COMPUTATIONAL PRINCIPLES UNDERLYING CONTEXTUAL MODULATIONS IN VISUAL PERCEPTION

Jiang Mao

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

2023

Supervisor of Dissertation

Alan A. Stocker, Associate Professor of Psychology

Graduate Group Chairperson

Russell Epstein, Professor of Psychology

Dissertation Committee

Geoffrey K. Aguirre, Professor of Neurology
Konrad Kording, PIK University Professor of Neuroscience

COMPUTATIONAL PRINCIPLES UNDERLYING CONTEXTUAL MODULATIONS IN VISUAL PERCEPTION

ACKNOWLEDGEMENT

First, I would like to thank my advisor, Alan Stocker, for his support and guidance throughout my journey in graduate school. His insightful advice and unwavering dedication to scientific rigor have been invaluable in shaping my research skills and scientific mindset. I would also like to thank my committee, Geoff Aguirre and Konrad Kording, for their invaluable suggestions. Their expertise and insights have greatly enriched my work.

I am thankful to other faculty members from whom I have learned a lot: David Brainard, Johannes Burge, Nicole Rust, Josh Gold, and Vijay Balasubramanian. I would also like to thank my undergraduate advisors, Jian Li and Zili Liu, who inspired my to choose this path of research.

I would like to thank members of the Stocker lab for sharing the journey with me. Long Luu welcomed me into the lab when I first started. Cheng Qiu was extremely helpful with her knowledge and expertise in research as well as experience in life, and she is also a great friend to hang out with. Ling-Qi Zhang has always been a great inspiration to me for his passion and insight in science and a lot of things. I've also had many interesting discussion with Long Ni, and talking to Mengting Fang has always been a delight.

I want to thank my fellow CNI members for creating a vibrant and supportive environment: Michael Barnette, Ben Chin, Takahiro Doi, Seha Kim, Noam Roth, Kyra Schapiro, Dave White, Catrina Hacker, Kara McGaughey, Anthony LoPrete, Carlos Rodriguez, Callista Dyer, Simon Bohn, Emily Meyer, Semin Oh, Briana Haggerty, Daniel Herrera. I deeply cherish the happy time we shared and enjoy all the random conversations in the office, although I only quietly lurked for most of the time. I would also like to thank my other friends at Penn: Tima Zeng, Wanling Zou, Zhengang Lu, Yihui Zhang, and Yucheng Kang. A special thanks goes to Lingxi Lu, an old friend who has been around for the longest time, for the many evenings when we talked about anything and everything.

Lastly, I would like to thank my family: my parents, Rong Zhu and Hengqing Mao, for their support throughout the years, and my fiance, Xinpeng Wang, for the company and encouragement through

good and bad times.

ABSTRACT

COMPUTATIONAL PRINCIPLES UNDERLYING CONTEXTUAL MODULATIONS IN VISUAL PERCEPTION

Jiang Mao

Alan A. Stocker

Perception is constantly modulated by context, including temporal context, spatial context, structural context, etc. The same stimulus may be perceived differently under different contexts. Studying the contextual modulation of perception may provides deep insight into the computational principles of the sensory system. The present thesis sheds light on the computational principles underlying these contextual modulation. In Chapter 2, I test the efficient coding principle during sensory adaptation (temporal context). I measure the orientation discrimination threshold in human observers and extract the difference in coding accuracy under different adaptation states. By comparing the extracted coding accuracy with the image statistics analyzed from the retinal input of freely behaving human subjects under natural conditions and the Fisher information in a recurrent neural network trained on natural scene videos to predict the next frame while performing the same task as human subjects, I provide evidence for the efficient coding explanation of the adaptation effect, namely, adaptation to the recent history of sensory input establishes efficient sensory representations for the next expected sensory input. In Chapter 3, I present results on the structural context in visual orientation perception. I propose a holistic matching model that assumes perception is a holistic inference process that simultaneously operates at all levels of the representational hierarchy. Validation against multiple existing psychophysical datasets and data from a new psychophysical experiment demonstrates that compared to previous models, our model provides a quantitatively accurate and detailed description of subjects' behavior, which includes categorical contextual effects that previous models have failed to even qualitatively account for. I also show that the model generalizes to other features and thus offers a universal explanation for categorical contextual modulation in low-level sensory perception. Together, this thesis advances

our understanding of visual perception under contextual modulations and provides insight into the underlying computational principles from a normative perspective on both the encoding and decoding process of perception.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

BACKGROUND

Perception of s stimulus feature never occurs in isolation. It is always influenced by the context within which the stimulus presents itself, including temporal context (Gibson and Radner, 1937; Muller et al., 1999), spatial context (O'Toole and Wenderoth, 1977), structural context (Young et al., 1987; Gershman et al., 2016; Liberman et al., 1957; Bae et al., 2015), the task sequence (Jazayeri and Movshon, 2007; Bronfman et al., 2015), etc. Many famous visual illusions take advantage of the fact that our visual system relies on context when perceiving a target stimulus feature. For example, a dark square brightly illuminated and a white square in the shadow can have the same brightness, but people inevitably see them as different colors due to the spatial context of the average brightness around them. This characteristic of our visual system may seem suboptimal in these situations. However, because there are intrinsic regularities in the natural world, incorporating context as a way of utilizing these regularities can improve our perception, whereas these illusions either require people to disregard the context or are artificially designed to break the natural regularities. Early studies in visual perception tend to isolate perception from the context by using a uniform neutral background when presenting the stimulus or by averaging out the context by randomizing the trial sequence. This approach was initially suitable for studying the "basic" perception because taking context into consideration can be complicated. More recently, however, research has been focused on contextual modulations to start to understand the adaptive computation involved in contextual perception. For example, task sequence can lead to confirmation bias where people's percept is biased to be more consistent with their own previous judgment (Jazayeri and Movshon, 2007), and Luu and Stocker (2018) showed that confirmation bias can be accounted for by a self-consistent Bayesian observer model with a prior consistent with the previous categorical judgment.

In this dissertation, I studied the influence of context on visual perception, mainly focusing on the temporal context and the structural context. Temporal context refers to the previous sensory input within a period of time prior to the present. The temporal contextual effect is commonly referred to

as adaptation, a series of changes in neural responses and behaviors following prolonged exposure to the same stimulus. Adaptation has been an important probe into the sensory representation and the dynamic coding principles of the visual system. However, many previous experiments were done with limited conditions and some questions remain unclear. In this dissertation, I ran a psychophysical experiment on visual orientation adaptation with the test stimulus spanning the entire orientation range and with a properly controlled null-adaptation condition, extracted the coding accuracy with an information theoretic data analysis, and tested the efficient coding hypothesis of adaptation using natural scene statistics and a recurrent neural network.

Structural context refers to the hierarchical structure a stimulus or a feature is in. It can be the grouping of multiple objects into a whole or the categorical identity of a feature. In particular, the categorical structure has been shown to influence many different perceptual features and has been studied extensively. However, there has not been a normative computational model of the structural contextual effect that both captures the hierarchical nature of the structural context and successfully explains the categorical biases. I proposed a holistic matching model that assumes that perception is a holistic inference process that simultaneously operates at all levels of the representational hierarchy and showed that it accurately accounts for the distribution of response from previous experiments, and successfully predicts biases that seem counter-intuitive under non-categorical estimation models.

1.1. Temporal Context

When you stare at a video of a moving waterfall for 30 seconds and then the video pauses, the waterfall will look like it is moving upward (Crane, 1988). This famous waterfall illusion demonstrates that our perception is heavily influenced by the temporal context, which is the sensory input we receive prior to the current stimulus. Both the neural representation of the sensory information and the perceptual behavior change with the temporal context. These changes are referred to as adaptation.

On the behavioral level, prolonged exposure to visual stimuli leads to systematic changes in the detection, discrimination and estimation of subsequent stimuli. These changes have been found in various domains of visual perception, including contrast (Hammett et al., 1994; Webster and Miya-

hara, 1997), orientation (Gibson and Radner, 1937; Regan and Beverley, 1985), direction (Phinney et al., 1997; Schrater and Simoncelli, 1998), face perception (Webster et al., 2004), etc. After adaptation, detection is selectively impaired for subsequent stimuli similar to the adaptor due to reduced neural responsivity to the adaptor (Blakemore and Campbell, 1969; Blakemore and Nachmias, 1971; Regan and Beverley, 1983). Discrimination, however, improves for stimuli similar to the adaptor, but is impaired for stimuli more different from the adaptor (Regan and Beverley, 1983, 1985; Phinney et al., 1997; Clifford et al., 2001; Stocker and Simoncelli, 2004). For estimation, adaptation leads to a repulsive bias near the adaptor, and for circular variable like orientation and motion direction, a smaller, narrower attractive bias far from the adaptor (Gibson and Radner, 1937; Mitchell and Muir, 1976; Clifford et al., 2000; Schrater and Simoncelli, 1998; Blakemore et al., 1970; Stocker and Simoncelli, 2009; Webster et al., 2004).

On the neural level, adaptation leads to reduced neural responsivity and changes in the shapes of the tuning curves, noise properties, and correlation between neurons (Kohn, 2007), and these changes could be different depending on the feature and the adaptation time. For adaptation to orientation, the amplitudes of neurons whose preferred orientations are close to the adaptor are reduced after adaptation, and their preferred orientations also shift away from the adaptor (Muller et al., 1999; Dragoi et al., 2000; Patterson et al., 2013). The bandwidths of tuning curves that peak around the adaptor increase after adaptation (Dragoi et al., 2000), but some studies also show that tuning curves sharpen around the adaptor (Muller et al., 1999; Patterson et al., 2013). Neurons with preferred orientations close to the adaptor have also been found to be less noisy after adaptation (Muller et al., 1999). From simultaneous recordings of multiple units, Gutnisky and Dragoi (2008) found that signal correlation, noise correlation, and the variability of correlation decrease after adaptation, and Fisher information increases both near and opposite of the adaptor. Besides adapting to a single feature value, studies have shown that the sensory system also adapts to stimulus statistics such as the mean or range of variations of the stimulus. Photoreceptors adjust their dynamic ranges according to the mean light intensity (Normann and Perlman, 1979). Retinal ganglion cells change their sensitivity to match the contrast in the visual environment (Baccus and Meister, 2002). Motion-sensitive H1 neurons in flies adjust their dynamic ranges to match the range

of variation in the stimulus (Brenner et al., 2000; Fairhall et al., 2001).

