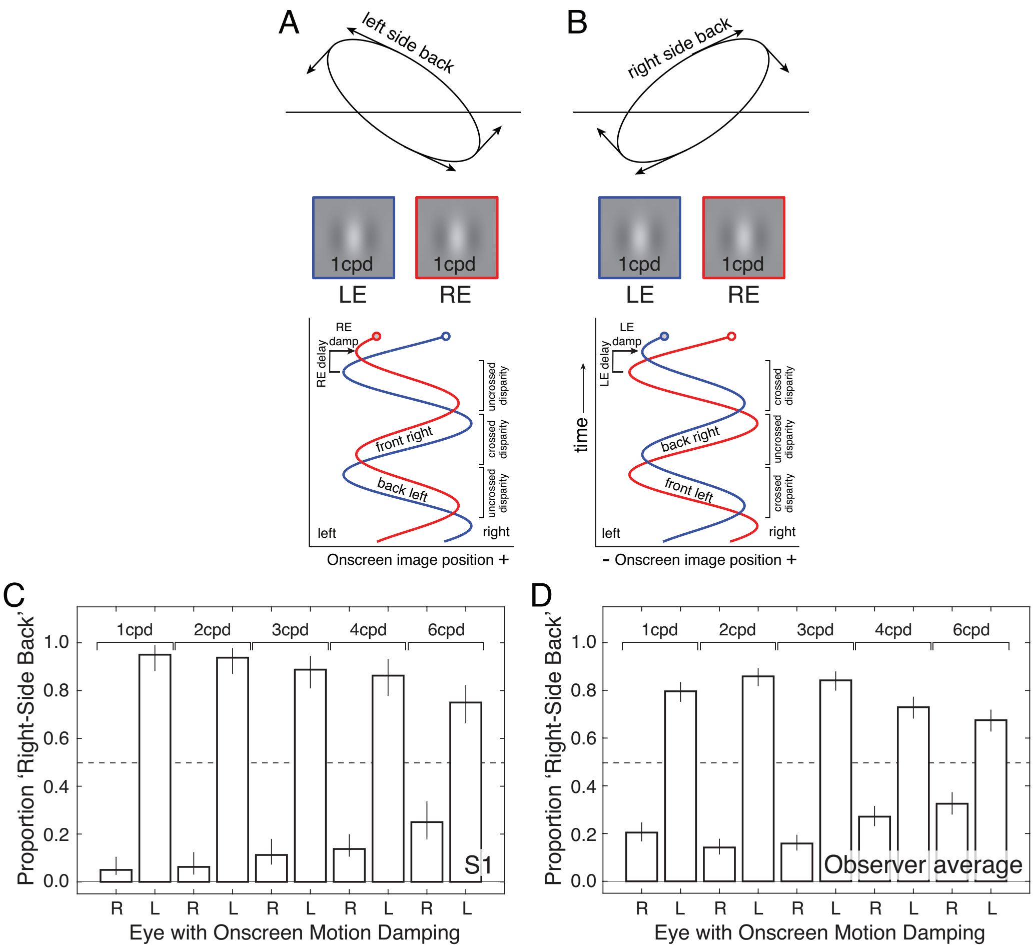

Figure 3.4. Experiment 2 stimuli, conditions, and results. **A** When the eyes are presented Gabors with the same spatial frequency, but the right-eye motion amplitude is damped (and delayed) onscreen, stereo-geometry specifies a near-elliptical motion trajectory that is oriented left-side back with respect to the screen. **B** When the eyes are presented Gabors with the same spatial frequency, but the left-eye motion amplitude is damped (and delayed) onscreen, stereo-geometry specifies a near-elliptical motion trajectory that is oriented right-side back with respect to the screen. **C** Mean-centered results from one observer. For all frequencies, responses are consistent with stereo-geometry-based predictions. Error bars indicate ±1 standard error. **D** Combined results across all observers.

Experiment 2: Onscreen damping with matched spatial frequencies in the two eyes

In Experiment 1, our hypothesis is that the effective motion amplitude in one eye was damped because a higher spatial frequency was presented to that eye. Assuming the hypothesis is correct, Experiment 1 therefore manipulated motion amplitude indirectly. Experiment 2 was designed to provide direct evidence that damping the motion signal in one eye changes the perceived orientation of the motion trajectory in depth with respect to the screen. Experiment 2 is distinguished from Experiment 1 by two major design changes. First, the motion amplitude was damped onscreen rather than manipulated indirectly via interocular spatial frequency differences (Fig. 3.4AB). The resulting onscreen disparities specified a motion trajectory in depth with a principal axis that was misaligned with the screen. Unlike in Experiment 1, identical Gabors—with the same spatial frequencies used in Experiment 1—were presented to the two eyes. Matched Gabors alleviate the fusion difficulties associated with mismatched Gabors. More importantly, because the Gabors were matched, neural processing delays and temporal integration periods should be matched between the eyes.

The task performed by each observer was the same as in Experiment 1. Observers reported both whether the perceived motion trajectory appeared to be oriented 'left-side back' or 'right-side back' with respect to the screen. When the right-eye onscreen motion amplitude was damped relative to the left-eye onscreen motion amplitude, observers more often reported trajectories that were oriented left-side back. When the left-eye onscreen motion amplitude was damped relative to the right-eye motion amplitude, observers more often reported trajectories that were oriented right-side back (Fig. 3.4CD; also see Fig. 3.S2). This data is similar to that collected in the first experiment. Recall that, under the working hypothesis, the mismatched Gabors in Experiment 1 should yield mismatched temporal integration periods between the eyes. The mismatched temporal integration periods, in turn, cause differential neural damping of the effective motion amplitude in the two eyes. The fact that the data exhibits similar patterns in the two experiments suggests that similar percepts result from differential onscreen damping, on one hand, and differential neural damping that results from mismatched Gabors in the two eyes, on the other.

Experiment 3: Estimating the magnitude of neural damping

Experiment 3 was designed to measure the amount of neural damping that is induced by mismatched frequencies in the two eyes. The logic of the design is as follows: If anomalous Pulfrich percepts are due to neural damping differences in addition to neural delays, it should be possible to find the onscreen damping differences that eliminate the perceived orientation of the motion-in-depth trajectory. These onscreen damping differences should be equal in magnitude but opposite in sign of the neural damping differences that are induced by mismatched spatial frequencies in the two eyes.

We collected psychometric functions with onscreen damping difference as the independent variable in each condition, and measured the proportion of times that observers reported motion trajectories that were oriented ‘right side back’ with respect to the screen. The resulting psychometric functions are shown in Fig. 3.5A (also see Fig. 3.S3). The point of subjective equality (PSE) in each condition indicates the onscreen damping that is equal in magnitude and opposite in sign to the corresponding neural damping. When the right eye had the higher spatial-frequency stimulus, the left-eye onscreen motion had to be damped to eliminate anomalous Pulfrich percepts, and vice versa (Fig. 3.5BC; also see Fig. 3.S4). This finding held across all tested frequency combinations. These results further support the hypothesis that stimulus-induced differences in temporal integration periods cause neural motion damping that can be neutralized by onscreen motion damping.

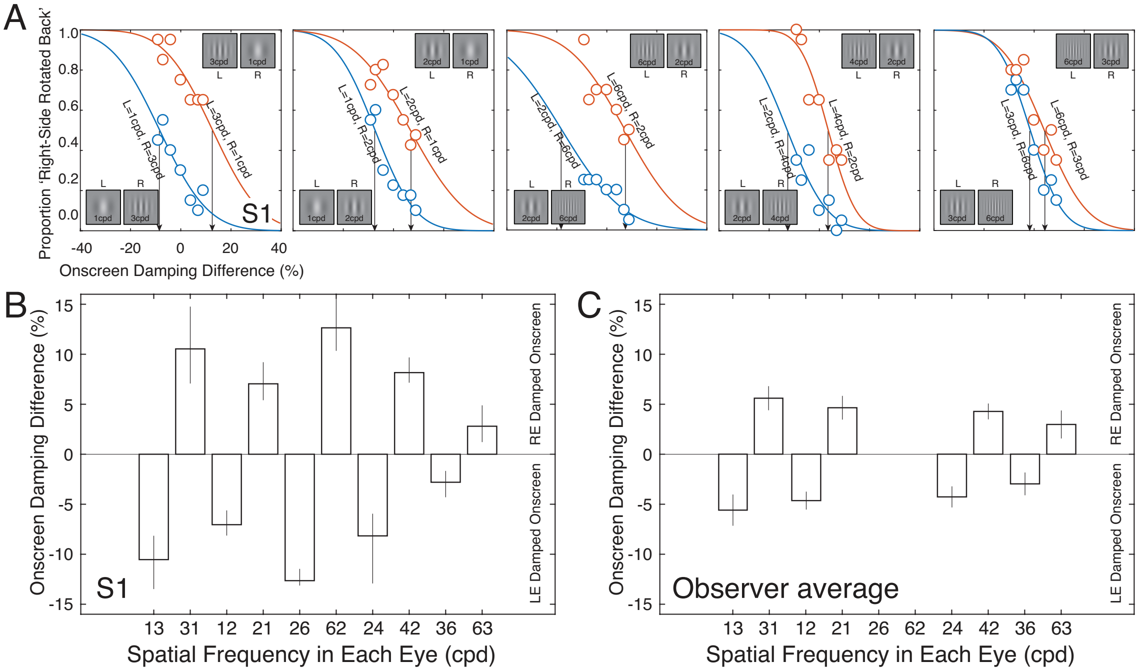

Figure 3.5. Experiment 3 stimuli, conditions, and results. **A** Psychometric functions from the first human observer for five different frequency pairs (i.e. 1cpd vs 3cpd, 1cpd vs 2cpd, 2cpd vs 6cpd, 2cpd vs 4cpd, and 3cpd vs 6cpd). When the left eye had the lower frequency stimulus, the psychometric functions were shifted consistently to the left (blue points and curves), indicating that the left-eye motion amplitude had to be damped onscreen to null the perceived orientation in depth. When the right eye had the lower frequency stimulus (red points and curves), the psychometric functions were shifted consistently to the right, indicating that the right-eye motion amplitude had to be damped onscreen to null the perceived orientations in depth. **B** Mean-centered points of subjective equality (PSEs; arrows in A) in each condition for one observer (see Methods). The PSEs are estimates of the amount of onscreen damping required to null the perceived orientations associated with different spatial frequencies in the two eyes. **C** Mean-centered PSEs averaged across all observers. Error bars represent bootstrapped ±1 standard errors on the points of subjective equality. Note that the 2cpd vs. 6cpd conditions are absent from this subplot. In the screening phase (see Methods), no observers other than observer S1 were able to perform the task in these conditions.

Experiment 4: Continuous target-tracking psychophysics

The results from the first three experiments i) establish that mismatched spatial frequencies in the two eyes cause anomalous Pulfrich percepts, ii) demonstrate that damping the onscreen motion amplitude in one eye causes anomalous Pulfrich percepts with matched spatial frequencies in the two eyes, and iii) show that onscreen damping can eliminate spatial-frequency-induced anomalous Pulfrich percepts. Together, these results suggest that different temporal integration periods between the eyes are the root cause of spatial-frequency-induced anomalous Pulfrich percepts. The evidence for this conclusion, however, is indirect. To gain more direct evidence that mismatched frequencies induce interocular differences in temporal integration, Experiment 4 made use of an entirely different paradigm: continuous target-tracking psychophysics (Bonnen et al., 2015).

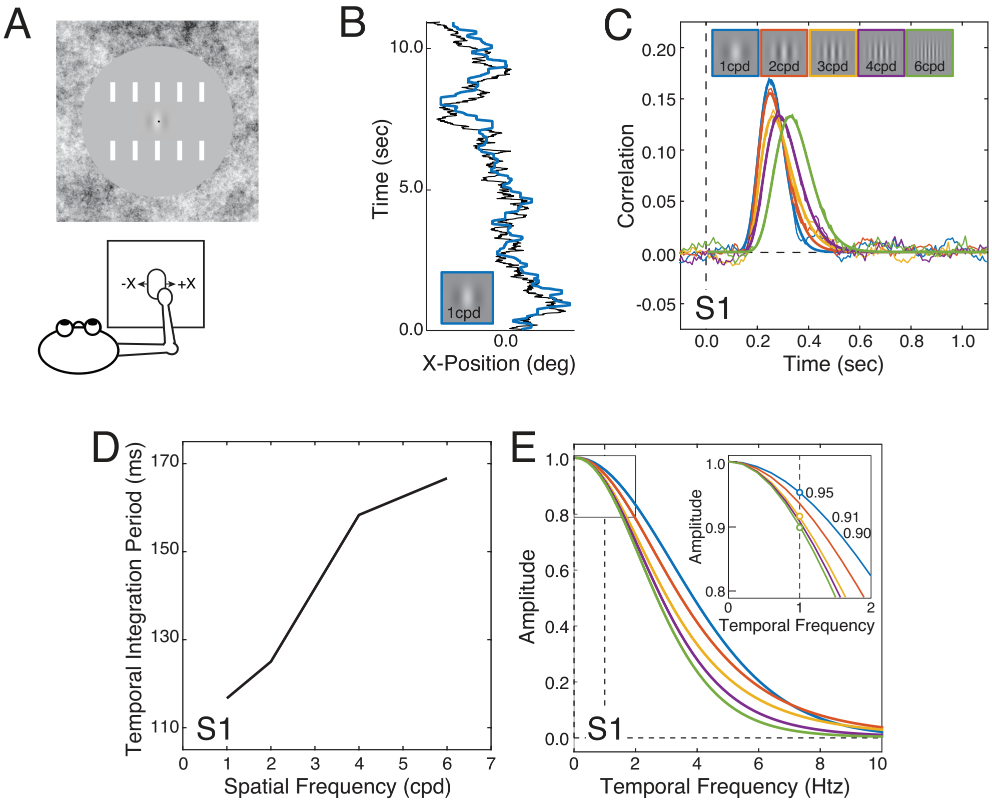

Figure 3.6. Effects of spatial frequency on target tracking performance for the first human observer. **A** On each trial, the observer tracked, with a mouse cursor (black dot at the center of the screen), a Gabor stimulus following a horizontal random walk across the center of the screen. **B** Example target-tracking performance on a single trial. The solid black trace indicates the horizontal random walk taken by the stimulus (1cpd Gabor). The blue trace indicates the position of the observer's cursor. **C** Cross-correlograms in the target tracking task derived from target-tracking performance. The cross-correlograms change systematically as a function of spatial frequency (colors). **D** The temporal integration period (i.e. full-width at half-height) increases from approximately 115ms to 165ms as spatial frequency increases from 1cpd to 6cpd. **E** The amplitude spectra of the cross-correlograms provide an estimate of the amount of effective motion damping for each of many temporal frequencies. The inset shows the estimated amount of visuomotor motion damping for each spatial frequency at 1.0htz, the temporal frequency of the motion stimulus in the 2AFC experiments.

Using a mouse, observers manually tracked one of five Gabor targets at a time (Fig. 3.6A). The Gabor targets had carrier spatial frequencies of 1cpd, 2cpd, 3cpd, 4cpd, and 6cpd. These spatial frequencies were matched to those used in the previous experiments. For example, spatial frequencies of 1cpd and 3cpd were used in the target-tracking task because conditions in the previous experiments involved presenting a 1cpd Gabor to one eye and a 3cpd Gabor to the other. Throughout each run, the target underwent a horizontal random walk on the screen (Fig. 3.6B). The task was performed without difficulty. The cross-correlation between the target and response motions provides information about the temporal processing of the visuo-motor system. If the visuomotor system is linear, the cross-correlogram equals an estimate of the temporal impulse response function when the target velocities are white noise, which they are here by design.

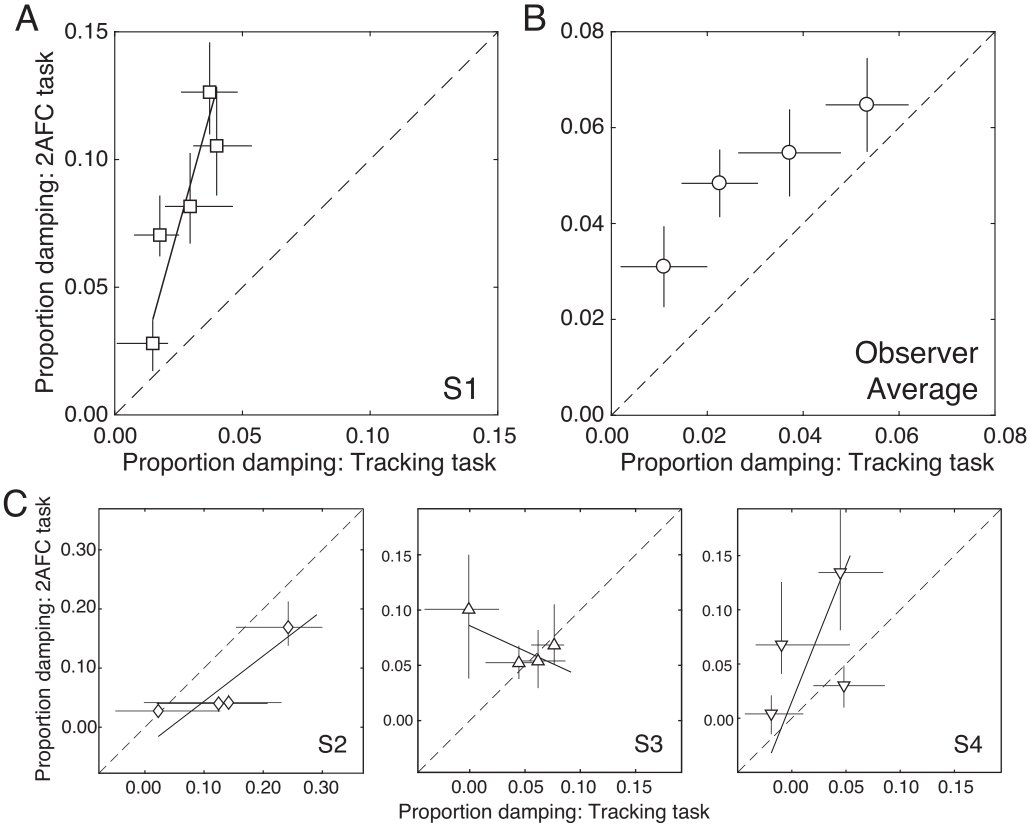

Figure 3.7. Comparison of 2AFC-based vs. target-tracking-based estimates of motion damping. **A** Results for the first observer. The best-fit line via weighted linear regression (solid), and the unity line (dashed) are also shown. Error bars on data points indicate 68% bootstrapped confidence intervals. **B** Average results across all observers (see Methods). **C** Results for each of the other three observers.

The cross-correlograms are broader in time as spatial frequency increases (Fig. 3.6CD). The amplitude spectra of the cross-correlograms indicate the proportion by which each spatial frequency is damped as a function of temporal frequency (Fig. 3.6E, see Methods). The inset shows at 1.0htz—the temporal frequency at which targets oscillated in the 2AFC forced-choice experiments (i.e. Exp. 1-3)—the motion amplitude of the visuomotor response in the tracking task decreases as spatial frequency increases. (The same is true for other temporal frequencies.) In other words, the amplitude of the visuomotor response is damped increasingly more as target spatial frequency increases.

To examine whether the visuomotor damping estimated in the target tracking task (Exp. 4, see Fig. 3.6E) can predict the effective sensory-perceptual motion damping estimated in the 2AFC forced-choice task (Exp. 3, see Fig. 3.5), we plotted the estimates of damping from the two experiments against each other. For the first human observer, the damping estimates are strongly correlated ($r=0.90$; $p<0.04$; Fig. 3.7A). The group average shows a similar trend ($r=0.98$; $p=0.02$; Fig. 3.7B). For all observers but one, the same qualitative pattern exists: sensory-perceptual- and tracking-based estimates of motion damping increase together. However, the slopes of the best-fitting lines vary substantially across observers (Fig. 3.7C). It will therefore be difficult, on an observer-by-observer basis, to predict the magnitude of the visuomotor motion damping in the tracking task from estimates of visual motion damping in the forced-choice task, or vice versa (see Discussion). Nevertheless, these results are consistent with the hypothesis that effective motion damping underlies anomalous Pulfrich percepts.

3.3 Methods

Participants

Four human observers participated in the experiment. Three observers were male and one observer was female. One was an author and the rest were naïve to the purposes of the experiment. All had normal or corrected to normal visual acuity (20/20), and normal stereoacuity as determined by the Titmus Stereo Test. The observers were aged 23, 26, 27, and 42 years old at the time of the measurements. All observers provided informed consent in accordance with the Declaration of Helsinki using a protocol approved by the Institutional Review Board at the University of Pennsylvania.

Apparatus

Stimuli were presented on a custom four-mirror stereoscope. Left- and right-eye images were presented on two identical Vpixx VIEWPixx LED monitors. The monitors were 52.2x29.1cm, with a spatial resolution of 1920x1080 pixels, a refresh rate of 120Hz, and a maximum luminance of 105.9cd/m^2. After light loss due to mirror reflections, the maximum luminance was 93.9cd/m^2. The gamma function of each monitor was linearized using custom software routines. A single AMD FirePro D500 graphics card with 3GB GDDR5 VRAM controlled both monitors to ensure that the left and right eye images were presented simultaneously. To overcome bandwidth limitations of the monitor cables, custom firmware was written so that a single color channel drove each monitor; the red channel drove the left monitor and the green channel drove the right monitor. The single-channel drive to each monitor was then split to all three channels for gray scale presentation.

Observers viewed the monitors through a pair of mirror cubes positioned one inter-ocular distance apart. The mirror cubes had 2.5cm openings. Given the eye positions relative to the openings, the field of view through the mirror cubes was ~15x15º. The outer mirrors were adjusted such that the vergence distance matched the 100cm distance of the monitors. This distance was confirmed both by a laser ruler measurement and by a visual comparison with a real target at 100cm. At this distance, each pixel subtended 1.09arcmin. Stimulus presentation was controlled via the Psychophysics Toolbox-3 (Brainard, 1997). Anti-aliasing enabled sub-pixel resolution permitting accurate presentations of disparities as small as 15-20arcsec. Heads were stabilized with a chin and forehead rest.

Forced-choice psychophysics target motion

For the forced-choice psychophysics experiments, we simulated the classic pendulum Pulfrich stimulus on the display. For each trial, the left- and right-eye on-screen bar positions in degrees of visual angle were given by

$$x_L(t) = E_L \cos(2\pi\omega \cdot (t + \Delta t) + \phi_0) \tag{1a}$$

$$x_R(t) = E_R \cos(2\pi\omega \cdot (t) + \phi_0) \tag{1b}$$

where $\mathrm{E_L}$ and $\mathrm{E_R}$ are the left- and right-eye motion amplitudes in degrees of visual angle, Δt is the on-screen delay between the left- and right-eye target images, ω is the temporal frequency of the target movement, ϕ_0 is the starting phase, and t is time in seconds.

The undamped motion amplitude was 1.5° of visual angle (3.0° total change in visual angle in each direction). The maximum onscreen motion damping in one eye (20%) corresponded to 80% (1.2° of visual angle) of the undamped amplitude in the other. The range of particular damping values was adjusted to the sensitivity of each observer. The on-screen interocular delays were set at ±25ms. The temporal frequency was 1 cycle per second. The starting phase ϕ_0 was randomly chosen on each trial to equal either 0 or π, which forced the stimuli to start either to the left or to the right of center.

When the onscreen interocular difference in motion amplitude equals zero and the onscreen interocular delay is zero, the target moves in the frontoparallel plane at the distance of the screen; the onscreen disparities are zero throughout the trial. If the interocular difference in motion amplitude is non-zero and/or if the interocular delay is non-zero spatial binocular disparities result, and the disparity-specified target follows a motion-in-depth trajectory outside the plane of the monitor. Differences in motion amplitude cause a disparity-specified misalignment in depth of the motion trajectory. Non-zero delays cause a disparity-specified elliptical trajectory of motion in depth. Negative delay values indicate the left-eye on-screen image is delayed relative to the right; positive delay values indicate the left eye on-screen image is advanced relative to the right.

The on-screen binocular disparity for a given interocular delay and damping as a function of time is given by

$$\Delta \mathrm{x(t)} = \mathrm{x_R(t)} - \mathrm{x_L(t)} = \sqrt{\mathrm{E_L^2} + \mathrm{E_R^2} - 2\mathrm{E_L E_R} \cos(2\pi\omega \cdot \Delta \mathrm{t})} \cos(\phi_0) \cdot \cos\left[2\pi\omega \mathrm{t} - \tan^{-1}\left[\frac{\mathrm{E_L} \sin\,(2\pi\omega\cdot\Delta \mathrm{t})}{\mathrm{E_R} - \mathrm{E_L} \sin\,(2\pi\omega\cdot\Delta \mathrm{t})}\right]\right] \qquad (2)$$

where negative disparities are crossed (i.e. nearer than the screen) and positive disparities are uncrossed (i.e. farther than the screen). The disparity takes on its maximum magnitude when the perceived stimulus is directly in front of the observer and the lateral movement is at its maximum speed. The maximum disparity in visual angle is given by $\Delta \mathrm{x}_{\max} = \sqrt{\mathrm{E}_\mathrm{L}^2 + \mathrm{E}_\mathrm{R}^2 - 2\mathrm{E}_\mathrm{L}\mathrm{E}_\mathrm{R}\cos{(2\pi\omega \cdot \Delta\mathrm{t})}}$and it occurs when $\mathrm{t} = \tan^{-1}\left[\frac{\mathrm{E}_\mathrm{L}\sin{(2\pi\omega\cdot\Delta\mathrm{t})}}{\mathrm{E}_\mathrm{R}-\mathrm{E}_\mathrm{L}\sin{(2\pi\omega\cdot\Delta\mathrm{t})}}\right]/2\pi\omega$. Note that we did not temporally manipulate when left- and right-eye images were presented on-screen; both eyes' images were presented coincidently on each monitor refresh. Rather, we calculated the disparity $\Delta x = \dot{x}\Delta t$ given the target velocity and the desired on-screen delay on each time step, and appropriately shifted the spatial positions of the left- and right-eye images.

Two sets of five vertically-oriented picket-fence bars (0.25x1.00º) flanked the region of the screen traversed by the target stimulus. The picket fences were specified by disparity to be at the screen distance. A 1/f noise texture, also defined by disparity to be at the screen distance, covered the periphery of the display. Both the picket fences and the 1/f noise texture served as stereoscopic references to the screen distance and helped to anchor vergence.

Before the target appeared on each trial, a small dot appeared 1.5º to the left of center or 1.5º right of center at the location of imminent target appearance. Observers were instructed to fixate the dot and then, after the target appeared, fixate and follow the target throughout the trial. In pilot experiments, we found that if observers did not follow the target with their eyes, the highest spatial frequency Gabor occasionally appeared to vanish during the trial at and near when it hit top speed (i.e. 6cpd). Observers reported whether the perceived motion trajectory was oriented left-side back or right-side back from frontoparallel. All experiments used a one-interval, two-alternative forced choice procedure.

Forced-choice psychophysics stimuli

The same Gabor targets were presented in the forced-choice psychophysical experiments as in the tracking experiments. A vertically-oriented Gabor is given by the product of a sinewave carrier and a Gaussian envelope

$$G(x,y) = \mathrm{gauss}(x,y;\sigma_x,\sigma_y)\cos(2\pi f x + \phi) \qquad (3)$$

where σ_x and σ_y are the standard deviation in X and Y of the Gaussian envelope, f is the frequency of the carrier, and ϕ is the phase. Five Gabor targets with different carrier frequencies were used: 1cpd, 2cpd, 3cpd, 4cpd, and 6cpd. All had the spatial size because all had same Gaussian envelope (σ_x = 0.39 and σ_y = 0.32). The octave bandwidths thus equaled 1.5, 0.7, 0.46, 0.35, 0.23 and the orientation bandwidths equaled 60º, 32º, 22º, 16º, and 11º, respectively. The phase of the carrier frequency was equal to 0.0 for all Gabor stimuli (i.e. all Gabors were in cosine phase).

Experiment 1 presented Gabors with different spatial frequencies in the two eyes. Data was collected in blocks with an intermixed design. For example, blocks containing conditions in which the left and right eyes were respectively presented 1cpd and 3cpd Gabors were intermixed with conditions in which the left and right eyes were presented 3cpd and 1cpd Gabors. In each condition, we used two values of interocular delay. The increased neural temporal integration period associated with high spatial frequencies served to dampen the effective motion amplitude in one eye relative to the other. Human observers have poor utrocular discrimination; humans have significant difficulty determining which eye is being presented a given stimulus (Blake & Cormack, 1979; Schwarzkopf et al., 2010). Intermixing conditions ensured that, on any given trial, observers—even non-naïve observers—were unclear about which eye was being presented which stimulus. Thus, observers—even non-naïve observers—would be unable to determine, on a given trial, which response was consistent with the experimental hypothesis.

Experiment 2 presented Gabors with the same spatial frequency in the two eyes, and used two interocular delays and two damping values. We chose damping values that made the orientation of the near-elliptical trajectory in depth ('left side back' vs. 'right side back') easy for the observers to identify.