These changes in perceptual behavior and neural representation with temporal context are not arbitrary. Many different coding principles have been proposed to explain the adaptation effect. One of the most prominent hypotheses is efficient coding, which argues that the system maximizes the amount of information it encodes about the sensory input under certain biological constraints (Barlow et al., 1961; Wainwright, 1999; Brenner et al., 2000; Sharpee et al., 2014). An idea derived from efficient coding is histogram equalization, which states that the probability of different neural response level are the same (Laughlin, 1981; Fairhall et al., 2001). However, it can be difficult to rigorously test the efficient coding hypothesis, because it requires a measurement of the neural code that is comprehensive enough to calculate the information in the neural system. Redundancy reduction is also a coding principle commonly considered in adaptation, where each neuron should fire as independent of each other as possible (Barlow et al., 1961; Laughlin, 1989; Atick and Redlich, 1990). Reducing redundancy can achieve efficient coding in the small noise regime. However, when the noise is large, redundancy might be preferable in terms of conveying the most information (Atick and Redlich, 1990; Kastner et al., 2015). Another coding principle related to redundancy reduction is predictive coding, which hypothesizes that instead of representing the input directly, the neural system makes prediction of the input and represents the prediction error between the input and the prediction (Rao and Ballard, 1999; Aitchison and Lengyel, 2017). Instead of focusing on an accurate representation, Wei et al. (2015); Młynarski and Hermundstad (2018, 2021) proposed adaptive coding schemes that are optimized for inference, or decoding, to account for adaptation in a dynamic environment. Besides the above normative explanations, another common mechanistic explanation of adaptation is fatigue, an intrinsic suppression of neuronal activities after being active for a prolonged period of time (Maffei et al., 1973; Albrecht et al., 1984; Ohzawa et al., 1985). Vinken et al. (2020) showed that complex neural and behavioral adaptation phenomena can emerge from a cascade of intrinsic suppression through a feedforward neural network by incorporating exponentially decaying fatigue into every unit of AlexNet, a deep neural network trained to recognize objects.

Efficient coding being the most prominent hypothesis among the coding principles of adaptation,

rigorously testing it in terms of information is an important question that attracts much attention of the field of neuroscience (Schwartz et al., 2007; Gutnisky and Dragoi, 2008; Seriès et al., 2009). However, as mentioned above, it can be difficult to calculate the information conveyed by the neural code based on the neural measurement. One possible method is the simultaneous recording of neural populations. Compared to measuring tuning curves of individual neurons, this method can also account for the noise in neural activities and the correlation among neurons, but it is still limited by the number of neurons that can be recorded simultaneously. Another possible method is by using fMRI and decoding the stimulus information from trial-by-trial BOLD signal (Van Bergen et al., 2015). Still, these neural measurements suffer from the ambiguity of the causal role of a particular neural population in representing the stimulus. However, one can bypass the neural measurement and use the behavioral measurement of discrimination threshold as a proxy to probe the encoding principles of adaptation. The more information about the stimulus in the sensory system, the better the coding accuracy, hence the lower the discrimination threshold (Fechner, 1966). It can be analytically proven that the discrimination threshold is inversely proportional to the Fisher information in the neural representation when the noise is small (Seriès et al., 2009; Wei and Stocker, 2017). By behaviorally measuring the discrimination threshold, one can compute the Fisher information and directly probe the efficient coding hypothesis of adaptation.

In Chapter 2 of this dissertation, I will present psychophysical results of discrimination threshold before and after adapting to visual orientations and the Fisher information extracted from the measured discrimination threshold. Previous experiments measuring discrimination thresholds after adaptation either did not measure the entire stimulus range, or measure the same test orientation while adapting to different adaptor orientation, neither of which allows the extraction of Fisher information. I was able to derive a universal adaptation kernel that can describe the encoding changes after adapting to any arbitrary orientation. Furthermore, I tested the efficient coding hypothesis of adaptation using natural scene statistics and a recurrent neural network inspired by predictive coding optimized to do video prediction.

1.2. Structural Context

The visual scene is not a plain collection of individual features. Features are organized by spatial, temporal and categorical structures. Some features or objects belong to certain groups or categories; some tend to be arranged in a certain way and fall into a relatively fixed scheme or template. The representation of these structural contexts aids and influences the perception of individual features in multiple ways.

For example, faces are highly structured visual stimuli that people have extensive exposure to since birth. Many studies have shown that people have holistic processing of faces and the perception of the whole face as a structural context influences the perception of parts. For example, the composite face illusion where people fail to identify identical top halves of a face when the bottom halves are different implies that the features of a face cannot be perceived in isolation of the face as a whole (Young et al., 1987; Hole, 1994; Rossion, 2013). The face inversion effect shows that presenting a face upside-down impairs people's perception of metric distances between facial features (Rhodes et al., 1993; Rossion, 2008), reflecting a disruption of the perception of the configuration or structure. Interestingly, face inversion can also reduce or completely break the composite face illusion (Hole, 1994; Goffaux and Rossion, 2006), suggesting an intact percept of features when the structural context is disrupted. These effects show that the face, as a prevalent structural context, exerts an substantial impact on the perception of its features.

Another example of structural context is the structure of motion. It carries important information about the environment that aids behaviors such as navigation, prediction, tracking, and pursuit. Imagine a person walking on a moving train: the train and the person together move towards one direction at a high speed; on top of that, the person moves relative to the train; and the limbs of the person move relative to the body in a stereotypical biological motion. Representing the motion of each component in the hierarchy of motion structure can be much more informative and easier than representing the absolute motion of each object in isolation. Actually, people are so good at utilizing the structure of motion to the point where it can be difficult to unsee the structure, and the percept of individual motion can be greatly influenced by the motion structure. A classic demonstration

is the Duncker wheel: two dots move in a way that resembles the hub and one dot on the rim of a rolling wheel (Duncker, 1938). Although no additional structural information is given, humans perceive a rolling wheel instead of two individual moving dots. When asked about the peripheral dot, some people even report a backward motion when it is directly below the hub, although they could correctly perceive the non-backing motion when the hub is not presented. While the Gestalt theory of common fate serves as a rudimentary explanation of grouping by motion: objects that move together are grouped together, some recent studies proposed more detailed normative computational models to solve the complex problem of hierarchical motion structure perception, provided normative and quantitative explanations of human motion illusions, and made testable predictions about motion perception that were later verified (Gershman et al., 2016; Bill et al., 2020b; Yang et al., 2021; Bill et al., 2022).

A more simple form of structural context is given by the categorical membership of a feature, a stimulus or an object. The effect of categories has been shown prominently in the perception of speech sounds (Liberman et al., 1957, 1961), colors (Davidoff et al., 1999; Winawer et al., 2007; Bae et al., 2015; Hardman et al., 2017), faces (Etcoff and Magee, 1992; Calder et al., 1996; Beale and Keil, 1995), even artificial categories learned within the process of the experiment (Goldstone, 1994; Goldstone et al., 2001). The categorical effect behaviorally manifests itself mainly in two ways. First, discrimination is better around the categorical boundaries than near the centers of the categories (Liberman et al., 1957; Kuhl, 1991; Winawer et al., 2007; Etcoff and Magee, 1992). The heterogeneity in discrimination accuracy is commonly viewed as a defining characteristics of categorical perception. Second, the estimation or reproduction of a stimulus is biased away from the categorical boundaries and towards category centers (Huttenlocher et al., 1991; Bae et al., 2015).

Many models have been proposed to explain the perceptual biases related to the category. One prominent type of models is the Bayesian inference model that assumes a hierarchical generative process and that each category has a prior that peaks at the center or the prototype of the category (Feldman et al., 2009; Kronrod et al., 2016; Landy et al., 2017). By marginalizing over all the categories, these models are essentially equivalent to a non-hierarchical Bayesian inference model

with a heterogeneous prior determined by the sum of stimulus priors given each category weighted by the probability of the category. Another popular group of models assumes that the observer infers the category of the stimulus first, then makes inference about the feature given the inferred category, therefore deviates from the normative formulation (Bae et al., 2015; Stocker and Simoncelli, 2007; Luu and Stocker, 2018). The prototype model can be seen as a simplified version of the Bayesian model, where category prototypes exert a pull on nearby stimulus, resulting in an attractive bias towards them (Grieser and Kuhl, 1989; Hardman et al., 2017). Besides the Bayesian models, other computational models such as the rate-distortion theory (RDT) based model have also been proposed for the categorical effect Sims et al. (2016).

Several studies have suggested that the perception of visual orientation is affected by a distinction of the cardinal/oblique categories (Rosielle and Cooper, 2001; Wakita, 2004). Visual search for one orientation among heterogeneous distractors is more efficient when the target and the distractors belong to different orientation categories (Wolfe et al., 1992). Differentiating oblique from parallel or perpendicular relative orientations in objects is easier than differentiating between angles within the same category (Rosielle and Cooper, 2001). Monkeys trained to discriminate between two orientations are able to generalize across orientations within the cardinal/oblique categories but not across categories (Wakita, 2004). The better discrimination at cardinal orientations compared to oblique orientations seems to be consistent with categorical perception separating clockwise and counterclockwise relative to vertical. However, the current, most sophisticated computational model of orientation perception has not taken categories into account (Wei and Stocker, 2015; Taylor and Bays, 2018). Orientation being one of the most tested features in visual perception, the responses in classic orientation estimation experiments has not been thoroughly and quantitatively explained by previous computational models, and some data cannot be even qualitatively accounted for by estimation models in general. Category, as a simple form of structural context, might play a role.

In Chapter 3 of this dissertation, I proposed a holistic matching model that incorporates the effect of categorical structure into orientation perception. The holistic matching model assumes that orientation estimation is a matching process that operates on both the feature level and the categorical

level. Just like the holistic processing of faces can distort the percept of individual features, or the representation of motion structure can bias the percept of the composite motion of a single object, the holistic representation across the entire hierarchy of the feature and its category can lead to biases in estimating the low-level feature. I will demonstrate the ability of the holistic matching model to account for the full distribution of the response from existing psychophysical data, and present new psychophysical experiment results consistent with the predictions of the model which cannot be qualitatively explained by estimation models.