Experiment 3 was designed to measure observer sensitivity to interocular differences in motion amplitude (i.e. damping). Experiment 3 thus measured full psychometric functions in each condition using the method of constant stimuli. Seven different levels of damping were collected for each function. The psychometric functions were fit with a cumulative Gaussian via maximum likelihood methods. The 50% point on the psychometric function—the point of subjective equality (PSE)—indicates the onscreen motion damping needed to null the relative motion damping due to spatial frequency differences. Observers ran 140 trials per condition (i.e. 140 trials per psychometric function) in counter-balanced blocks of 70 trials each.

Mean-centering of effects

Data from Experiments 1-3 were mean-centered for pairs of matched conditions. Matched conditions were those involving the same spatial frequencies (e.g. 1cpd vs. 3cpd and 3cpd vs. 1cpd). The proportion of 'right-side back' responses or effective damping was mean-centered across matched conditions according to the following equation:

$$\Psi_L^* = \Psi_L - \left(\Psi_L + \Psi_R\right)/2 + \Psi_0 \quad (4a)$$

$$\Psi_R^* = \Psi_R - \left(\Psi_L + \Psi_R\right)/2 + \Psi_0 \quad (4b)$$

For Exp. 1, Ψ represents the proportion of 'right-side back' responses, where Ψ_L and Ψ_R respectively correspond to conditions where the left eye, or the right eye, were presented the lower spatial frequency. For Exp. 2, Ψ also represents the proportion of 'right-side back' responses, and Ψ_L and Ψ_R respectively correspond to conditions where the onscreen motion was damped in the left eye, or in the right eye. For Exp. 3, Ψ represents the psychophysical estimate of onscreen motion damping, $\hat{D}$, that is required to null the neural damping, and Ψ_L and Ψ_R correspond to the condition in which onscreen motion was damped in the left-eye, or in the right eye, respectively. In Experiments 1 and 2, Ψ_0 has a value of 0.5. In Experiment 3, Ψ_0 has a value of 0%.

Binomial test for significance

Under our working hypothesis, 'right-side back' responses should be reported more often when the effective motion-amplitude in the left eye is smaller than that in the right eye. Similarly, when the effective motion-amplitude in the right eye is smaller than that in the left eye, 'right-side back' responses should be reported less often. The null hypothesis predicts that there will be no difference in the proportion of 'right-side back' responses across two matched conditions (e.g. 1cpd vs. 3cpd and 3cpd vs. 1cpd). To determine whether the proportions of 'right-side back' responses differed significantly from those predicted by the null hypothesis, we used a binominal test. Under the null hypothesis, the probability, p, of the observed response proportions is given by

$$p = \sum_{i=0}^{n-k} \left[\binom{n}{i} \pi_0^i \left(1-\pi_0^i\right)^{n-i} \sum_{j=i+k}^{n} \binom{n}{j} \pi_0^j \left(1-\pi_0^j\right)^{n-j} \right] \tag{5}$$

where n is the number of trials in a given condition, π_o is the probability of the observer responding 'right-side back' in each of the two matched conditions under the null hypothesis (i.e., 0.5), and k is the difference in the number of 'right-side back' responses between two matched conditions.

Reliability-weighted averaging of estimated motion damping

PSEs estimates in Experiment 3 were averaged across observers using reliability-weighted averaging:

$$\hat{D} = \sum_{i=1}^{N} r_i \hat{D}_i / \sum_{i=1}^{N} r_i , \qquad r = 1/s^2 \qquad \text{(6a,b)}$$

where $\hat{D}$ is the estimate of motion damping for a given condition, averaged across all observers. N is the number of observers and s is the standard error of motion damping estimates (as determined by 68% bootstrapped confidence intervals). Reliability-weighted averaging takes into consideration differences in the reliability of damping estimates across observers. These differences in reliability arise because some observers are more sensitive to onscreen motion damping than others. It is well-known from signal detection theory that greater sensitivity in a task is associated with more reliable estimates of the point of subjective equality (here, estimates of motion damping).

Estimated relationship between forced-choice- and target-tracking-based motion damping estimates

The relationship between 2AFC-based and target-tracking-based estimates of motion damping was fit with a line via weighted linear regression. Since estimates of motion damping from both tasks have associated uncertainty, simple linear regression is not appropriate. This is because simple linear regression assumes that one of the variables is independent, and thus has no associated uncertainty. We fit the parameters of the best-fit line with maximum likelihood methods using numerical optimization. The cost function was

$$c = \sum_{i=1}^{N}\left[\left(\hat{D}_{tracking,i} - \bar{D}_{tracking,i}\right)^{2} / \sigma^{2}_{tracking,i}\right] + \sum_{i=1}^{N}\left[\left(\hat{D}_{2AFC,i} - a - b\bar{D}_{tracking,i}\right)^{2} / \sigma^{2}_{2AFC,i}\right] \quad (7)$$

where N is the total number of conditions for an observer, $\hat{D}$ is the experiment-derived estimate of motion damping for a given condition, $\bar{D}$ is a free parameter indicating the expected amount of motion damping for a given condition, σ is the standard error of the motion damping estimate for a given condition (as determined by 68% bootstrapped confidence intervals), a is the y-intercept of the best fit line, and b is the slope of the best fit line.

Observer screening

Before inclusion in the main experiments, observers were screened for their ability to perform the task when the spatial frequencies in the two eyes differed by a factor of three. During this screening phase, the onscreen motion amplitude differed in the two eyes by a large amount of up to 20%. These onscreen amplitude differences caused the stereo-specified motion trajectory to be misaligned with the screen. If an observer was unable to correctly report the direction of the stereo-specified misalignment at least 80% of the time, no further data was collected from that observer. Four out of eight screened observers were excluded from the study on this basis. The excluded observers all reported difficulty fusing and difficulty seeing any stereo-specified depth at all. The pilot data is consistent with these reports.

Target-tracking procedure

Tracking data was collected from each observer in blocks of individual runs. Each run was initiated with a mouse click, which caused the target and a small dark mouse cursor to appear in the center of the screen. After a stationary period of 500ms, the target began a one-dimensional horizontal random walk (i.e. Brownian motion) for eleven seconds. The task was to track the target as accurately as possible with a small dark mouse cursor. Blocks contained intermixed runs from each of the four conditions.

Target-tracking psychophysics: Onscreen stimuli

Data was collected in five conditions, each of which was distinguished by a different target Gabor stimulus. Each Gabor target had one of five different carrier frequencies: 1cpd, 2cpd, 3cpd, 4cpd, and 6cpd. All Gabor targets shared the same Gaussian envelope (σ_x=0.39º & σ_y=0.32º), and subtended approximately 2.0ºx2.0º of visual angle (i.e. five sigma). Hence, in the five conditions, the octave bandwidths equaled 1.5, 0.7, 0.46, 0.35, and 0.23 and the orientation bandwidths equaled 60º, 32º, 22º, 16º, and 11º, respectively. Data was collected in five intermixed blocks of twenty runs each for a total 20 runs per condition.

Target-tracking psychophysics: Target motion

For the tracking experiments, the target stimulus performed a random walk on a gray background subtending 10.0x7.5º of visual angle, and was surrounded by a static field of 1/f noise. The region of the screen traversed by the target was flanked by two horizontal sets of thirteen vertically-oriented picket fence bars (Fig. 3.6A).

The x-positions of the target on each time step $t+1$ were generated as follows

$$x\left(t+1\right)=x\left(t\right)+\varepsilon_x \;; \quad \varepsilon_x \sim N\left(0,Q\right) \tag{8}$$

where ε_x is a random sample of Gaussian noise and Q is the drift variance. The random sample determines the change in target position between the current and the next time step. The drift variance determines the expected magnitude of the position change on each time step, and hence the overall variance of the random walk. The variance of the walk positions across multiple walks $\sigma^2\left(t\right)=Qt$ is equal to the product of the drift variance and the number of elapsed time steps. The value of the drift variance in our task (0.8mm per time step) was chosen to be as large as possible such that each walk would traverse as much ground as possible while maintaining the expectation that less than one walk out of 500 (i.e. less than one per human observer throughout the experiment) would escape the horizontal extent of the gray background area (176x131mm) before the 11 second trial completed.

The effective on-screen positions of the images are obtained by convolving the on-screen target images with the temporal impulse response function

$$\tilde{x}(t) = x(t) * h(t) \qquad (9)$$

where $h(t)$ is a temporal impulse response function corresponding to a specific frequency. Convolving the target velocities with the impulse response function gives the velocities of the effective target images. Integrating these velocities across time gives the effective target positions.

To determine the impulse response function relating the target and response, we computed the zero-mean normalized cross-correlations between the target and response velocities

$$\rho\left(\tau;\dot{x},\dot{\tilde{x}}\right)=\frac{1}{\left\|\dot{x}\left(t\right)\right\|\left\|\dot{\tilde{x}}\left(t\right)\right\|}\left[\sum_{t=1}^{N}\left(\dot{x}\left(t\right)-\overline{\dot{x}}\right)\left(\dot{\tilde{x}}\left(t+\tau\right)-\overline{\dot{\tilde{x}}}\right)\right] \qquad (10)$$

where τ is the lag, $\dot{x}$ and $\dot{\tilde{x}}$ are the target and response velocities. Assuming a linear system, when the input time series (i.e. the target velocities) is white, as it is here by design, the cross-correlation with the response gives the impulse response function of the system.

To compute the normalized cross-correlations, we did not include the first second of each eleven second tracking run so that observers reached steady state tracking performance. The mean cross-correlation functions shown in the figures were obtained by first computing the normalized cross-correlation in each run (Eqn. 10), and then averaging these cross-correlograms across runs in each condition.

Gamma distribution fits to mean cross-correlograms

To summarize the mean cross-correlograms, we fit a Gamma distribution function using maximum likelihood methods. The form of the fitted function was given by

$$\rho(\tau) = A\left[1/\left(\Gamma(s)m^{s}\right)\right](\tau - d)^{s-1}\exp\left[-(\tau - d)/m\right] \qquad (11)$$

where A is the amplitude, and m, s, and d are the parameters determining the shape and scale of the fit. The mode (i.e. peak) of the function is given by ms. We use the mode as our measure of delay. The full-width at half-height can be used as a measure of the temporal integration period, and can be computed via numeric methods. The damping associated with a given fitted function is given by the value of the normalized amplitude spectrum at the temporal frequency of the stimulus, which in the current experiments is one cycle per second.

3.4 Discussion

In this manuscript, we presented evidence that anomalous Pulfrich percepts—illusory motion trajectories in depth misaligned with the true direction of motion—are caused by interocular differences in temporal integration periods in the two eyes. This specific perceptual effect, and the reasons it occurs, have more general implications.

The integration of multiple complementary streams of incoming information with different temporal dynamics is fundamental to the performance of biological systems. In most cases, sensory-perceptual systems successfully solve this temporal binding problem, and compute accurate estimates of environmental properties. In some cases, the visual system fails to compensate for temporally mismatched signals, and inaccurate estimates result. Such cases are instructive. They can help reveal fundamental properties about the temporal nature of sensory signals, and make plain the striking perceptual consequences of insufficient compensatory mechanisms.

In this discussion section, we contextualize the anomalous Pulfrich effect with reference to other areas of vision research, consider how visual and visuomotor measures of performance are related, and discuss potential future directions.

Analogy to the Geometric effect in surface orientation perception

Horizontal minification (or magnification) of the image in one eye causes the misperception of surface orientation. This phenomenon is known as the Geometric effect (Banks & Backus, 1998; Ogle, 1950). The Geometric effect occurs because the horizontal minification in one eye distorts the patterns of binocular disparity such that they specify a surface slant that is different from the actual surface slant. For example, when a frontoparallel surface is viewed with a horizontal minifier in front of the right eye, the surface is perceived to be slanted left-side back. If the left-eye image is minified, the same surface is perceived to be slanted right-side back (Fig. 3.8).

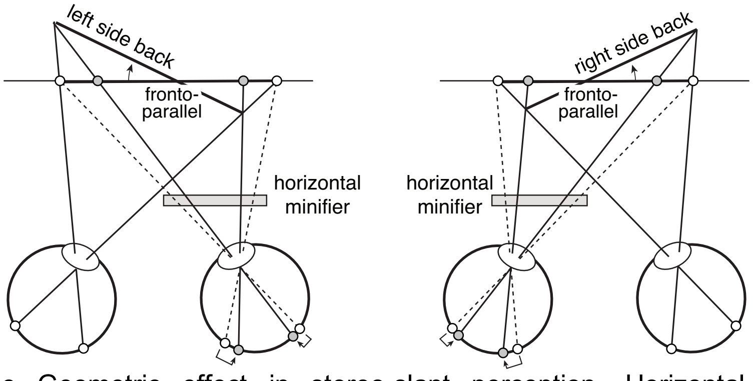

Figure 3.8. The Geometric effect in stereo-slant perception. Horizontal minification (or magnification) distorts the pattern of binocular disparities such that the disparity-specified orientation of the surface appears rotated in depth. If the horizontal minifier is in front of the right eye, a frontal surface straight-ahead is perceived left side back. If the horizontal minifier is in front of the left eye, a frontal surface straight-ahead is perceived right side back. The same principles account for both the Geometric effect and anomalous Pulfrich percepts.

The principles behind the Geometric effect mirror the principles behind the anomalous Pulfrich effect. An obvious analogy can be drawn between right- or left-eye motion damping and right- or left-eye horizontal minification. Anomalous Pulfrich percepts are caused by motion that is differentially damped between the two eyes. Indeed, if the effective image motion is damped but not delayed in one eye relative to the other, the disparity-specified motion trajectory lies in the plane of the slanted surface specified by disparities caused by the Geometric effect.

Preservation of sensory processing dynamics in motor movements

The current manuscript reports a series of results that strongly suggest that different spatial frequencies are processed with different temporal integration periods, and that these differences underlie anomalous Pulfrich percepts. Linking the target-tracking results to sensory-perceptual processing requires an assumption. The assumption is that changes in the ability of an observer to track a target across different target stimuli reflect changes in the sensory-perceptual processing of the stimuli as opposed to changes in the motor response. Multiple studies have shown this assumption holds in various situations. Motor variation in smooth-pursuit eye movements is due overwhelmingly to sensory errors (Osborne, Lisberger, & Bialek, 2005). Changes in the width of the cross-correlogram associating target and hand movements during target-tracking are linked to the sensitivity of visual target location discrimination (Bonnen et al., 2015). Delays in visual processing match delays in the motor response of both the eye (Lee et al., 2016), and the hand (Burge & Cormack, 2020; Lee et al., 2016). However, it appears from the present experiments that differences in the visual temporal integration period are not always faithfully preserved in the motor response of the hand.

Experiments 1-3 used traditional forced-choice psychophysical techniques to establish the anomalous Pulfrich phenomenon and quantify the effective motion damping that is caused by differences in temporal processing induced by different spatial frequencies. Experiment 4 used continuous target-tracking psychophysics to collect more direct evidence that different spatial frequencies are indeed associated with different temporal integration periods. The average estimates of motion damping across human observers from the target-tracking task very nearly matched those from the forced-choice task (see Fig. 3.7B). But there was significant inter-observer variability regarding how the two sets of estimates were related (see Fig. 3.7A,C). In two of four observers, the forced-choice-based estimates were systematically larger than the tracking-based estimates. In one observer, the reverse was true. And in the remaining observer, the estimates were nearly matched, except for an apparent outlier.

The finding that forced-choice- and target-tracking-based estimates of damping are correlated but do not exactly agree for individual observers warrants further study. Our analysis assumes that the motor component of the visuomotor response can be accurately modeled with convolution, a linear open-loop computation. It is likely that there are benefits to modeling visuomotor performance in the target-tracking task as a closed-loop system, given that visual feedback is integral to good performance in many visuomotor tasks. It is also possible that convolution does not accurately capture how the motor system translates visual input into a motor response. If so, other (possibly nonlinear) operations will be required to accurately model the motor contribution to performance. These, and related, issues are under active investigation.

Computational challenges of mismatched temporal processing

The visual system must constantly deal with the problem of staggered information arrival. We have focused on the perceptual consequences of temporal processing differences associated with mismatched spatial frequency content in the two eyes. Interocular differences in spatial frequency content commonly occur in natural viewing. During binocular viewing of surfaces that are slanted about a vertical axis, for example, the spatial frequencies tend to be higher in one eye than the other. These differences, while extremely common, tend to be relatively small. For a surface at a distance of 30cm and a slant of 72º, the corresponding frequencies in the two eyes will differ by approximately a factor of two (i.e. horizontal size ratios of 0.5 or 2.0, depending on whether the surface is slanted left- or right-side back). For more distant and less slanted surfaces, which are more common in natural viewing (Adams et al., 2016; Backus, Banks, van Ee, & Crowell, 1999; Burge, McCann, & Geisler, 2016; Kim & Burge, 2018; 2020; Yang & Purves, 2003), the ratio tends to be substantially smaller. However, typical natural images have broadband 1/f spectra, and frequencies above the contrast detection threshold typically vary by a factor of ten or more. Thus, the temporal binding problem may be a more acute computational challenge within each eye's image than between the images in the two eyes. In spite of this challenge, the visual system usually generates (largely) accurate estimates of environmental properties.

Measuring the temporal processing constraints of the nervous system, and developing normative theory for how different streams of information should be integrated to achieve accurate perceptual estimates, will help advance our understanding of how the spatial-frequency binding problem is resolved by biological systems (Burge et al., 2019). Incorporating these solutions into image-computable ideal observers for sensory-perceptual tasks with natural stimuli is a potentially fruitful future direction for neuroscience and vision research (Burge, 2020; Burge & Geisler, 2011; 2012; 2014; 2015; Chin & Burge, 2020).

Conclusion

The problem of binding temporally damped and temporally staggered information is not a niche problem. It is not at all specific to the combination of information from different spatial frequency channels, as we have focused on in this paper. The visual system must resolve temporal differences between luminance and chromatic signals, high and low luminance signals, and high and low contrast signals. More generally, the different senses—visual, auditory, vestibular, proprioceptive, tactile—transmit signals possessing substantially different temporal properties. These signals must also be combined to form accurate, temporally coherent percepts. Future work will investigate how sensory-perceptual systems solve the temporal binding problem within and across senses.

3.5 Supplement

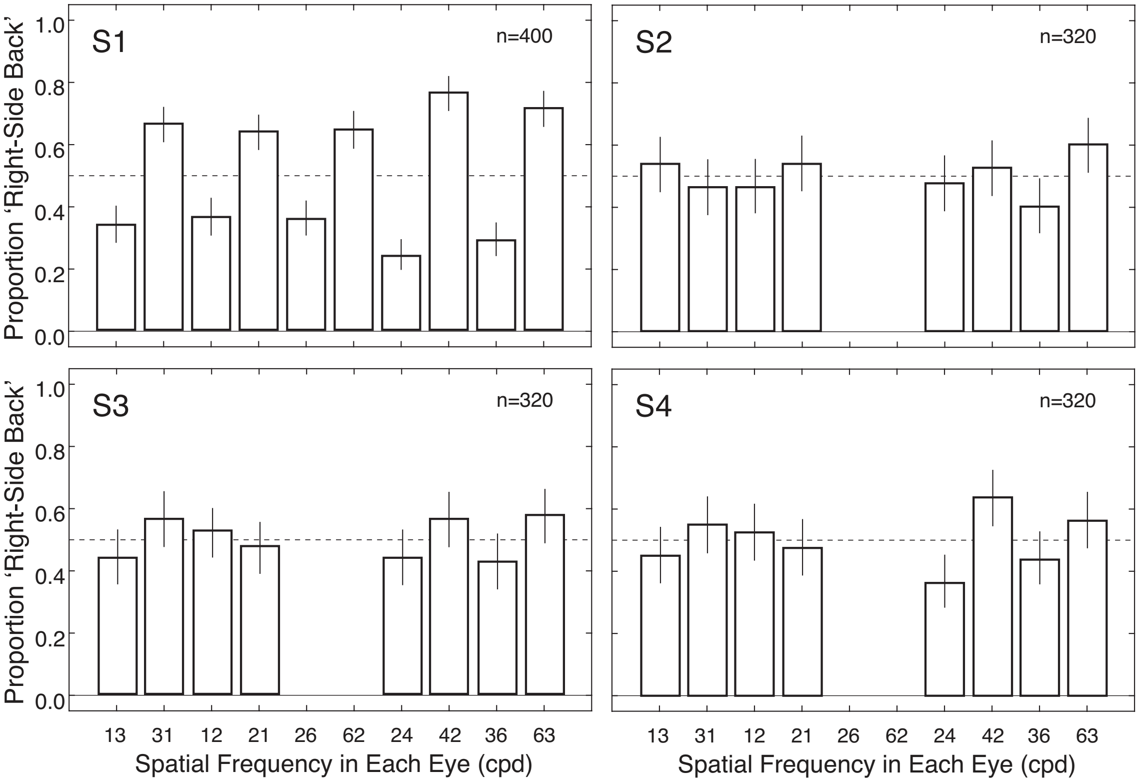

Figure 3.S1. Experiment 1 results for all observers and conditions.

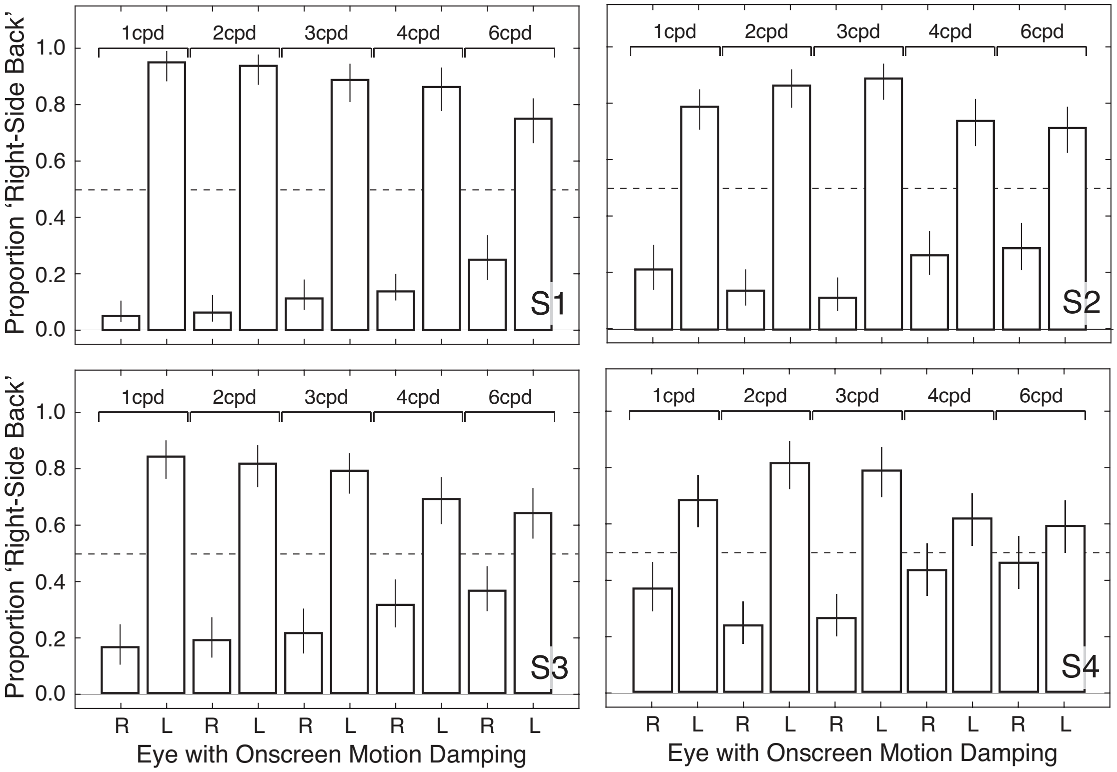

Figure 3.S2. Experiment 2 results for all observers and conditions.

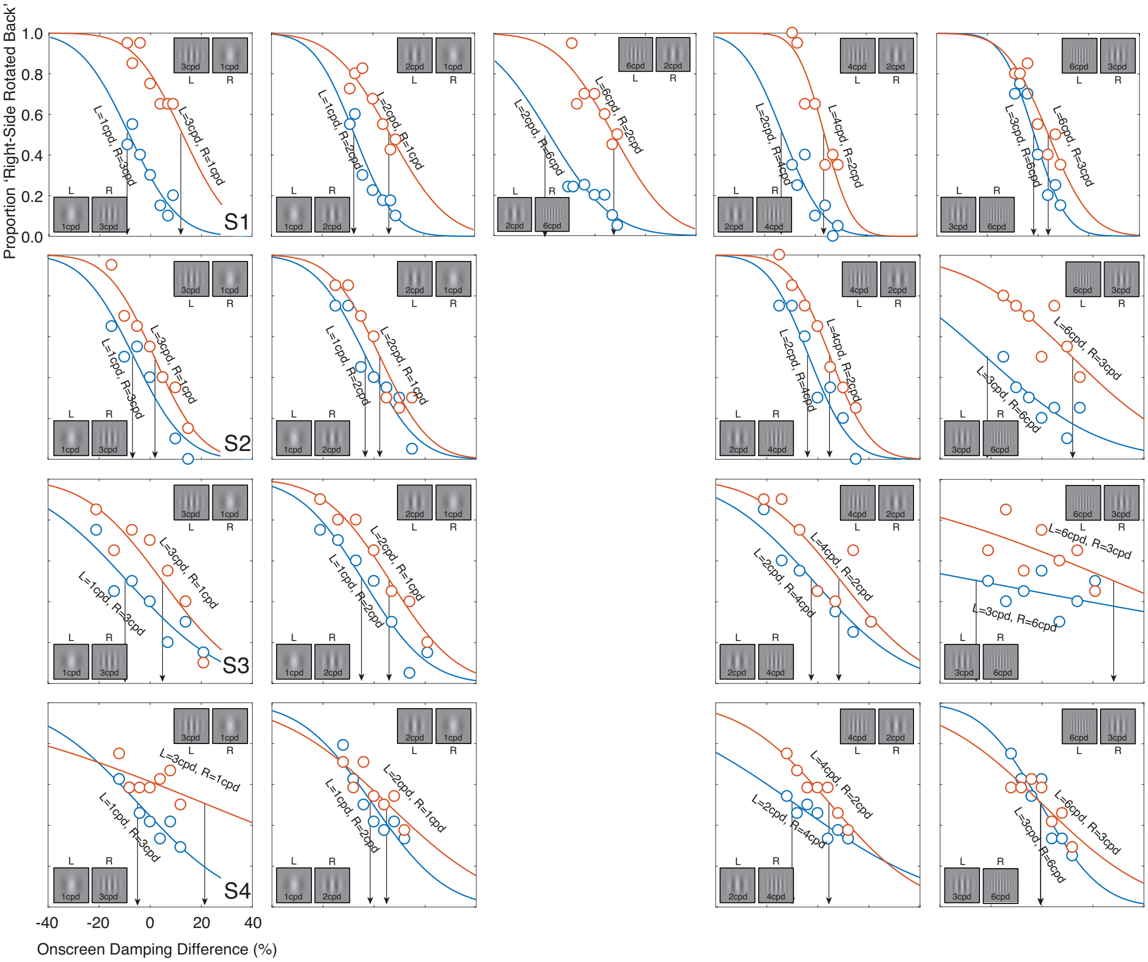

Figure 3.S3. Experiment 3 stimuli, conditions, and psychometric functions for all observers and conditions. The data from all four observers follow the same qualitative pattern.