CHAPTER 2

PERCEPTUAL ADAPTATION LEADS TO CHANGES IN ENCODING ACCURACY OPTIMIZED FOR FUTURE SENSORY INPUT

2.1. Abstract

Our visual system continually adapts to its sensory environment. As a result, both the encoding accuracy of sensory information and perceptual behavior change according to the recent history of sensory input. Here we tested the hypothesis that adaptation to the recent history of sensory input optimally prepares the perceptual system for the future, i.e. establishes efficient sensory representations for the next expected sensory input. We first quantitatively characterized the precise changes in neural encoding accuracy induced by adaptation by measuring discrimination thresholds for visual orientation after adaptation in a psychophysical experiment. Using an information theoretic data analysis, we then extracted the characteristics form of how encoding accuracy changed due to adaptation as a function of stimulus orientation relative to the adaptor orientation. We found that encoding accuracy was substantially higher at the adaptor orientation compared to the control condition. We then asked whether this increase in accuracy at the adaptor orientation is predicted by the natural visual input statistics. In the retinal input of freely behavior humans subjects under natural conditions, we found that after a relatively stable visual input over a short time-window, the distribution of local orientations at fixation in the next frame are peaked at the mean orientation over the time-window. We further tested the hypothesis with PredNet, a recurrent neural network trained on natural scene videos to predict the next frame. We input image sequences similar to those used in the human adaptation experiment and found that PredNet exhibited the same increase in encoding accuracy at the adaptor orientation as observed in human subjects. Together, our results suggest that adaptation induced changes in encoding accuracy and perceptual behavior reflect the visual systems attempt to be best possibly prepared for future sensory input.

2.2. Introduction

Adaptation has been shown to influence stimulus encoding on the single-neuron level (Dragoi et al., 2000, 2002; Patterson et al., 2013), the population level (Gutnisky and Dragoi, 2008) and the perceptual level (Regan and Beverley, 1985; Clifford et al., 2001; Mitchell and Muir, 1976). Adaptation has been mainly thought of as an efficient coding mechanism that adjusts the operational regime of the visual system to optimally represent and process sensory information based on the spatiotemporal statistical regularities of the recent sensory input (Barlow et al., 1961; Wainwright, 1999; Brenner et al., 2000; Sharpee et al., 2014). Here we aim to test the efficient coding hypothesis, namely, adaptation establishes efficient sensory representations for the next expected sensory input.

First, we need to characterize the change in sensory representation after adaptation. Fisher information represents the coding accuracy of the sensory representation, or the amount of coding resources allocated to each stimulus. With the assumption of uniformly loose Cramer-Rao bound, Fisher information can be computed from discrimination threshold (Seriès et al., 2009). So in order to find out the change in encoding, we can look at the change in discrimination threshold after adaptation.

Many previous studies have measured the adaptation effect on discrimination threshold, but none of them measured the impact of adapting to a specific orientation to the entire range of orientations, which permits the characterization of the distribution of Fisher information among different orientations. Some have shown that adapting to a single stimulus will decrease the discrimination threshold at the adaptor, and increase the threshold for stimulus near but different from the adaptor (Regan and Beverley, 1983, 1985; Clifford et al., 2001; Phinney et al., 1997). For circular variables like orientation, some studies found that the discrimination threshold also decreases when the test is opposite of the adaptor (Clifford et al., 2001; Dragoi et al., 2002), but this result is not conclusive (Westheimer and Gee, 2002; Clifford et al., 2003). However, many studies used different adaptors with the same test stimulus instead of one adaptors with different test stimulus (Clifford et al., 2001); some studies used one adaptor and different test stimuli but did not test the entire stimulus range (Regan and Beverley, 1985). In order to fully characterize the change in sensory representa-

tion induced by adaptation, one should test the entire stimulus range under the same adaptation state.

We then asked whether the observed change in sensory representation is efficient for the next expected sensory stimulus. Coding is efficient when the Fisher information is matched to the stimulus distribution (Wei and Stocker, 2015). Due to the continuous nature of natural scene, the next stimulus is likely to be similar to recent history (Dragoi et al., 2002; Felsen et al., 2005; van Bergen and Jehee, 2019). We analyzed the the spatiotemporal stimulus distributions in natural scene in retinal coordinate using natural videos recorded simultaneously with the eye movement of the observer. If the orientation in the next frame is more likely to be similar to the orientation in recent history at the same retinal location, increased coding accuracy at the adaptor will prove to be efficient for future sensory input in natural viewing conditions.

The efficient coding hypothesis can be further tested by investigating artificial neural network trained on natural scene images and videos. Previous studies have shown that deep neural network is able to pick up natural scene statistics and encode certain stimulus feature efficiently (Benjamin et al., 2019). In particular, PredNet is a recurrent neural network that predicts the next frame in a video (Lotter et al., 2016). PredNet was inspired by the concept of "predictive coding" in neuroscience, with each layer making local predictions of incoming stimuli and forwarding the prediction errors to the subsequent layer. Previous studies have found that PredNet trained with natural scene videos perceives similar illusory motion as human (Watanabe et al., 2018; Kobayashi et al., 2022). If PredNet also shows similar adaptation effect as human, because there is no local adaptation mechanism (e.g. neural gain change), it suggests that the adaptation induced changes in encoding may be optimal for the perception of incoming stimulus with regard to the spatiotemporal regularities of natural scene.

In this study, we investigated the change in the distribution of Fisher information after adaptation. We hypothesize that adaptation reallocates the coding resources, or in other words, changes the distribution of Fisher information, but does not change the total amount of coding resources, and the reallocation is the same for different adaptors. We tested the discrimination threshold of visual

orientation after adapting the participants with stimuli containing certain orientation components. The adaptors were either oblique adaptors containing only one oblique orientation, or null adaptors containing all orientations as a control condition. We found that compared to the control condition, adapting to the oblique adaptor decreases the discrimination threshold at the adaptor orientation and the orthogonal orientation, and increases the discrimination threshold near the adaptor orientation. By fitting a reallocation model to the data, we found that the total Fisher information is the same under different adaptation states, and that the reallocation of Fisher information can be described by a single adaptation kernel for adaptors different orientations, confirming our hypothesis.

Our second hypothesis is that the changes in encoding induced by adaptation is efficient for the representation of the next stimulus. We analyzed spatiotemporal orientation distributions in retinal input of freely behavior humans subjects under natural conditions. We found that after a relatively stable visual input over a short time-window, the distribution of local orientations at fixation in the next frame are peaked at the mean orientation over the time-window. Thus an increase in encoding accuracy at the adaptor represents an efficient encoding of the next stimulus under these natural images statistics. We further tested the hypothesis with the PredNet trained on natural scene videos. We computed the Fisher information of image orientation in the network after being presented with short sequence of images with a single orientation (adaptor). PredNet exhibited the same increase in encoding accuracy at the adaptor orientation as observed in human subjects, suggesting that this increase is optimal for the representation of future sensory input given the spatiotemporal image statistics in natural scene.

2.3. Results

2.3.1. Reallocation Model

The representational resource in the perceptual system is limited, and is usually not uniformly distributed among stimuli. We propose that adaptation reallocates the coding resource among the stimulus relative to the adaptor without changing the total amount of representational resource. More specifically, we quantify the representational resource or encoding accuracy with Fisher in-

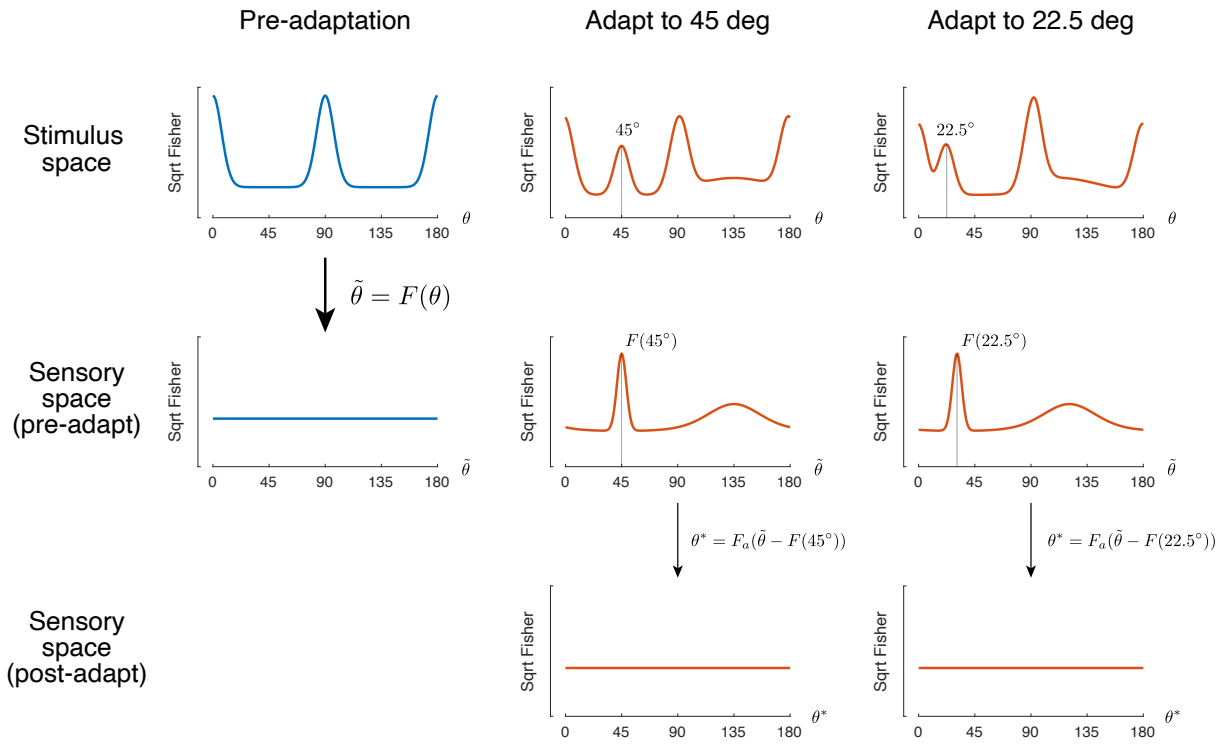

Figure 2.1: Reallocation Model. Before adaptation (left), the Fisher information has a certain distribution in the stimulus, which can be transformed to a uniform distribution in a sensory space. After adaptation (middle and right), the Fisher information is re-distributed according to an adaptation kernel, which takes the same shape for different adaptors but is centered at the respective orientation.

formation. Before adaptation, Fisher information is higher around cardinal orientations, leading to better discriminability at cardinal compared to oblique orientations (Caelli et al., 1983) (Fig. 2.1). By applying a transformation $\tilde{\theta} = F(\theta)$ where F is the cumulative distribution of the square root of Fisher information, we can transform to a sensory space where the distribution of Fisher information is uniform (Wei and Stocker, 2015). After adapting to a single orientation, the Fisher information is re-distributed according to an adaptation kernel. We assume that the adaptation kernel, when formulated in the pre-adaptation sensory space, takes the same shape for different adaptor orientations but is only shifted according to the adaptor. Now, the pre-adaptation sensory space is not a uniform space any more. Again, we can transform to a post-adaptation sensory space where the distribution of Fisher information is uniform by going through a transformation F_a that reflects the adaptation kernel. We assume homogeneous von Mises sensory noise in the uniform sensory space. In a discrimination task, a noisy sensory measurement is made for each stimulus according to the sensory noise, and the observer makes an optimal decision by comparing the percept of the test stimulus and the comparison stimulus (see Methods). The resulting discrimination threshold reflects the distribution of Fisher information in the sensory representation (Seriès et al., 2009).