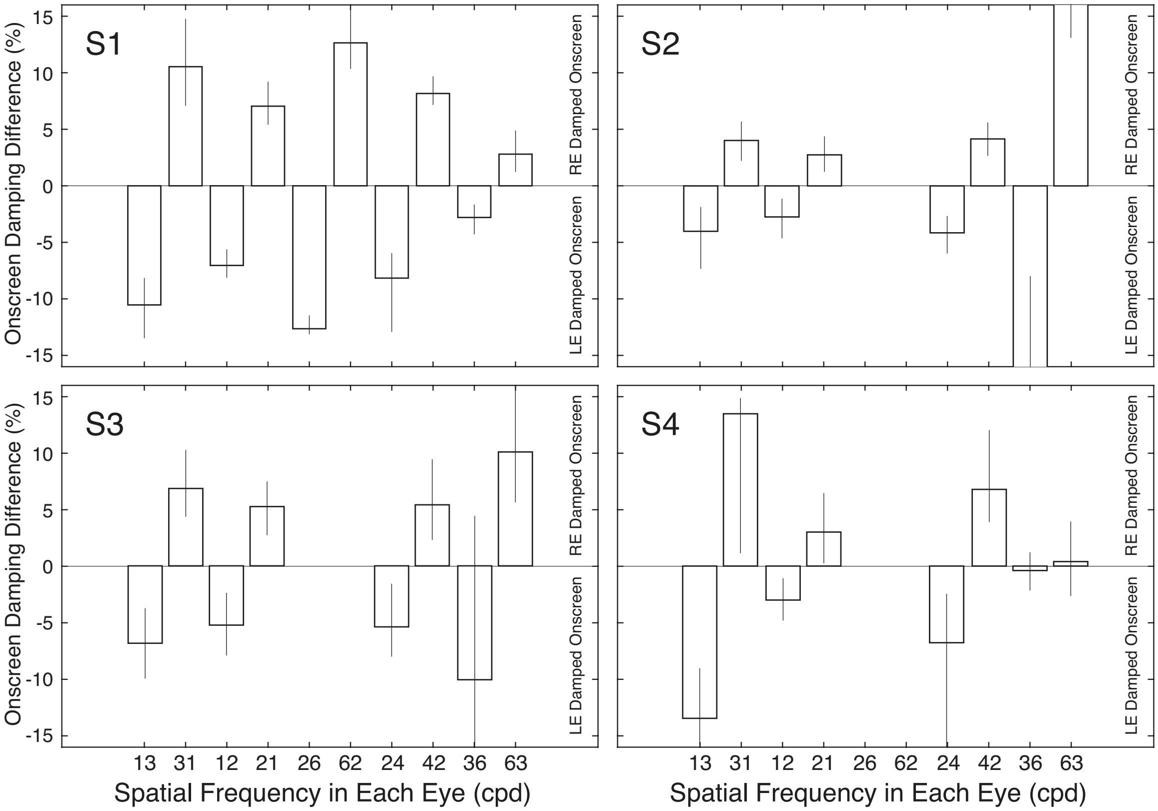

Figure 3.S4. Experiment 3 points of subjective equality (PSEs) for all observers and conditions. The data from all four observers follow the same qualitative pattern. Note that for observer S2, the PSEs for the rightmost two conditions are larger than the scale of the plot.

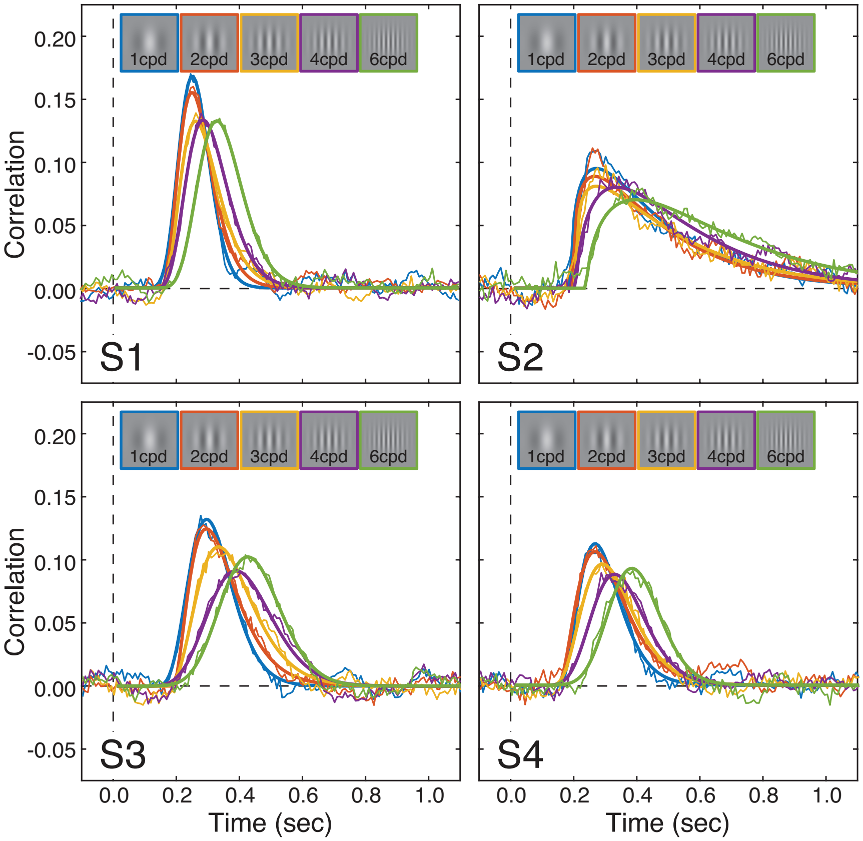

Figure 3.S5. Experiment 4 stimuli and cross-correlograms for all observers and conditions.

CHAPTER 4

ABSTRACT

BINDING OF CHROMATIC SIGNALS ACROSS TIME DURING VISUAL MOTION PERCEPTION

Benjamin M. Chin*

Johannes Burge

*work performed jointly with Michael Barnett

A core problem for the visual system to solve is the binding of sensory signals across time. In the retina, light is encoded by three types of cone photoreceptors. Of these, the S-cone signals have the longest temporal processing delays, and L-cone signals the shortest. This predicts that when visual stimuli move, spatial position signals driven by S-cones should lag behind signals driven by L-cones. We investigate how the visual system binds L-cone and S-cone spatial modulations when they are both present in a moving stimulus. Three observers tracked, with a cursor, the position of a chromatic Gabor conducting a horizontal random walk. The Gabor was composed of L-cone-directed and S-cone-directed modulations whose ratios define polar angles (i.e. color directions) in cone contrast space. We measured tracking performance for stimuli in twelve color directions, with six log-spaced contrasts in each direction. To analyze the data, we computed the cross-correlation between target and tracking velocities. This yields an estimate of the impulse response function associated with the signals that drive tracking. We use time-to-peak of the estimated impulse response functions to estimate processing latency for each stimulus condition. For all subjects, we found that i) temporal lag decreases as contrast increases, for all color directions, and ii) nominally L-cone isolating stimuli are associated with smaller lags than nominally S-cone isolating stimuli, when contrast is equated. A model based on two underlying chromatic mechanisms

accounts for the data well. Each mechanism is determined by a weighted sum of nominal L- and S-cone contrasts. For all observers, one mechanism dominates for the majority of color directions. The dominant mechanism weights L-cone contrast 30-60x higher than weights on S-cone contrast. Future work will examine how the visual system binds other kinds of signals, such as the output of spatial frequency selective channels.

4.1 Introduction

The Binding Problem in perception refers to the problem of integrating information from a multitude of signals. When we perceive an object, we combine its various properties—shape, color, motion, and many other properties—into a coherent percept. At a finer level, perceiving any one of these properties involves the integration of multiple signals. Shape perception involves the integration of contours, and color perception involves the integration of outputs from at least three chromatic pathways.

Much research on the Binding Problem over the past several decades has focused on binding between properties. Significant effort has been made to probe the limits of binding between properties. Binding failures such as illusory conjunctions (ICs) have generated strong interest. For example, a red triangle and a green square might be misperceived as a green triangle and a red square. These errors are typically induced by brief stimulus exposures or rapid temporal modulations of the stimulus (Bartels & Zeki, 2006). Binding between properties is also considered important for visual search. Certain conjunctions of properties can cause visual search to be serial rather than parallel, such as when the search target is a green letter ‘H’ among green letter ‘X’ and brown letter ‘H’ distractor stimuli (Treisman & Gelade, 1980).

Multisensory cue combination is another manifestation of the Binding Problem that has inspired a large amount of research. When observers receive conflicting estimates about an object property, such as size, from different sensory modalities, such as vision and touch, they often do not notice the discrepancy. Instead, they combine the conflicting estimates into a single estimate (Ernst & Banks, 2002). In many cases, the estimates are combined in a Bayes-optimal manner; the final estimate is a weighted sum of estimates from each sensory modality, with the weights determined by the relative reliability of the estimates. Such Bayes-optimal cue combination computations have also been found to apply when combining cues within a sensory modality, such as when using binocular disparity and texture cues to estimate 3D slant (Hillis et al., 2004).

Perceiving even a single property, such as color, requires the Binding Problem to be solved. In the first stage of color processing, light energy is transduced into electrical signals by three classes of cone photoreceptors (the L-, M-, and S- cones) located in the retina. The spectral sensitivity functions of each cone type have been well-characterized by the color science community (Stockman & Sharpe, 2000). This enables the Method of Silent Substitution: the precise modulation of cone activity by presenting visual stimuli with the appropriate light wavelength spectra. The ability to precisely modulate cone activity has facilitated a large body of research on post-receptoral mechanisms that combine cone signals to support color perception. It is widely believed that there are three cone-opponent mechanisms: two chromatic mechanisms, a red-green (L-M) and a blue-yellow (S-(L+M)) mechanism, and an achromatic mechanism (L+M) (Jameson & Hurvich, 1955; Krauskopf, Williams, & Heeley, 1982, Derrington, Krauskopf, & Lennie, 1984).

Previous work has investigated the processing latencies of the cone-opponent pathways. Reaction times to modulations in the (S-(L+M)) pathway are slower than those to modulations in the L-M opponent pathway (Smithson & Mollon, 2004; McKeefry, Parry, & Murray, 2003). These results are consistent with neurophysiological evidence suggesting a sluggish S-cone pathway (Cottaris & De Valois, 1998). These latency differences raise the question of how the visual system binds signals from the different cone types in time: a Temporal Binding Problem.

The Temporal Binding Problem is particularly relevant to the perception of chromatic stimuli that move. Most chromatic stimuli modulate activity all three cone types to varying degrees. Under photopic conditions, the cones constitute the earliest stage of motion processing. For a moving stimulus, position signals derived from S-cone modulations should lag behind position signals derived from L-cone and M-cone modulations due to the longer processing latency for S-cones. Thus, without solving the Temporal Binding Problem, a single moving stimulus would appear as multiple moving stimuli at different positions. But human beings typically perceive moving stimuli to be rigid. This suggests a single, bound percept. We were particularly interested in the temporal processing characteristics associated with the rigid stimulus.

4.2 Results

We leveraged a recently developed psychophysical paradigm, continuous target-tracking, to investigate how the Temporal Binding Problem is solved between L-cone and S-cone modulations. Due to the novel nature of the task, we made no assumptions about the mechanisms binding signals from the two cone types. Instead, we set out to uncover the mechanisms from the experimental data. Observers tracked Gabor targets comprised of both L-cone and S-cone spatial modulations. We determine that when a stimulus is comprised of both L-cone and S-cone modulations, the visual system heavily prioritizes L-cone signals during target-tracking. We present a descriptive mechanistic model that quantifies the relative contributions of L-cone and S-cone modulations.

All chromatic stimuli lay in a two-dimensional space of L-cone and S-cone modulations relative to a grey background (Fig. 4.1A). The angle θ specifies a chromatic axis in this space. Twelve chromatic axes between –86.25° and 90° were investigated, with six contrast levels tested per axis. The 0° axis corresponds to pure L-cone modulations, and the 90° axis corresponds to pure S-cone modulations. Cone contrast modulations were spatial in nature; all stimuli were 1 cyc/° Gabors. Notably, because all Gabors were in sine phase, each stimulus was symmetric, containing both positive and negative modulations of equal magnitude across its chromatic axis.

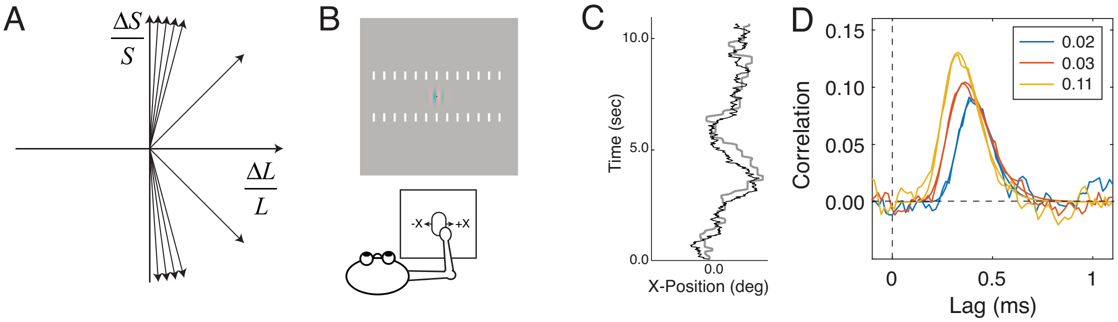

Figure 4.1. Stimuli, task, and analysis of tracking data. **A** Directions of cone contrast modulation for stimuli used in the experiment, shown in the L-S plane. The origin corresponds to the gray background. **B** On each trial, the observer tracked, with a mouse cursor, a Gabor stimulus following a horizontal random walk across the center of the screen. **C** Example target-tracking performance on a single trial. The solid black trace indicates the horizontal random walk taken by the stimulus. The gray trace indicates the position of the observer's cursor. **D** Cross-correlograms in the target tracking task derived from target tracking performance, for three levels of pure L-cone modulations (0° in Fig. 4.1A). The cross-correlograms become narrower, and have shorter times to peak, as contrast increases.

On each trial, a Gabor target underwent a horizontal random walk on the screen (Fig. 4.1BC). Cross-correlation between the target and response motions yields a unimodal cross-correlogram (Fig. 4.1D). Assuming the visuomotor system is linear, this cross-correlogram equals the temporal impulse response function of the visuomotor system to the target. The time-to-peak of the cross-correlogram provides an estimate of response lag in the visuomotor system.

We investigated how lag changes as a function of chromatic modulation direction and cone contrast. Figure 4.2 shows how lag changes with contrast for 12 directions between –86.25° and 90° in the space of L-cone and S-cone modulations. Across all three observers, lag decreases with cone contrast. Pure S-cone modulations (90°) have the highest lags across a large range of contrasts, while pure L-cone modulations (0°) have the lowest lags. Response lags along non-cardinal directions in the space (45°, 75°, etc.) lie in between response lags to pure S-cone and pure-L-cone modulations.