2.3.2. Experiment

We measured observers' discrimination threshold to orientation under different adaptation states with a 2AFC discrimination task (Fig 2.2a). At the beginning of each block, participants viewed the adaptor for one minute. Then in each trial, there was a top-up adaptation of 5s before the test and comparison stimuli were presented. Participants responded which stimulus was more clockwise (or counterclockwise). We tested two types of adaptors. The oblique adaptor is white noise filtered with spatial frequency band and a narrow orientation band. The control adaptor is white noise filtered with the same spatial frequency band but contains all the orientations. We tested two oblique adaptors: 45 deg and 22.5 deg. We set the adaptor orientation to be oblique instead of cardinal to avoid possible ceiling effect because discrimination threshold is the lowest at cardinal without adaptation (Caelli et al., 1983) and is expected to lower if adapting there (Regan and Beverley, 1985; Clifford et al., 2001).

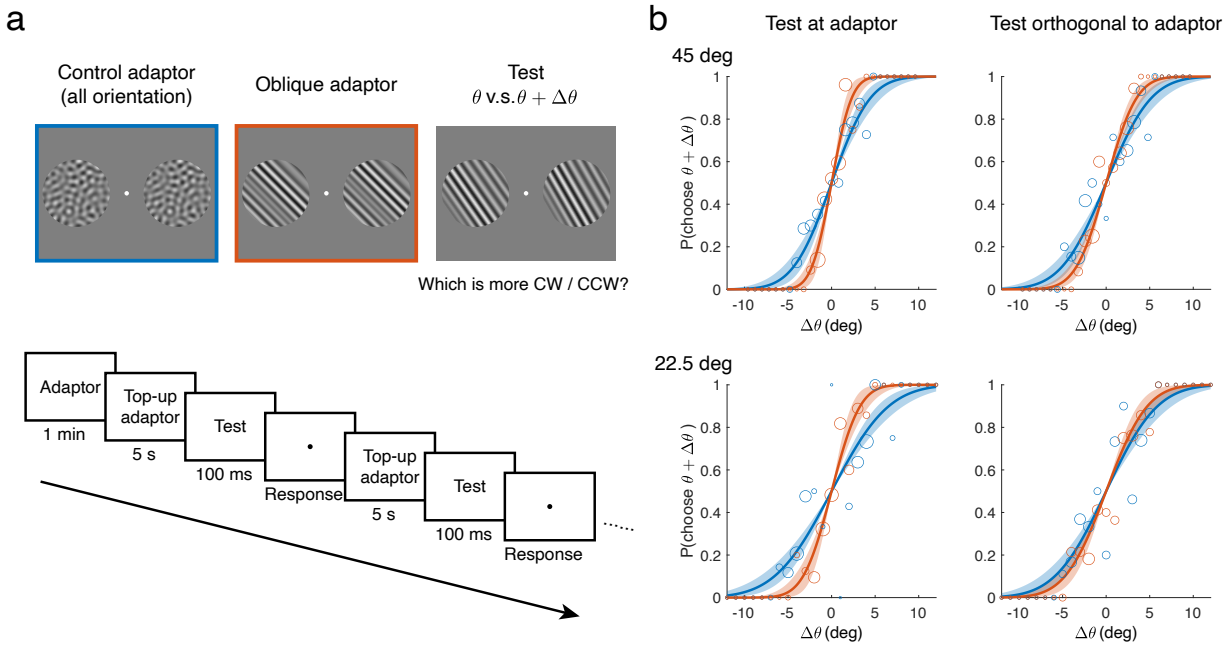

Figure 2.2: Experiment procedure and psychometric curves. (a) Experiment procedure. At the beginning of each block, there was an adaptation period of 1 minute. At the beginning of each trial, there was a top-up adaptation of 5s. Then participants viewed two oriented gratings and responded which one is more clockwise or counterclockwise. There were two types of adaptor: the control adaptor and oblique adaptor. (b) Data and fitted psychometric curves for one example subject. Compared to the control adaptor (blue), after adapting to the oblique adaptor (orange), the psychometric curve becomes steeper when the test is at the adaptor or orthogonal to the adaptor. The size of data points indicates the number of trials. Shaded areas represent 95% confidence intervals from 1000 bootstrap samples of the data.

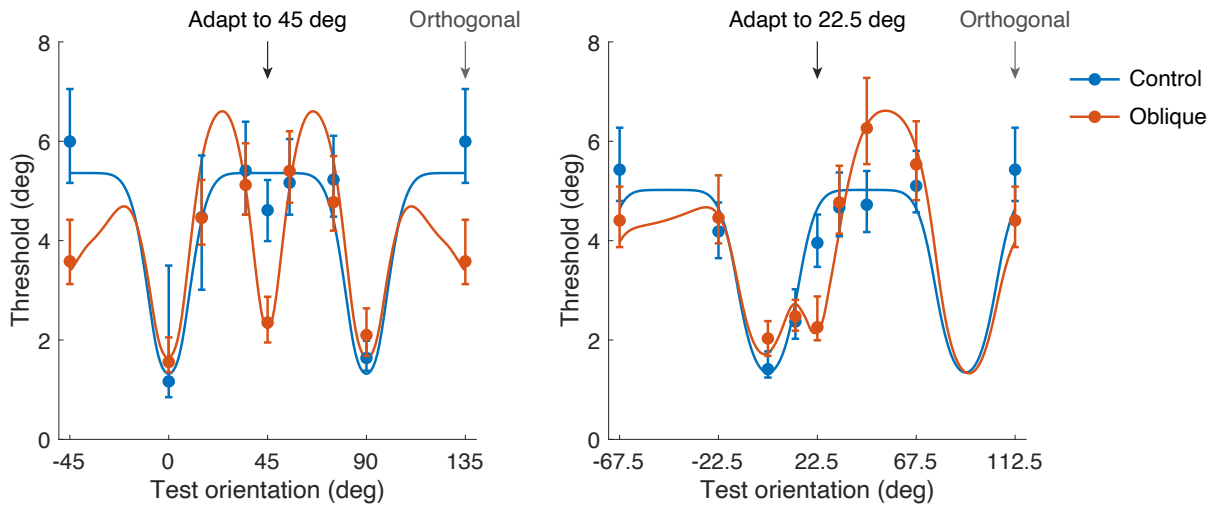

Figure 2.3: Discrimination threshold data and model fits averaged across subjects. In the control condition, the threshold is lower at cardinal orientations. In the oblique adaptor condition, the threshold decreased at the adaptor and orthogonal to the adaptor, and increased away from the adaptor. Error bars represent 95% confidence intervals from 1000 bootstrap samples of the data.

We fitted psychometric curves to the 2AFC experiment data (Fig. 2.2b), and extracted 75% discrimination thresholds (Fig. 2.3). In the control condition, discrimination threshold was lowest at cardinal orientation as shown in previous studies (Caelli et al., 1983). Compared to the control condition, after adapting to a single-orientation adaptor, the discrimination threshold decreased at the adaptor and increased slightly away from the adaptor, which is consistent with previous studies (Regan and Beverley, 1985; Clifford et al., 2001). The discrimination threshold at the orientation orthogonal to the adaptor also decreased after adaptation, which has not been consistently shown in previous studies (Clifford et al., 2001; Westheimer and Gee, 2002; Clifford et al., 2003; Dragoi et al., 2002). These results are consistent for both 45 deg and 22.5 deg adaptors and across subjects (Fig 2.4).