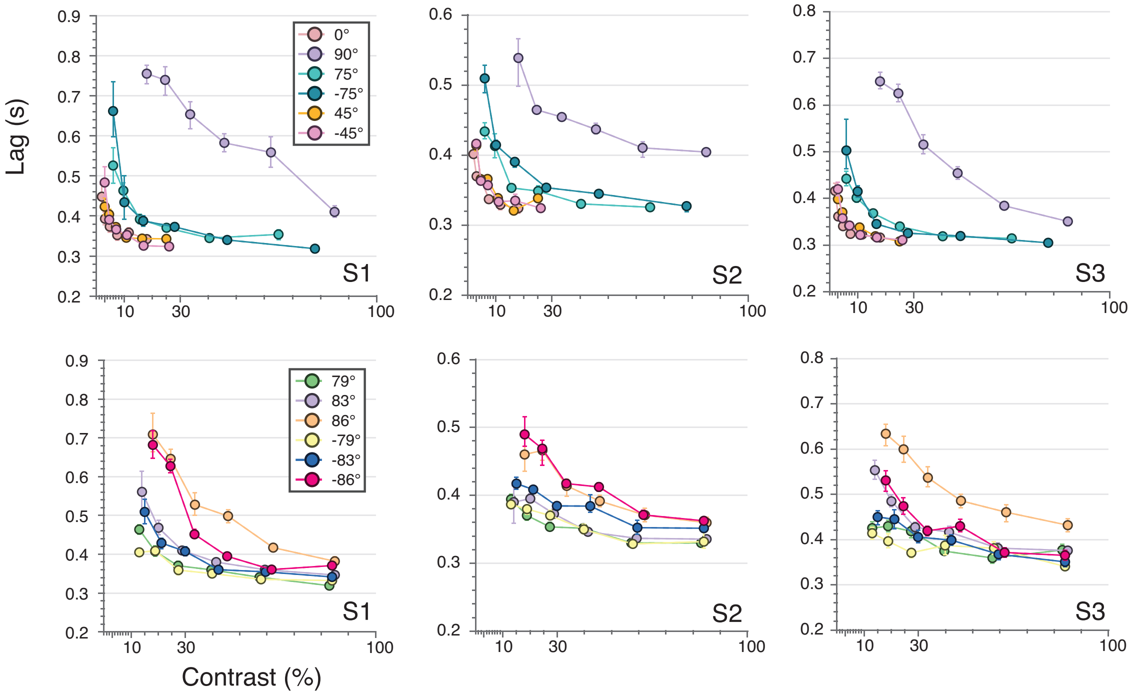

Figure 4.2. Estimated response lag as a function of cone contrast, for all chromatic directions and observers. Error bars represent 68% bootstrapped confidence intervals on estimated response lag. Response lag decreases with cone contrast, and increases as chromatic direction approaches 90°.

Response lags along the ±45° and ±75° chromatic axes are more similar to pure L-cone modulations than pure S-cone modulations. Only for steep chromatic axes greater than 75° or less than -75° do response lags become more similar to pure S-cone modulations. We quantify the relative dominance of L-cone modulations with a two-mechanism, two-stage model of response lag. The first stage of the model describes the color mechanisms underlying target-tracking behavior:

$$m_1 = \left| a_1 c_L + b_1 c_S \right| \tag{1}$$

$$m_2 = \left| a_2 c_L + b_2 c_S \right| \tag{2}$$

where m_1 and m_2 are both outputs of color mechanisms, c_L and c_S are contrasts of the L-cone and S-cone modulations respectively, a_1 is the weight on L-cone contrast for the first color mechanism, a_2 is the weight on L-cone contrast for the second color mechanism, b_1 is the weight on S-cone contrast for the first color mechanism, and b_2 is the weight on S-cone contrast for the second color mechanism. Each color mechanism output is equivalent to the rectified dot product of the stimulus vector $\left(c_L, c_S\right)$ with the weight vector $\left(a, b\right)$. Thus, it is proportional to the projection of the stimulus vector onto the weight vector (Fig. 4.3A). The outputs of both color mechanisms are then combined into a final mechanism output:

$$m = \max\left(m_1, m_2\right) \tag{3}$$

The final mechanism output m is then converted into response lag via an exponential decay function (Fig. 4.3A):

$$l = A\exp\left(-m\right) + l_0 \tag{4}$$

where l is response lag, A is a parameter controlling the slope of the function, and l_0 is the minimum lag achievable.

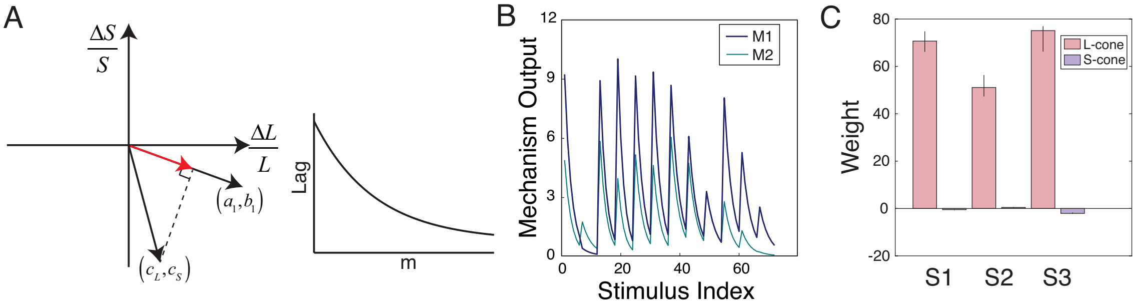

Figure 4.3. Model of lag as a function of chromatic direction. **A** Schematic for illustrating the operation of a color mechanism. The output of the color mechanism is proportional to the projection of the stimulus vector onto the weight vector. The projection is indicated by the length of the red arrow. **B** Output of both color mechanisms for each stimulus index, for observer S2. Each stimulus index corresponds to a specific combination of chromatic direction and contrast. The output of one mechanism is consistently higher than that of the other mechanism, for the majority of stimuli. There is a small subset of stimuli for which the other mechanism has higher output. **C** Weights on L-cone and S-cone contrast for the dominant color mechanism, for all three observers. Weights on L-cone contrast are 30-60 times larger than weights on S-cone contrast. Error bars represent 68% bootstrapped CIs on L-cone and S-cone weights.

We fit the model to response lag data from all three observers separately, using root mean squared error (RMSE) between observer latencies and model predictions as the cost function. The free parameters were a_1, a_2, b_1, b_2, A, and l_0. Model fits are shown in Figure 4.S1. The model captures the pattern of data across all chromatic axes.

An examination of the color mechanism outputs given the fit values for parameters a_1, b_1, a_2, and b_2 reveals that one of the two mechanisms has a larger response than the other for the majority of chromatic directions, and thus dominates the model response (Fig. 4.3B). We find that for the more dominant color mechanism, the L-cone weight a_1 is much larger than the S-cone weight b_1. This pattern is consistent across all observers (Fig. 4.3C). These findings indicate that during target-tracking, the binding process of the visual system preferentially weights L-cone modulations over S-cone modulations.

4.3 Methods

Subjects

Three human observers (two male, one female) between 18 and 65 years of age participated in the experiment: 2 were authors, and the third was naïve to the purposes of the experiment. All had normal or corrected-to-normal acuity. The research protocol was approved by the Institutional Review Board of the University of Pennsylvania and was in accordance with the Declaration of Helsinki. The study was preregistered. All experiments were performed in MATLAB 2017a using Psychtoolbox version 3.0.12 (Brainard, 1997). Psychophysical data are presented for each individual human observer. Bootstrapped SEs or CIs are presented on all data points unless otherwise noted.

Equipment

Stimuli were presented on a ViewSonic G220fb 40.2cm X 30.3cm cathode ray tube monitor with 1280 X 1024 pixel resolution, and a refresh rate of 60 Hz. At the 92.5

cm viewing distance, the monitor subtended a FOV of 24.5° X 18.6° of visual angle. The observer's head was stabilized with a chin-and-forehead rest. Primaries of the CRT were measured using a PR-650 spectroradiometer (PhotoResearch).

Stimuli

Stimuli were colored Gabor patches comprised of bipolar L-cone and S-cone directed spatial modulations around a constant background light (mean luminance ~30.75 cd/m^2, chromaticity: x≈0.326, y≈0.372). All Gabor patches were in sine phase and had carrier spatial frequencies of 1 cpd. All Gabor had an octave bandwidth of 1.5, corresponding to a Gaussian window with a standard deviation of 0.6°. Stimuli were created at 12 directions in the L-S plane. Within a direction, stimuli were created at six log-spaced contrast levels. The maximum and minimum contrast levels for each direction are shown in Tables 3.1 and 3.2. Cone contrast values for all stimuli were computed using the Stockman and Sharpe 2° cone fundamentals. We use the convention that the cone contrast values shown in Tables 1 and 2 represent the cone contrast corresponding to the peak modulation of the carrier sinewave for each Gabor.

Table 3.1: Chromaticity coordinates of Stimuli in the L-S plane

	-75°	-45°	0°	45°	75°	90°
Maximum Contrast	78%	26%	18%	25%	65%	85%
Minimum Contrast	6%	3%	2%	3%	6%	18%

Table 3.2: Chromaticity coordinates of Stimuli in the L-S plane

	-86.25°	-82.5°	-78.75°	78.75°	82.5°	86.25°
Maximum Contrast	78%	26%	18%	25%	65%	85%
Minimum Contrast	18%	15%	13%	13%	14%	18%

Target-tracking procedure

Tracking data was collected from each observer in blocks of individual runs. Each run was initiated with a mouse click, which caused the target and a small dark mouse cursor to appear in the center of the screen. After a stationary period of 500ms, the target began a one-dimensional horizontal random walk (i.e. Brownian motion) for eleven seconds. The task was to track the target as accurately as possible with a small dark mouse cursor.

Target-tracking psychophysics: Onscreen stimuli

Data was collected in 72 conditions, with each condition defined by its chromatic direction and contrast. Data was collected in 40 intermixed blocks of 36 runs each for a total 20 runs per condition. For each observer, data was collected over 6 hours, split among 4 sessions of 1.5 hours. Data for the following directions was collected in the first 2 sessions: -75°, -45°, 0°, 45°, 75°, and 90°. Data for the remaining 2 sessions was collected in the subsequent 2 sessions.

Target-tracking psychophysics: Target motion

For the tracking experiments, the target stimulus performed a random walk on a gray background. The region of the screen traversed by the target was flanked by two horizontal sets of thirteen vertically-oriented picket fence bars (Fig. 4.2A).

The x-positions of the target on each time step $t+1$ were generated as follows

$$x(t+1) = x(t) + \varepsilon_x \; ; \quad \varepsilon_x \sim N(0,Q) \tag{5}$$

where ε_x is a random sample of Gaussian noise and Q is the drift variance. The random sample determines the change in target position between the current and the next time step. The drift variance determines the expected magnitude of the position change on each time step, and hence the overall variance of the random walk. The variance of the walk positions across multiple walks $\sigma^2(t) = Qt$ is equal to the product of the drift variance and the number of elapsed time steps. The value of the drift variance in our task (0.8mm per time step) was set such that the mean of the velocity distribution was ~4 °/s and that 90% of the velocity distribution fell within ±10 °/s.

The effective on-screen positions of the images are obtained by convolving the on-screen target images with the temporal impulse response function

$$\tilde{x}(t) = x(t) * h(t) \tag{6}$$

where $h(t)$ is a temporal impulse response function corresponding to a specific frequency. Convolving the target velocities with the impulse response function gives the velocities of the effective target images. Integrating these velocities across time gives the effective target positions.

To determine the impulse response function relating the target and response, we computed the zero-mean normalized cross-correlations between the target and response velocities

$$\rho\left(\tau;\dot{x},\dot{\tilde{x}}\right)=\frac{1}{\left\|\dot{x}\left(t\right)\right\|\left\|\dot{\tilde{x}}\left(t\right)\right\|}\left[\sum_{t=1}^{N}\left(\dot{x}\left(t\right)-\bar{\dot{x}}\right)\left(\dot{\tilde{x}}\left(t+\tau\right)-\bar{\dot{\tilde{x}}}\right)\right] \tag{7}$$

where τ is the lag, $\dot{x}$ and $\dot{\tilde{x}}$ are the target and response velocities. Assuming a linear system, when the input time series (i.e. the target velocities) is white, as it is here by design, the cross-correlation with the response gives the impulse response function of the system.

To compute the normalized cross-correlations, we did not include the first second of each eleven second tracking run so that observers reached steady state tracking performance. The mean cross-correlation functions shown in the figures were obtained by first computing the normalized cross-correlation in each run (Eqn. 7), and then averaging these cross-correlograms across runs in each condition.

Log-normal fits to mean cross-correlograms

To summarize the mean cross-correlograms, we fit a log-Gaussian-shaped function using maximum likelihood methods. The form of the fitted function was given by

$$\rho\left(\tau\right)=A\exp\left[-0.5\left(\left(\ln\left(\tau\right)-m\right)/s\right)^{2}\right] \tag{8}$$

where A is the amplitude, and m and s are the parameters determining the shape and scale of the fit, respectively. The mode (i.e. peak) of the function can be used as a measure of delay, and is given by $\exp\left(m\right)$.

4.4 Discussion

Summary of results and relation to other work

We have investigated how processing latency, as quantified by the target-tracking task, is influenced by L-cone and S-cone modulations. Our findings can be summarized as follows: 1) S-cone modulations require more processing time than L-cone modulations, 2) L-cone modulations primarily determine processing latency when both L-cone and S-cone modulations are present in a stimulus, and 3) a single post-receptoral mechanism captures the pattern of processing latency data for a majority of angles in the L-S plane.

The finding that S-cone modulations are associated with longer processing latency than L-cone modulations coheres with findings by McKeefry, Parry, and Murray (2003), as well as Smithson and Mollon (2004). A key distinguishing feature of the present research is the absence of assumptions about the mechanisms underlying continuous target-tracking behavior; modulations targeted individual cone classes rather than putative post-receptoral color mechanisms. Previous studies, on the other hand, modulated their stimuli along the L-M and S-(L+M) cone-opponent pathways. Our results suggest that processing latency in continuous target tracking can be best accounted for by two mechanisms that differ from the canonical cone-opponent pathways.

General Discussion

From an ecological standpoint, there is an inherent tradeoff between prioritizing L-cone modulations and prioritizing S-cone modulations. Information arriving from L-cone modulations arrives sooner, supporting quicker action responses. On the other hand, S-cone modulations also contain information about the stimulus. It is likely that in some tasks, it is beneficial for the visual system to wait for information from S-cone modulations to arrive. The current study suggests that in the task of continuous target-tracking, the visual system prioritizes information that arrives sooner, rather than integrating information from slower-arriving signals.

Research on multisensory integration and cue combination, which are manifestations of the Binding Problem, has identified normative principles governing the integration of information. In many cases, when a sensory-perceptual system receives conflicting estimates of a stimulus property, the estimates are combined via a weighted average that is optimal from a Bayesian perspective (Ernst & Banks, 2002; Hillis et al., 2004). This suggests that optimality is a useful lens through which human perception can be understood. Theoretical work might explore whether the preferential weighting of L-cone modulations in the current study can be predicted by normative principles, or whether such a weighting scheme is suboptimal.

Future directions

It has long been known that S-cone modulations are more difficult to detect than L-cone modulations (Eskew et al., 1999). Future work will thus examine the extent to which slower processing latencies and downweighting of S-cone modulations can be explained by the lower detection sensitivity of the visual system to S-cone modulations.