2.3.3. Model Fit and Comparison

As shown in the previous section, discriminability improves after adaptation both at the adaptor and orthogonal to the adaptor. So we assume that the adaptation kernel, which determines the ratio of Fisher information after and before adaptation in the pre-adaptation sensory space, has two peaks at and orthogonal to the adaptor respectively, and the same adaptation kernel applies

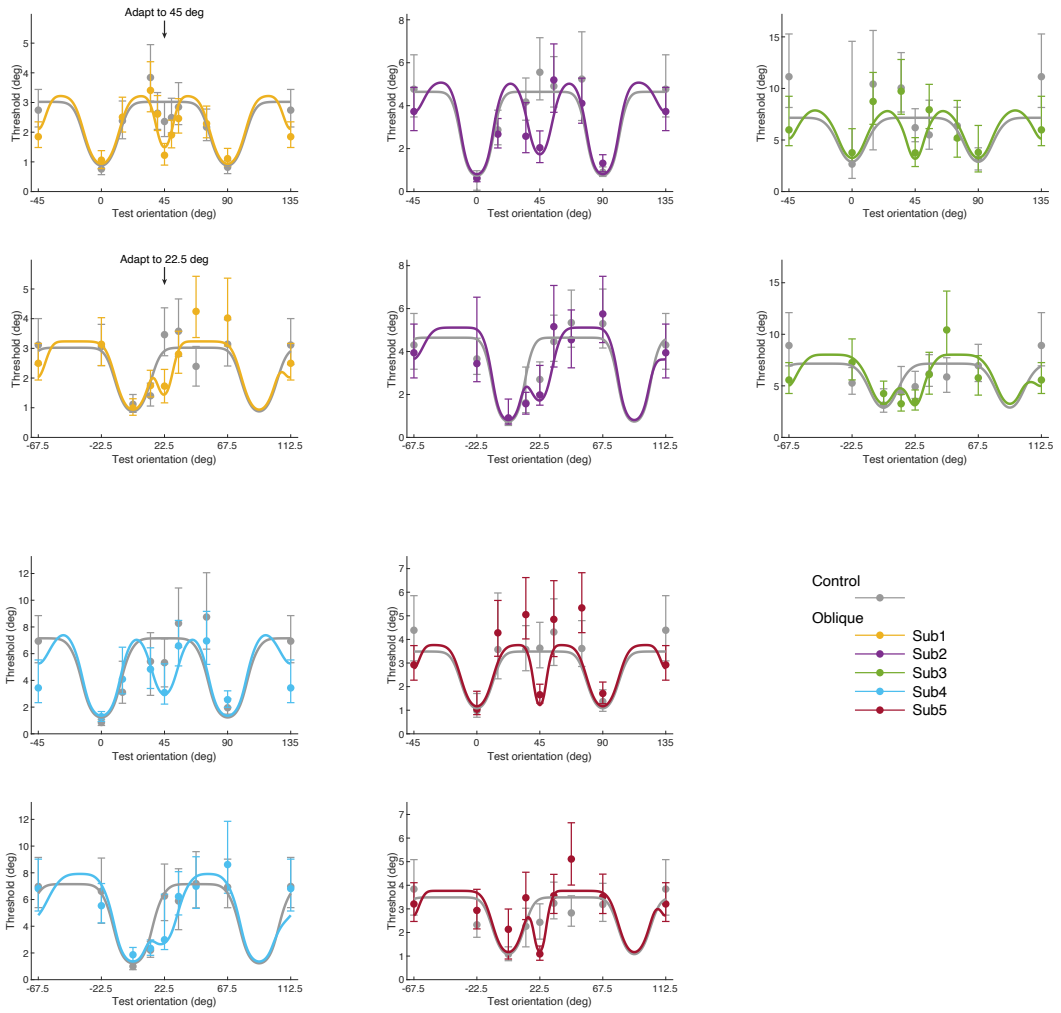

Figure 2.4: Discrimination threshold data and model fits for individual subjects. Most subjects show decreased threshold at and orthogonal to the adaptor, and increased threshold away from adaptor. Error bars represent 95% confidence intervals from 1000 bootstrap samples of the data.

to both 45 deg and 22.5 deg adaptors. We also assume that the total Fisher information does not change after adaptation, meaning the average sensory noise as expressed by the width of the sensory measurement distribution in the uniform sensory space remains constant. We fit the model individually to each subject. We first fit to the control adaptor condition, which determines the average sensory noise and the distribution of Fisher information before adaptation. Then we fit to the 45 deg and 22.5 deg adaptor conditions jointly to determine the adaptation kernel.

Figure 2.3 shows the mean discrimination threshold across subjects predicted by the model, and Figure 2.4 shows the fit to individual subjects. The model can fit not only the improvement of discrimination at the adaptor and orthogonal to the adaptor, but also the increased discrimination threshold at test orientations slightly different from the adaptors. We compare the current model (2-peak) with a model that assumes coding improvement only at the adaptor, which means the adaptation kernel only has one peak (1-peak), a two-peak model that allows the total Fisher information to change after adaptation (2-peak + Fisher), and a two-peak model that allows the adaptation kernel to be different for different adaptor orientations (2-peak + kernel)(Fig. 2.5). When penalized for the number of free parameters, the 2-peak model fits the data best for most participants (Fig. 2.5a). When the total Fisher information is allowed to change, the fitted total Fisher information under control and oblique adaptation is very similar, validating our hypothesis that the total representation resource is fixed and does not change with the adaptation state (Fig. 2.5b). When the adaptation kernels for different oblique adaptors are allowed to be different, the fitted kernels are similar to each other and to the kernel when constrained to be the same for different oblique adaptors (Fig. 2.5c), suggesting that the sensory system adapts to different adaptors in the same way, with the same adaptation kernel.

2.3.4. Natural Scene Statistics

We proposed that adaptation changes the encoding of stimulus due to efficient coding of the upcoming stimulus. Natural scene is continuous in time and space, so the next stimulus is bound to be similar to the previous stimulus (Dragoi et al., 2002; van Bergen and Jehee, 2019). If the stimulus is stable for a relatively long time, the next stimulus might be more likely to be the same,

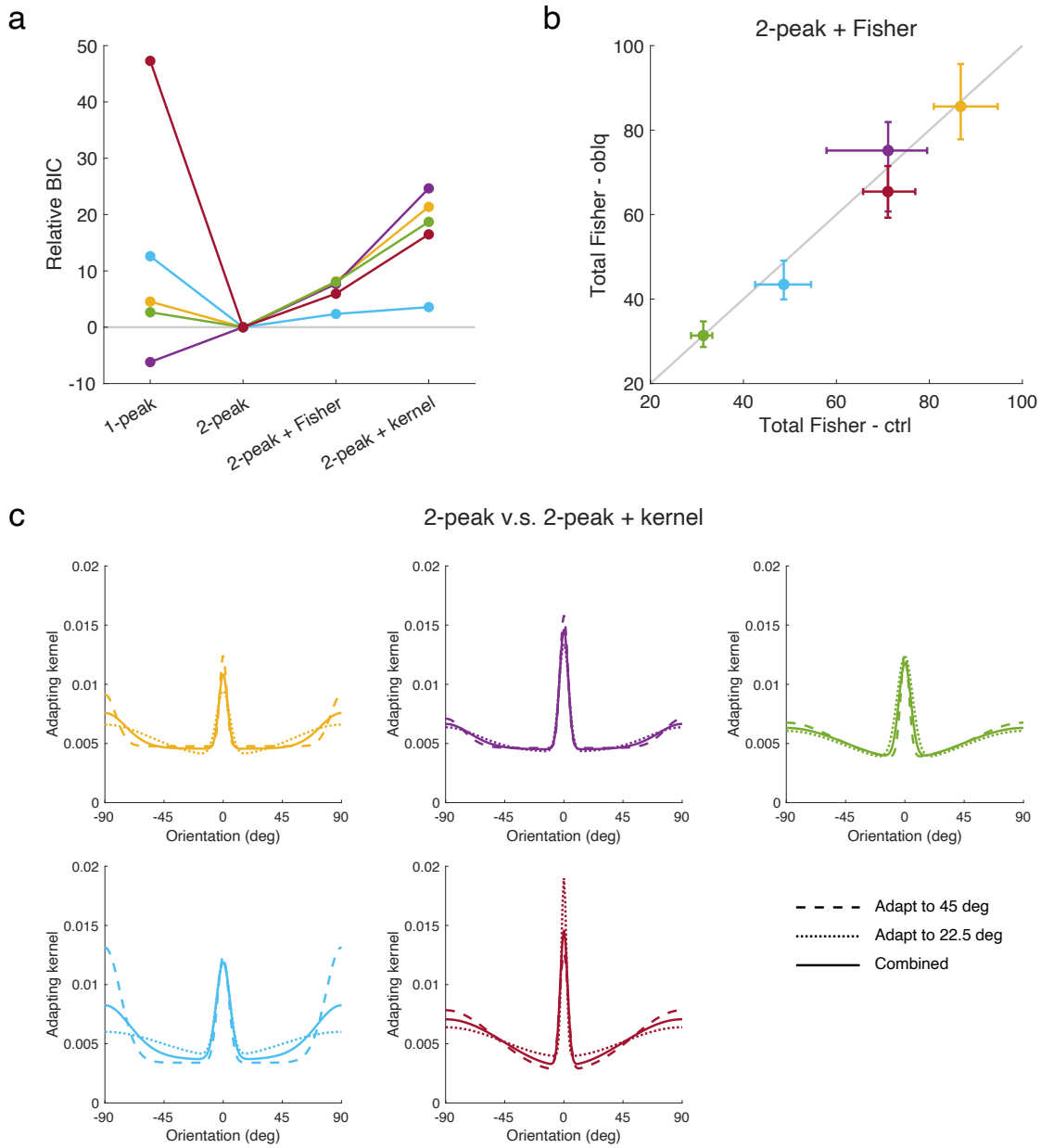

Figure 2.5: Model comparison. (a) BIC of different models relative to the 2-peak model for each subject. Most subjects are best described by the 2-peak model. (b) Total Fisher information for each subject under control and oblique adaptor condition when they are allowed to vary separately (2-peak + Fisher model). All subjects fall close to the unity line, suggesting that adaptation state does not alter the total coding resources. Error bars represent 95% confidence intervals from 100 bootstrap samples of the data. (c) Adaptation kernels for each subject fitted to 45 deg adaptor (dashed line), 22.5 deg adaptor (dotted line), and both combined (solid line). The kernels are similar within most subjects, suggesting that adaptation kernel is the same for adaptors with different orientations. Different colors represent different subjects as in Fig. 2.4.

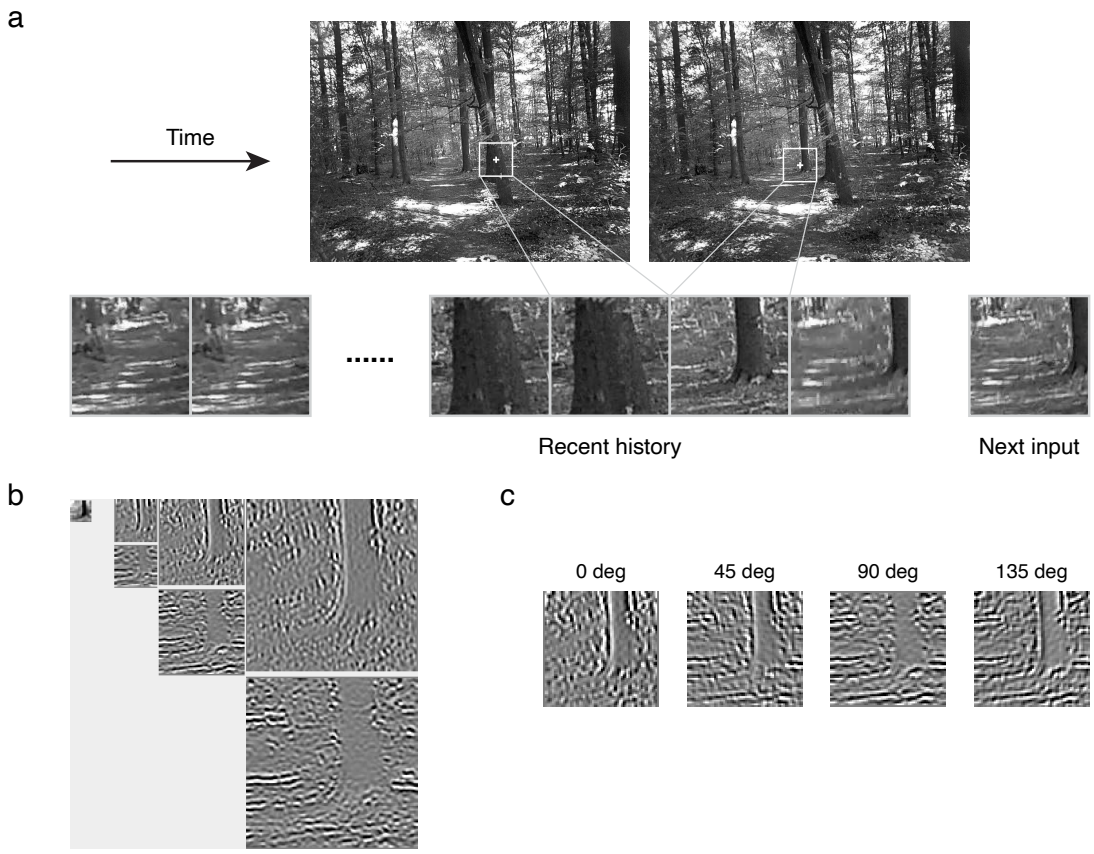

Figure 2.6: Natural scene video dataset and the steerable pyramid. (a) We extracted the area centered at the fixation from each frame of the natural scene videos and examined the next frame as future sensory input based on a recent short-term history. (b) The steerable pyramid with Kth-order directional derivative operators decomposes an image into K+1 orientation channels on multiple spatial frequency subbands. Shown are the decomposition of an image into three spatial frequency subbands with 1st-order directional derivative operators and the low-pass residual. (c) The response of the image to the filter on the second subbbands steered to different orientations. By steering the orientation filter, we can compute the strength of any orientation at every position of the image. Video frames are from an unpublished dataset from Constantin Rothkopf.

so accordingly increasing the coding accuracy for stimulus similar to the adaptor would be efficient. In order to verify this hypothesis, we examined the natural scene videos filmed with head-fixed camera by people walking in the woods with their eye-movement recorded (unpublished dataset from Constantin Rothkopf). We extracted a 6*6 deg patch at the center of fixation from each frame, and examined the orientation in the next frame relative to a recent history of 3 seconds (Fig. 2.6a). We extracted the orientation at each position using the steerable pyramid (Simoncelli and Freeman, 1995). The steerable pyramid is a linear multi-scale, multi-orientation image decomposition tool. The basis functions of the steerable pyramid are Kth-order directional derivative operators. It decomposes the image into a pyramid of multiple spatial frequency subbands and K+1 orientation channels (Fig. 2.6b). The decomposition is steerable in that the image can be equivalently decomposed with rotated basis functions without artifacts. The steerable pyramid allows us to extract local orientations at different spatial frequency levels by rotating the orientation filters and looking for the rotation angle that results in the largest value at each position (Fig. 2.6c). After extracting the orientation at each position of each frame, we then computed the mean and variance of orientation within a short-term history of 3s and find the time and position where the variance over the short history is small (circular variance smaller than 0.1). We calculated the difference of the orientation in the next frame and the mean orientation in the recent short-term history, and found that the orientation at the same position in the next frame is mostly likely to be similar to the mean orientation in the immediate past (Fig. 2.7). The smaller the history variance, the more concentrated the distribution of the next orientation, and this is consistent across spatial frequency levels (Fig. 2.8). Such statistical regularity implies that reallocating Fisher information towards the adaptor is consistent with the efficient coding hypothesis in natural viewing conditions.

2.3.5. Predictive Neural Network

We further tested the hypothesis that adaptation optimally prepares the perceptual system for the future with PredNet, a recurrent neural network trained on natural scene videos to predict the next frame (Lotter et al., 2016). PredNet was inspired by the concept of "predictive coding" in neuroscience, which assumes that the neural system does not directly encode the input; instead, it makes prediction about the input and represent its deviation from the true input, usually with top-down

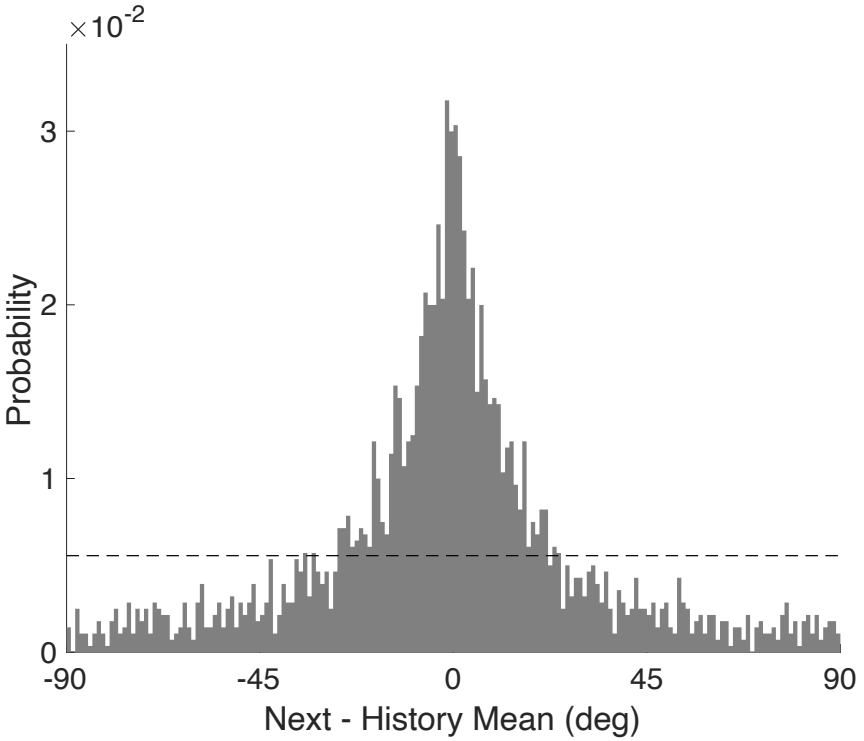

Figure 2.7: Distribution of orientation in the next frame relative to history mean when the variance of orientation during the immediate past is small in natural scene. The distribution centers narrowly around zero, showing that the future sensory input is very likely to be similar to a stable recent history. Data are analyzed from an unpublished dataset from Constantin Rothkopf.

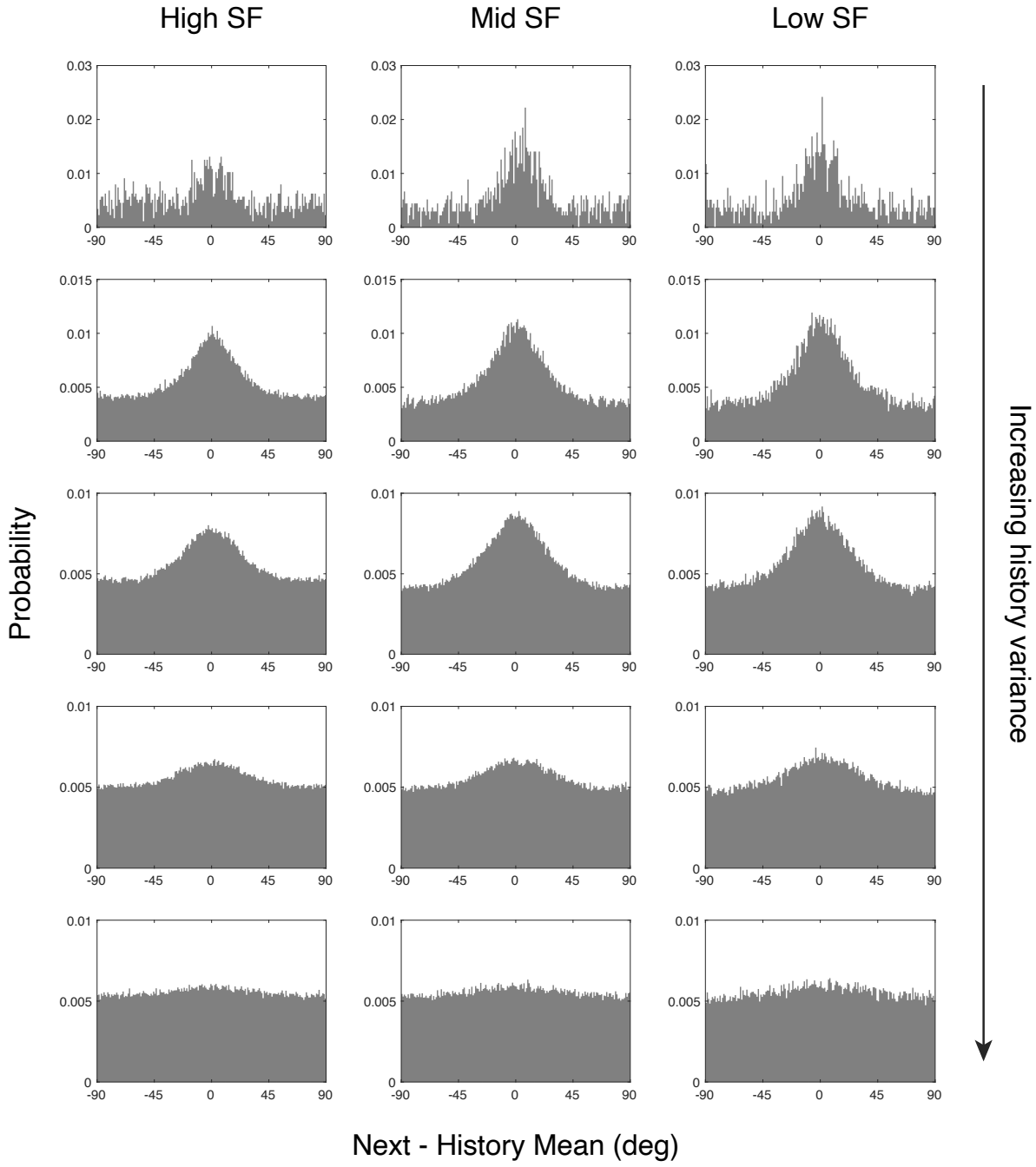

Figure 2.8: Distribution of orientation in the next frame relative to history mean for different spatial frequencies and different history variance. The three spatial frequency levels were calculated from level 2 to 4 of the steerable pyramid. Variance are circular variance ranging from 0 to 1 and binned into bins of 0.2. For all spatial frequency levels, the distribution of orientation in the next frame is more concentrated around the history mean as the variance becomes smaller.