For each observer in our study, there is a direction in the L-S plane that is orthogonal to the dominant color mechanism determined for that observer. Under our model, that chromatic direction should elicit no response from the dominant color mechanism; a so-called ‘null’ direction. Future work will examine whether the null direction for each observer corresponds to that observer’s true tritan line, which might deviate from the theoretical tritan line. Psychophysical work has demonstrated that the true tritan line can be empirically determined via an adaptation paradigm (Smithson, Sumner, & Mollon, 2003).

Color is not the only dimension of visual stimuli along which significant differences in processing latency have been found. Differences in processing latency for different spatial frequencies have been well established: higher spatial frequencies are processed with longer latency (Parker, 1980; Mihaylova, Stomonyakov, & Vassilev, 1998). Future work will explore how the temporal binding problem is solved for compound Gabors consisting of more than one carrier spatial frequency. Of particular interest is the question of whether low spatial frequencies are prioritized during the binding process, owing to their shorter temporal latencies.

4.5 Supplement

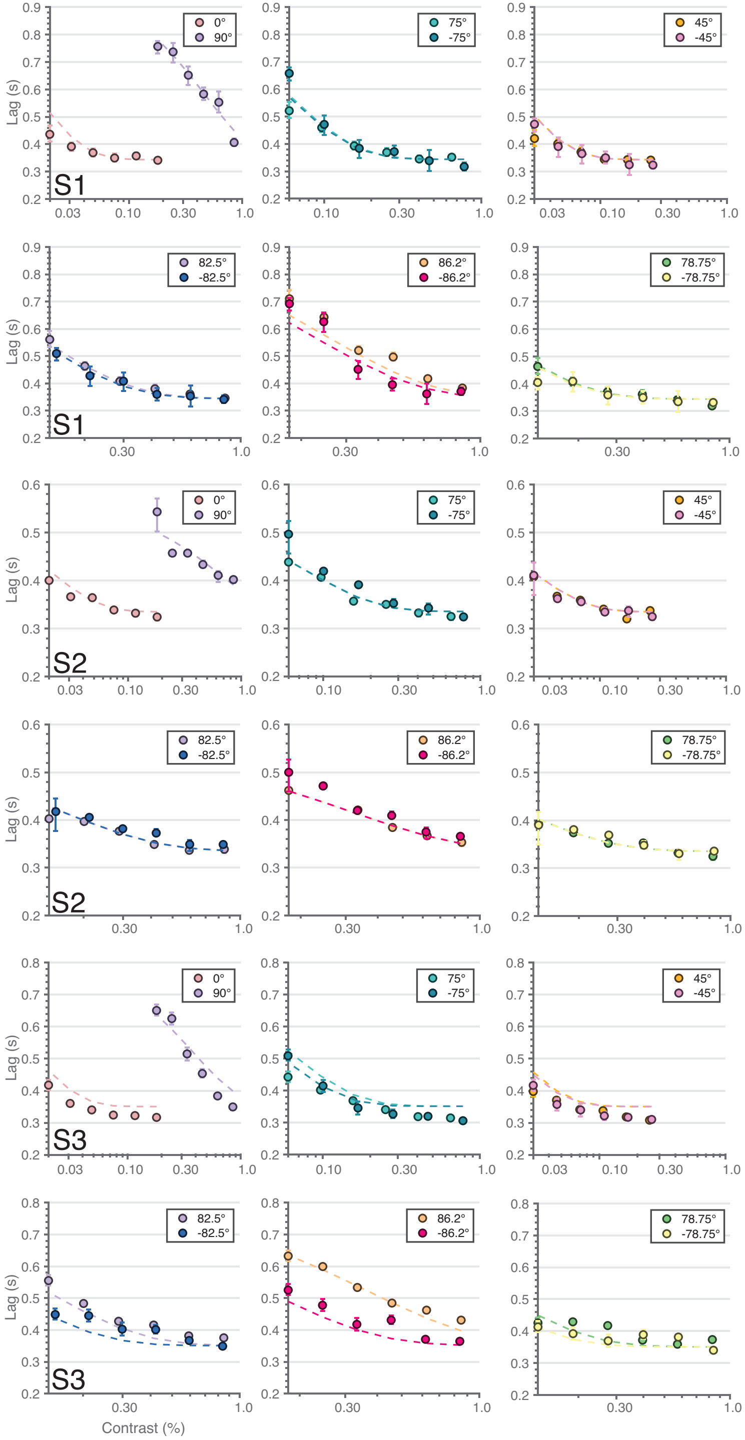

Figure 4.S1. Fits of the two-mechanism model to response lag data for all observers (S1, S2, and S3) and all conditions. Dotted lines represent fits of the model. Solid points represent observer data. Error bars represent 68% bootstrapped confidence intervals on estimated response lag. Data from pairs of chromatic directions were split across plots for viewing clarity.

CHAPTER 5: SUMMARY OF CONTRIBUTIONS

I have investigated the relationship between basic properties of visual stimuli and the perception of motion. It has long been known in the field of vision science that properties such as spatial frequency and color impact motion perception, but a full quantitative characterization of their impact has yet to be achieved. The work discussed in this dissertation constitutes an important step towards achieving such a characterization. The present research is rooted in hypotheses about the computations performed by the visual system to extract motion. These hypotheses yield testable predictions about how different stimulus properties influence motion perception. The predictions have been validated by a rich set of psychophysical data collected across an array of interlocking experiments.

5.1 Contributions of Chapter 2

Normative Bayesian ideal observer models have exploded in popularity in recent years, as a means to predict and understand neural properties and behavioral performance. However, many recent ideal observers have dispensed with a characteristic that was key to their early success. Early ideal observer models were image-computable: they explicitly modeled the flow of visual information from the retinal image to the perceptual estimate, making optimal use of the statistics relating task-relevant image features to task-relevant latent variables. In the 1980s and early 1990s, image-computable ideal observer models markedly advanced spatial vision and visual neuroscience, yielding deep insights about simple tasks like target detection or orientation discrimination in noise.

Image-computable ideal observer analysis fell out of favor in the 1990s because it resisted successful application to more complicated tasks (e.g. motion estimation) with more complicated (e.g. naturalistic) stimuli. With natural images, the probabilistic relationship between image features and the task-relevant variable is generally not known. Indeed, many modern Bayesian ideal observers cannot be directly applied to the patterns of light falling in the retinas, and instead must rely on assumptions (which may be incorrect) about the information available in the encoded image. An exception to this trend is our recent previous work (Burge & Geisler, 2015), in which we developed an image-computable ideal observer for speed estimation with naturalistic stimuli and used it to fit human performance. This provided us with a measure of how well human performance compares to the ideal observer, but left open the more fundamental question of why human performance falls short.

Chapter 2 answers that question, providing a suite of new tools and mathematical results in the process that should benefit research going forward. We develop an experimental protocol that can distinguish two distinct sources of human inefficiency: internal noise and suboptimal computations. A complementary computational model predicts the behavioral signatures of each source without fitting parameters to the data. By confirming the predictions, we find that human observers perform near-optimal computations on natural stimuli for estimating speed, underperforming the ideal because of noise rather than the systematic misuse of the available information. Furthermore, we find that human behavioral variability is majorly impacted by external sources of uncertainty (i.e. stimulus variability). External variability i) shapes the optimal computations, ii) dictates the pattern of human performance, and iii) predicts the partition of behavioral variability (i.e. the relative importance of external and internal variability). These findings motivate continued efforts to understand how natural stimulus variability impacts perceptual performance. They are part of a larger trend in vision and systems neuroscience to characterize the impact of typically unstudied sources of uncertainty on behavioral and neural measures.

5.2 Contributions of Chapter 3

The Pulfrich effect is a striking visual illusion that has been known for 100 years. The effect is well-understood: differences in processing latency between the eyes cause oscillating targets in the frontal plane to be misperceived as moving along a near-elliptical motion trajectory in depth. These differences in processing latency can be induced by interocular differences in luminance (Pulfrich, 1922) or blur (Burge, Rodriguez-Lopez, & Dorronsoro, 2019). This explanation does not account for all Pulfrich-like phenomena, however.

Anomalous Pulfrich percepts have been occasionally reported (Emerson & Pesta, 1992; Harker & O’neal, 1967; Trincker,1953; Weale, 1954). Specifically, observers sometimes report perceiving near-elliptical motion trajectories that are misaligned in depth relative to the true direction of motion. The standard explanation does not account for these percepts; interocular differences in processing latency predict perceived motion trajectories that are aligned with the true path of motion. Although various explanations have been proposed regarding the cause of anomalous Pulfrich percepts, no scientific consensus exists.

Chapter 3 proposes a novel explanation for anomalous Pulfrich percepts and validates it with results from a set of interlocking experiments. The anomalous Pulfrich effect is caused by interocular differences in temporal integration periods. For oscillating motion, these differences in the temporal integration period effectively damp the motion amplitude in one eye relative to the other. Under these circumstances, stereo-geometry predicts the illusory misorientation of the motion trajectory in depth. The anomalous Pulfrich effect can therefore be thought of as a dynamic analog to the ‘geometric effect’ in stereo-slant perception (Ogle, 1950).

Our findings motivate continued efforts to characterize the variations in temporal processing properties across different visual stimuli. Our work demonstrates that differences in temporal processing pose challenges to the visual system, and in some cases, impact its ability to accurately estimate properties of the environment. Given that, in most circumstances, the visual system computes (largely) accurate estimates of motion and other properties of the environment, it is thus important to understand the compensatory computations that the visual system uses to solve these challenges. We expect the current paper to be an early contribution to a series of papers on the general topic.

5.3 Contributions of Chapter 4

The Binding Problem is ubiquitous in perception; how should sensory-perceptual signals with different characteristics be combined into a coherent percept? Much work in perceptual psychology has investigated how human beings solve the Binding Problem between properties, such as between motion and color, or between color and orientation. The popularity of such work perhaps belies the fact that estimating even a single property, spatial position, requires the Binding Problem to be solved. In the case of estimating spatial position, the Binding Problem arises early in the visual system; how are signals from L, M, and S cones in the retina bound across time during stimulus motion? The question is pertinent because it is well known that S-cones have longer processing latency than L-cones. This implies that during motion, spatial position estimates derived from S-cone signals should lag behind spatial position estimates derived from L-cone signals. We find that the visual system prioritizes L-cone signals during stimulus motion by having observers continuously track visual stimuli comprised of both L-cone and S-cone spatial modulations. Our findings motivate future work to understand how the visual system binds signals across time for other perceptual tasks and stimulus properties.

BIBLIOGRAPHY

Abbey, C. K., & Eckstein, M. P. (2014). Observer efficiency in free-localization tasks with correlated noise. *Frontiers in Psychology*, *5*(345).

Adams, W. J., Elder, J. H., Graf, E. W., Leyland, J., Lugtigheid, A. J., & Muryy, A. (2016). The Southampton-York Natural Scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports*, *6*(35805).

Adelson, E. H., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, *2*, 284–299.

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

Anstis, S. M. (1977). Apparent Movement. In R. H. Perception, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of Sensory Physiology: Vol. VIII*. Springer-Verlag.

Anstis, S. M. (1980). *The perception of apparent movement* (P. R. S. L. S. B, Trans.; p. 290,153-168).

Backus, B. T., Banks, M. S., Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*(6), 1143–1170.

Badcock, D. R., & Derrington, A. M. (1985). Detecting the displacement of periodic patterns. *Vision Research*, *25*(9), 1253–1258. https://doi.org/10.1016/0042-6989(85)90040-9

Bair, W., & Movshon, J. A. (2004). Adaptive temporal integration of motion in direction-selective neurons in macaque visual cortex. *Journal of Neuroscience*, *24*(33), 7305–7323. https://doi.org/10.1523/JNEUROSCI.0554-04.2004

Banks, M. S., & Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vision Research*, *38*(2), 187–194.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research*, *27*, 1915–1924.

Bartels, A., & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research*, *46*, 2280–2286. https://doi.org/10.1016/j.visres.2005.11.017.

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

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Verlag.

Blake, R., & Cormack, R. H. (1979). On utrocular discrimination. *Perception & Psychophysics*, *26*, 53–68.

Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision*, *15*(3), 1–16. https://doi.org/10.1167/15.3.14

Bonnen, K., Huk, A. C., & Cormack, L. K. (2017). Dynamic mechanisms of visually guided 3D motion tracking. *Journal of Neurophysiology*, *118*(3), 1515–1531. https://doi.org/10.1152/jn.00831.2016

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, *10*, 433–436.

Burge, J. (2020). Image-computable ideal observers for tasks with natural stimuli. *Annual Review of Vision Science*, *6*, 491–517. https://doi.org/10.1146/annurev-vision-030320-041134

Burge, J., & Cormack, L. K. (2020). *Target tracking reveals the time course of visual processing with millisecond-scale precision*. https://doi.org/10.1101/2020.08.05.238642

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *Journal of Neuroscience*, *30*, 7269–7280.

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

Burge, J., & Geisler, W. S. (2012). Optimal defocus estimates from individual images for autofocusing a digital camera. In *Proceedings of the SPIE* (p. 82990). https://doi.org/10.1117/12.912066.

Burge, J., & Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal of Vision*, *14*.

Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts human performance. *Nature Communications*, *6*(7900).

Burge, J., & Jaini, P. (2017). Accuracy Maximization Analysis for Sensory-Perceptual Tasks: Computational Improvements, Filter Robustness, and Coding Advantages for Scaled Additive Noise. *PLOS Computational Biology*, *13:e1005281*.

Burge, J., McCann, B. C., & Geisler, W. S. (2016). Estimating 3D tilt from local image cues in natural scenes. *Journal of Vision*, *16*(2).

Burge, J., Rodriguez-Lopez, V., & Dorronsoro, C. (2019). Monovision and the Misperception of Motion. *Current Biology*, *29*, 2586–2592.

Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer Inconsistency*. Journal of the Optical Society of America*, *5*, 617–627.

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

C, B. D. (1981). Temporal summation of moving images by the human visual system. *Proc. R. Sot. Lond. E*, *211*, 321–339.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat Rev Neurosci*, *13*, 51–62.

Cottaris, N. P., & De Valois, R. L. (1998). Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature*, *29;395(6705):896-900*. https://doi.org/10.1038/27666.

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology*, *357*, 241–265. https://doi.org/10.1113/jphysiol.1984.sp015499

Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences*, *95*, 13988–13993.

Drugowitsch, J., Wyart, V., Devauchelle, A.-D., & Koechlin, E. (2016). Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality. *Neuron*, *92*, 1398–1411.

Emerson, P. L., & Pesta, B. J. (1992). A generalized visual latency explanation of the Pulfrich phenomenon. *Perception & Psychophysics*, *51*(4), 319–327. https://doi.org/10.3758/bf03211625

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Fleming, R. W., & Storrs, K. R. (2019). Learning to see stuff. *Current Opinion in Behavioral Sciences*, *30*, 100–108.

Frazor, R. A., Albrecht, D. G., Geisler, W. S., & Crane, A. M. (2004). Visual cortex neurons of monkeys and cats: Temporal dynamics of the spatial frequency response function. *Journal of Neurophysiology*, *91*(6), 2607–2627. https://doi.org/10.1152/jn.00858.2003

Frechette, E. S., Sher, A., Grivich, M. I., Petrusca, D., Litke, A. M., & Chichilnisky, E. J. (2005). Fidelity of the ensemble code for visual motion in primate retina. *Journal of Neurophysiology*, *94*, 119–135.

Freeman, T. C. A., Champion, R. A., & Warren, P. A. (2010). A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology*, *20*, 757–762.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. *J Comp Neurol*, *201*, 519–539.

Gattass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience*, *8*, 1831–1845.

Gegenfurtner, K., & Sharpe, L. T. (Eds.). (1999). Chromatic detection and discrimination. In *Color Vision: From Genes to Perception* (pp. 345–368). Cambridge University Press.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267–314.

Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks. *Journal of Vision*, *9*(17).

Geisler, W. S., & Perry, J. S. (2009). Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience*, *26*, 109–121.

Gekas, N., Meso, A. I., Masson, G. S., & Mamassian, P. (2017). A Normalization Mechanism for Estimating Visual Motion across Speeds and Scales. *Current Biology*, *27*, 1514–1520.

Goettker, A., Braun, D. I., Schütz, A. C., & Gegenfurtner, K. R. (2018). Execution of saccadic eye movements affects speed perception. *Proceedings of the National Academy of Sciences*, *115*, 2240–2245.

Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). *Characterizing perceptual learning with external noise*. Cognitive Science.

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Publishing Group*, *17*, 858–865.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (p. 1).

Harker, G. S., & O'neal, O. L. (1967). Some observations and measurements of the Pulfrich phenomenon. *Perception & Psychophysics*, *2*, 438–440.

Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision Research*, *18*(11), 1579–1586. https://doi.org/10.1016/0042-6989(78)90014-7

Hecht, S., Shlaer, S., & Pirenne, M. H. (1942). Energy, Quanta, and Vision. *J Gen Physiol*, *25*, 819–840.

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

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *1;4(12):967-92*. https://doi.org/10.1167/4.12.1.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Neurophysiology*, *160*, 106–154.

Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, *64*(6, Pt.1), 384–404. https://doi.org/10.1037/h0041403

Iyer, A., & Burge, J. (2019). The statistics of how natural images drive the responses of neurons. *Journal of Vision*, *19*(4).

Jaini, P., & Burge, J. (2017). Linking normative models of natural tasks to descriptive models of neural response. *Journal of Vision*, *17*(16).

Jogan, M., & Stocker, A. A. (2015). Signal Integration in Human Visual Speed Perception. *Journal of Neuroscience*, *35*, 9381–9390.

Kane, D., Bex, P., & Dakin, S. (2011). Quantifying "the aperture problem" for judgments of motion direction in natural scenes. *Journal of Vision*, *11*.

Kim, S., & Burge, J. (2018). The lawful imprecision of human surface tilt estimation in natural scenes. *eLife*, *7*.

Kim, S., & Burge, J. (2020). Natural scene statistics predict how humans pool information across space in surface tilt estimation. *PLoS Computational Biology*, *16*(6), 1007947–26. https://doi.org/10.1371/journal.pcbi.1007947

Knöll, J., Pillow, J. W., & Huk, A. C. (2018). Lawful tracking of visual motion in humans, macaques, and marmosets in a naturalistic, continuous, and untrained behavioral context. *Proceedings of the National Academy of Sciences*, *115*, 10486–10494. https://doi.org/10.1073/pnas.1807192115/-/DCSupplemental

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, *51*, 1457–1483.

Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, *1982;22(9):1123-31. doi*, 10 1016 0042-6989 82 90077-3.

Laming, D. R. J. (1979). Choice reaction performance following an error. *Acta Psychologica*, *43*, 199–224.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412.

Lappin, J. S., & Bell, H. H. (1972). Perceptual differentiation of sequential visual patterns. *Perception & Psychophysics*, *12*, 129–134.

Lee, J., Joshua, M., Medina, J. F., & Lisberger, S. G. (2016). Signal, Noise, and Variation in Neural and Sensory- Motor Latency. *Neuron*, *90*(1), 165–176. https://doi.org/10.1016/j.neuron.2016.02.012

Levi, D. M., Harwerth, R. S., & Manny, R. E. (1979). Suprathreshold spatial frequency detection and binocular interaction in strabismic and anisometropic amblyopia. *Investigative Ophthalmology & Visual Science*, *18*(7), 714–725.

Li, R. W., Klein, S. A., & Levi, D. M. (2006). The receptive field and internal noise for position acuity change with feature separation. *Journal of Vision*, *6*, 311–321.

Li, X., Lu, Z.-L., Xu, P., Jin, J., & Zhou, Y. (2003). Generating high gray-level resolution monochrome displays with conventional computer graphics cards and color monitors. *J Neurosci Methods*, *130*, 9–18.

Lit, A. (1949). The magnitude of the Pulfrich stereophenomenon as a function of binocular differences of intensity at various levels of illumination. *The American Journal of Psychology*, *62*(2), 159–181.

Lorenceau, J., & Alais, D. (2001). Form constraints in motion binding. *Nature Neuroscience*, *4*(7), 745–751. https://doi.org/10.1038/89543

Lu, Z.-L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, *35*(19), 2697–2722. https://doi.org/10.1016/0042-6989(95)00025-U

Lyu, S., & Simoncelli, E. P. (2009). Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures. *IEEE Trans Pattern Anal Mach Intell*, *31*, 693–706.

Marx, M. S., & May, J. G. (1983). The relationship between temporal integration and persistence. *Vision Research*, *23*(10), 1101–1106. https://doi.org/10.1016/0042-6989(83)90022-6

McKeefry, D. J., Parry, N. R., & Murray, I. J. (2003). Simple reaction times in color space: The influence of chromaticity, contrast, and cone opponency. *Invest Ophthalmol Vis Sci*, *May;44(5):2267-76. doi*, 10 1167 02-0772.

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

Mihaylova, M., Stomonyakov, V., & Vassilev, A. (1999). Peripheral and central delay in processing high spatial frequencies: Reaction time and VEP latency studies. *Vision Research*, *39*(4), 699–705. https://doi.org/10.1016/S0042-6989(98)00165-5

Min, S. H., Reynaud, A., & Hess, R. F. (2020). Interocular Differences in Spatial Frequency Influence the Pulfrich Effect. *Vision*, *4*(20), 1–13. https://doi.org/10.3390/vision4010020

Mulligan, J. B., Stevenson, S. B., & Cormack, L. K. (2013). *Reflexive and voluntary control of smooth eye movements* (B. E. Rogowitz, T. N. Pappas, & H. Ridder, Eds.; Vol. 8651, p. 86510 1-22). https://doi.org/10.1117/12.2010333

Nachmias, J. (1967). Effect of exposure duration on visual contrast sensitivity with square-wave gratings. *Journal of the Optical Society of America*, *57*(3), 421–427. https://doi.org/10.1364/JOSA.57.000421

Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Research*, *46*, 2465–2474.

Nitzany, E. I., & Victor, J. D. (2014). The statistics of local motion signals in naturalistic movies. *Journal of Vision*, *14*.

Nover, H., Anderson, C. H., & DeAngelis, G. C. (2005). A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *Journal of Neuroscience*, *25*, 10049–10060.

Ogle, K. N. (1950). *Researches in binocular vision*. W B Saunders.

Osborne, L. C., Hohl, S. S., Bialek, W., & Lisberger, S. G. (2007). Time course of precision in smooth-pursuit eye movements of monkeys. *Journal of Neuroscience*, *27*, 2987–2998.

Osborne, L. C., & Lisberger, S. G. (2009). Spatial and temporal integration of visual motion signals for smooth pursuit eye movements in monkeys. *Journal of Neurophysiology*, *102*(4), 2013–2025. https://doi.org/10.1152/jn.00611.2009

Osborne, L. C., Lisberger, S. G., & Bialek, W. (2005). A sensory source for motor variation. *Nature*, *437*(7057), 412–416. https://doi.org/10.1038/nature03961

Parker, D. M. (1980). Simple reaction times to the onset, offset, and contrast reversal of sinusoidal grating stimuli. *Perception & Psychophysics*, *28*(4), 365–368. https://doi.org/10.3758/BF03204396

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America*, *2*, 1508–1532.

Pelli, D. G. (1991). Noise in the visual system may be early. In J. A. Movshon & eds (Eds.), *Computational Models of Visual Processing (Landy MS* (pp. 147–152). MIT Press.

Perrone, J. A., & Thiele, A. (2001). Speed skills: Measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience*, *4*, 526–532.

Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The neural representation of speed in macaque area MT/V5. *Journal of Neuroscience*, *23*, 5650–5661.

Priebe, N. J., & Lisberger, S. G. (2004). Estimating target speed from the population response in visual area MT. *Journal of Neuroscience*, *24*, 1907–1916.

Pulfrich, C. (1922). Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. *Die Naturwissenschaften*, *10*(35), 553–564.

Reynaud, A., & Hess, R. F. (2017). Interocular contrast difference drives illusory 3D percept. *Scientific Reports*, *7*(1), 5587. https://doi.org/10.1038/s41598-017-06151-w

Rodriguez-Lopez, V., Dorronsoro, C., & Burge, J. (2020). Contact lenses, the reverse Pulfrich effect, and anti-Pulfrich monovision corrections. *Scientific Reports*, *10*, 1–16. https://doi.org/10.1038/s41598-020-71395-y

Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, *49*, 2415–2441.

Rucci, M., & Poletti, M. (2015). Control and Functions of Fixational Eye Movements. *Annu Rev Vis Sci*, *1*, 499–518.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*, 1421–1431.

Santen, J., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America. A, Optics and Image Science*, *2*, 300–321. https://doi.org/10.1364/JOSAA.2.000300.

Schneeweis, D. M., & Schnapf, J. L. (1995). Photovoltage of rods and cones in the macaque retina. *Science*, *268*, 1053–1056.

Schrater, P. R., Knill, D. C., & Simoncelli, E. P. (2000). Mechanisms of visual motion detection. *Nature Neuroscience*, *3*, 64–68.

Schrater, P. R., Knill, D. C., & Simoncelli, E. P. (2001). Perceiving visual expansion without optic flow: Abstract: Nature. *Nature*, *410*, 816–819.

Schütt, H. H., & Wichmann, F. A. (2017). An image-computable psychophysical spatial vision model. *Journal of Vision*, *17*(12).

Schwarzkopf, D. S., Schindler, A., & Rees, G. (2010). Knowing with which eye we see: Utrocular discrimination and eye-specific signals in human visual cortex. *PLoS One*, *5*(10). https://doi.org/10.1371/journal.pone.0013775

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

Sebastian, S., Burge, J., & Geisler, W. S. (2015). Defocus blur discrimination in natural images with natural optics. *Journal of Vision*, *15*(16).

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

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research*, *38*, 743–761.

Sinha, B., SR, W, de R. van S., & R. (2018). *Optimal local estimates of visual motion in a natural environment*.

Smithson, H. E., & Mollon, J. D. (2004). Is the S-opponent chromatic sub-system sluggish? *Vision Research*, *44*(25), 2919–2929. https://doi.org/10.1016/j.visres.2004.06.022

Smithson, H. E., Sumner, P., & Mollon, J. D. (2003). How to find a tritan line? In J. D. Mollon, J. Pokorny, & K. Knoblauch (Eds.), *Normal and defective colour vision* (pp. 279–287). Oxford University Press.

Spering, M., Kerzel, D., Braun, D. I., Hawken, M. J., & Gegenfurtner, K. R. (2005). Effects of contrast on smooth pursuit eye movements. *Journal of Vision*, *5*, 455–465.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*, 578–585.

Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, *10*(1016/s0042-6989(00)00021-3), 10814758.

Thibos, L. N., Ye, M., Zhang, X., & Bradley, A. (1992). The chromatic eye: A new reduced-eye model of ocular chromatic aberration in humans. *Appl Opt*, *31*, 3594–3600.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, *22*, 377–380.

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

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Exp Brain Res*, *41*, 414–419.

Tomko, G. J., & Crapper, D. R. (1974). Neuronal variability: Non-stationary responses to identical visual stimuli. *Brain Res*, *79*, 405–418.

Traer, J., & McDermott, J. H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences*, *113:E7856–E7865*.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Trincker, D. (1953). Brightness-darkness adaptation and special vision. I. Phenomenology of the Pulfrich's effect with reference to asymmetry-phenomenon. *Pflugers Archiv Fur Die Gesamte Physiologie Des Menschen Und Der Tiere*, *257*(1), 48–69. https://doi.org/10.1007/BF00363411

Turano, K. A., & Heidenreich, S. M. (1999). Eye movements affect the perceived speed of visual motion. *Vision Research*, *39*, 1177–1187.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.

Ullman, S. (1979). *The Interpretation of Visual Motion*. MIT U. Press.

Vassilev, A., Mihaylova, M., & Bonnet, C. (2002). On the delay in processing high spatial frequency visual information: Reaction time and VEP latency study of the effect of local intensity of stimulation. *Vision Research*, *42*(7), 851–864.

Watson, A., & Ahumada, A. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America. A, Optics and Image Science*, *2*, 322–341. https://doi.org/10.1364/JOSAA.2.000322.

Weale, R. A. (1954). Theory of the Pulfrich effect. *Ophthalmologica. Journal International D'ophtalmologie. International Journal of Ophthalmology. Zeitschrift Fur Augenheilkunde*, *128*(6), 380–388. https://doi.org/10.1159/000302399

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. *Nature Publishing Group*, *18*, 1509–1517.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*, 598–604.

Williams, D. R. (1985). Visibility of interference fringes near the resolution limit. *Journal of the Optical Society of America*, 2, 1087–1093.

Wilson, J. A., & Anstis, S. M. (1969). Visual delay as a function of luminance. *The American Journal of Psychology*, *82*(3), 350–358.

Wyszecki, G., & Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulas*. John Wiley & Sons.

Yang, Z., & Purves, D. (2003). Image/source statistics of surfaces in natural scenes. *Network (Bristol, England*, *14*(3), 371–390.

Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., & Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nature Publishing Group*, *20*, 1285–1292.

Ziemba, C. M., Freeman, J., Movshon, J. A., & Simoncelli, E. P. (2016). Selectivity and tolerance for visual texture in macaque V2. *Proceedings of the National Academy of Sciences*, *113:E3140–E3149*.