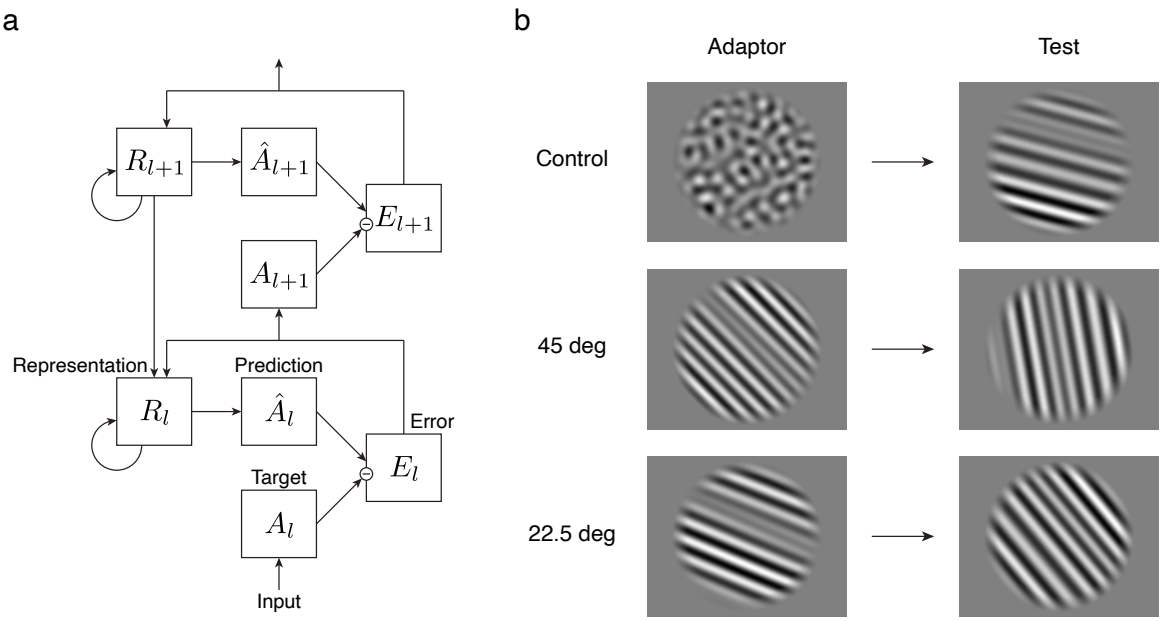

Figure 2.9: PredNet architecture and adaptation experiment in PredNet. (a) Architecture of PredNet (Lotter et al., 2016). Each layer of PredNet consists of four sub-layers: a representation layer (R_l), a prediction layer ($\hat{A}_l$), an input or target layer (A_l), and an error layer (E_l). The representation layer takes feedback from the error layer and the higher representation layer and makes predictions about the next input. The error layer takes the difference between the prediction and the input and passes on to the next layer. The network used in this paper has four large layers in total. (b) Adaptation experiment in PredNet. In each trial, PredNet is presented with four frames of adaptor followed by one frame of test stimulus. There are three types of adaptors: control adaptor, 45 deg and 22.5 deg adaptors. Test stimuli are filtered white noise patterns with different orientations.

connections conveying local predictions of incoming stimuli and bottom-up signals of the deviations from the predictions. There are four sub-layers in each layer of PredNet: a recurrent representation layer (R_l), a prediction layer ($\hat{A}_l$), an input or target layer (A_l), and an error layer (E_l) (Fig. 2.9a). The representation layer takes feedback from the error layer and the higher representation layer and generates predictions of the input layer; the error layer calculates the deviation of the prediction layer from the input layer and passes on to the next input layer. We measured the adaptation effect in PredNet by running an experiment similar to the human experiment on the network (Fig. 2.9b). We input a sequence of adaptor frames followed by a test frame, then we computed the Fisher information in the lowest representation layer in response to different orientations in the test frame, and compared results for different adaptors. Because PredNet does not have any built-in local adaptation mechanism (e.g. neural gain change), any adaptation effect we may observe in PredNet is due to the statistical regularities in natural videos and the task of the network to best predict the next frame.

We took the PredNet pretrained by Lotter et al. (2016) using natural videos from the KITTI dataset (Geiger et al., 2013). Similar to the human experiment, we tested three different adaptors: the control adaptor containing all orientations, and the oblique adaptors containing only 45 or 22.5 deg orientation; and the test images were noise patterns with different orientations (Fig. 2.9b). We computed the Fisher information to different orientations θ in each adaptation conditions using the activation $f(\theta)$ in the first representation layer (R_1) in response to the test frame assuming independent Gaussian noise (Fig. 2.10a):

$$J(\theta) = \left\| \frac{\partial f}{\partial \theta} \right\|_2^2 . \tag{2.1}$$

Similar to the results from human subjects, in the control condition, PredNet has higher Fisher information at cardinal orientations and lower Fisher information around oblique orientations; after adapting to an oriented adaptor, Fisher information exhibits a peak around the adaptor orientation. Finally, we computed the ratio of Fisher information between the oblique adaptor conditions and the control adaptor condition (Fig. 2.10b). We found that after adapting to a single orientation,

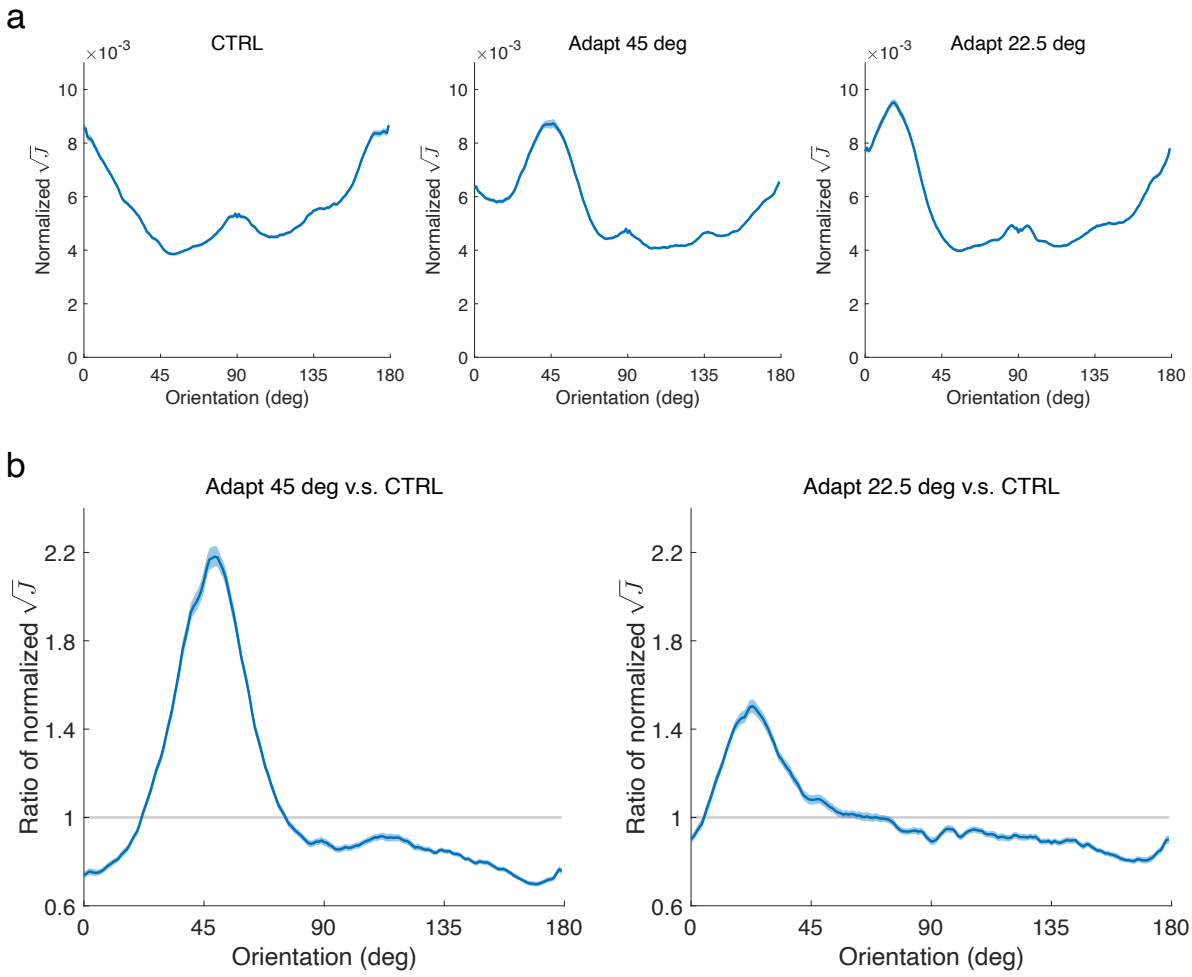

Figure 2.10: Fisher information recovered from the first representational layer of PredNet after adaptation. (a) Normalized square root of Fisher information averaged across test sequences. In the control condition, Fisher information is higher at cardinal orientations; after adapting to a single orientation, Fisher information near the adaptor is the highest. (b) Ratio of mean normalized square root of Fisher information between adapting to oriented versus control adaptor. Compared to the control condition, Fisher information is higher at the adaptor orientation and lower away from the adaptor, which is consistent with human behavior. Shaded areas represent 95% confidence intervals from 1000 bootstrap samples of the data.

the Fisher information in PredNet increases at the adaptor and decreases away from the adaptor compared to the control condition. Thus PredNet exhibits similar adaptation effect to human observers. This suggests that the reallocation of Fisher information towards the adaptor is beneficial for the representation of the next stimulus.

2.4. Discussion

We measured the discrimination threshold of orientation across the entire range of orientation after adapting to a single adaptor, and extracted the change in coding accuracy after adapting to different adaptors. We found that after adapting to a single orientation, discrimination decreased (Fisher information increased) at the adaptor and opposite of the adaptor compared to adapting to a control adaptor. Further modeling of the data revealed that one adapting kernel can describe the change in Fisher information for adaptors of the same format but with different feature values. By analyzing natural viewing input statistics, we found that increased Fisher information at the adaptor is efficient for encoding future sensory input in natural viewing condition. By conducting similar experiment on a recurrent neural network trained to predict the next frame of a video (PredNet), we showed that the neural network exhibited a reallocation of coding resources similar to that of human observers, which suggested that such resource reallocation is optimal for representing future sensory input. Previous studies have also measured the adaptation effect on orientation discrimination. However, some of them only tested orientations close to the adaptor and failed to test the entire 180 deg range (Regan and Beverley, 1985). Others tested the same orientation after adapting to different orientations (Clifford et al., 2001), with the implicit assumption that adaptation effect depends on the difference between the adaptor and the test orientation and is invariant to the absolute orientation, which is explicitly tested by our experiment. The significance of our work lies in mapping out the changes in encoding accuracy across the entire feature space after the system adapts to a single adaptor, and using a combination of natural scene statistics analysis and recurrent neural network to verify the efficient coding account of adaptation.

We replicated the debated finding of Clifford et al. (2001) that discrimination threshold decreases after adapting to the orthogonal orientation with two different adaptor orientations, although this

effect was not significant for all subjects in all adaptors, and the effect size is smaller than when the test orientation is the same as the adaptor orientation. We were able to verify the efficient coding hypothesis for the improved discrimination at the adaptor, but the efficient coding explanation cannot account for the orthogonal improvement. From a mechanistic point of view, Dragoi et al. (2002) showed that sharpened neural selectivity near the orthogonal orientations could explain the enhanced identification of orthogonal orientations after brief adaptation. Schwartz et al. (2007) also showed that the repulsive shift of orientation tuning curve away from the adaptor (Dragoi et al., 2000; Patterson et al., 2013) could lead to increased Fisher information at the orthogonal orientation. However, a normative explanation of the orthogonal improvement remains unclear. It is also possible that the orthogonal improvement is due to biological constraints: the wiring of neurons might constrain that the coding of orthogonal orientations changes simultaneously.

We measured natural scene statistics from natural scene videos filmed by head-mounted camera on human participants walking in the woods paired with eye-tracking. Compared to the history dependent distribution of orientation in natural videos measured in van Bergen and Jehee (2019), the dataset we used were filmed from humans' view point (instead of cat or static camera) and included potential regularities induced by active viewing through head motion and eye movement. One drawback of the current dataset is that it only consists of forest scenes. The statistics would be more comprehensive if we included some city scenes and indoor scenes in the analysis.

We tested the hypothesis that adaptation is optimal for the representation of future sensory input by demonstrating that the change in coding accuracy in PredNet after adaptation is similar to the adaptation effect in human observers. The fact that PredNet is trained on natural videos to predict future frames and no adaptation mechanism is built into the network implies that adaptation emerges from optimizing the representation of future sensory input based on natural statistics, supporting the efficient coding hypothesis. Artificial neural networks have been a powerful model to study the mechanism and normative explanation of neuronal and behavioral phenomena (Olshausen and Field, 1996; Singer et al., 2018). If a network built with a certain mechanism or trained to optimize a certain functional goal shows some neural properties or behavioral patterns, we can

imply that such neural properties or behaviors might be a result of said mechanism or functional goal. While we to approach the adaptation problem from a normative point of view, other studies have used artificial neural network to study adaptation from a mechanistic perspective. Vinken et al. (2020) introduced fatigue to each unit in AlexNet and showed that neurophysiological and perceptual properties of adaptation such as novelty detection, tuning curve shifts and perceptual bias readily emerge from the propagation of activation-based intrinsic suppression, suggesting that fatigue is a possible mechanism of adaptation. This result does not necessarily contradict the efficient coding explanation. While efficient coding might be the coding principle that the neural system is optimize for, fatigue can be the mechanism by which the system achieves efficient coding.

Conclusions The efficient coding hypothesis has been a prominent explanation of sensory adaptation, but a comprehensive neural measurement that allows the calculation of Fisher information makes it difficult to verify the efficient coding principle in the strictest term of maximizing information transmission. The present study mapped out the change in coding accuracy across the feature space after adapting to a single feature value through a psychophysical experiment, and found a universal parametric description of the reallocation of coding resources for different adaptors. Further analysis of natural scene statistics and artificial neural network provided support for the efficient coding hypothesis of adaptation: adaptation induced changes in encoding accuracy and perceptual behavior reflect the visual systems' attempt to best possibly represent the next expected sensory input.

2.5. Methods

2.5.1. Experimental methods

Subjects. 5 subjects participated in the experiment, one of which was the author. All subjects had corrected-to-normal vision.

Apparatus. The experiment was run using Matlab and PsychToolbox (Brainard and Vision, 1997). Participants sat in a dark room and viewed stimuli on a VPixx3D screen (1920*1080 pixels resolution, 120 Hz refresh rate) at a 89 cm distance. A circular aperture (26 cm diameter) was placed on the screen to occlude the edges of the screen, removing cardinal orientation cues.

Stimuli. All stimuli were presented on a gray background. The stimuli consisted of filtered white noise patterns. The control adaptor were filtered by a band-pass filter with a uniform profile within the spatial frequency range of 3.75 – 5.25 cpd. The oblique adaptor and test stimulus were further filtered by an orientation filter with a symmetric wrapped Laplace profile centered at the desired orientation with a standard deviation of 1.4 deg, in addition to the same band-pass filter as the control adaptor. Stimuli had 80% contrast. Stimuli were 2 deg in diameter, presented 1.67 deg to the left or right of fixation. A fixation dot was presented at the center of the screen throughout the experiment.

Procedure. Subjects were instructed to fixate at the fixation dot throughout the experiment. At the beginning of a block, subjects viewed adaptation stimuli for 60s. Each trial consisted of a top-up adaptation (5s), a blank interval (0.35s), a test screen (0.1s), and a response period. During the adaptation period, two identical adaptor patterns were presented on both sides of fixation, updating every 1.25s, with a 0.05s blank period between presentations. In the test screen, a test stimulus and a reference stimulus with slightly different orientations were presented to the left and right of fixation randomly. During the response period, subjects pressed one of two buttons on a gamepad to indicate which stimulus in the test screen was more clockwise (or counterclockwise, interleaved across blocks).

The experiment consists of two parts, each of which contains an oblique adaptor condition and a control adaptor condition. In the first half of the experiment, the oblique adaptor was oriented at ±45 deg. The test orientations were [0, ±10, ±30, ±45, 90] deg relative to the adaptor orientation (subject 1 had ±5 deg in addition). In the second half of the experiment, the oblique adaptor was oriented at ±22.5 deg. The test orientations were [0, ±10, ±22.5, ±45, 90] deg relative to the adaptor orientation. Test orientations were randomized across trial. The reference orientation varied according to a 2-up-1-down staircase procedure in 25 equal steps within a ±9.6 deg, ±15 deg, ±18 deg, or ±24 deg range relative to the test orientation, depending on the performance of each subject in the training session prior to the experiment.

Subjects completed 192 trials for each test orientation in each adaptation condition (216 trials for

subject 1 in the first half of the experiment). In each half of the experiment, subjects completed the control adaptor condition first, oblique adaptor condition next. In the oblique adaptor condition, subjects completed 4 blocks of one adaptor orientation, then 4 blocks of the opposite adaptor orientation (6 blocks each for subject 1 in the first half), e.g. first 45 deg then -45 deg. The order of the two adaptor orientations were counterbalanced across subjects. The control adaptor condition consisted of 8 blocks (12 blocks for subject 1 in the first half). Each block lasted for about 25min, depending on the response time of each subject. Blocks with different adaptors (control versus oblique adaptor condition, or opposite oblique adaptor orientations) were completed at least one day apart.

2.5.2. Data analysis

In the main analysis, all orientations in blocks with -45 or -22.5 deg adaptor were mirrored with respect to the vertical orientation (0 deg) to be combined with data from blocks with 45 or 22.5 deg adaptor, including the control condition with the assumed adaptor orientations.

Psychometric curves were obtained by fitting cumulative Gaussian distributions to the data. We assumed a mean of 0 deg for the Gaussian distribution and zero lapse rate. Thresholds were 75% discrimination thresholds from the fitted psychometric curves.

2.5.3. Modeling

Pre-adaptation. The following derivation follows Wei and Stocker (2015). Let θ be the orientation of the test stimulus and m its sensory measurement in a given trial. Before adapting to a single orientation, or in the control condition in our experiment, the discrimination threshold is typically lower at cardinal orientations, which implies higher Fisher information at cardinal orientations. We assume that the square root of Fisher information distribution is proportional to the weighted sum of a uniform distribution and two identical von Mises distributions centered at two cardinal orientations:

$$\sqrt{I_{\text{ctrl}}(\theta)} \propto k \ \text{vm}(\theta; 0, \kappa) + k \ \text{vm}(\theta; \pi, \kappa) + \frac{1-2k}{2\pi} \ , \tag{2.2}$$

where κ represents the width of Fisher information around cardinal orientations, and k represents the intensity. Note that because angles are defined on a 2π range, while orientation is define on a range of π, θ in this equation represents orientation multiplied by 2, and 0 and π represents cardinal orientations. We use this convention in the following.

Consider a sensory space in which Fisher information is uniform. The mapping $\tilde{\theta} = F(\theta)$ from stimulus to this sensory space is the cumulative of the square root Fisher distribution, thus $F(\theta) \propto \int \sqrt{I_{\text{ctrl}}(\theta)}\mathrm{d}\theta$. Assume homogeneous von Mises likelihood in the sensory space:

$$p(\tilde{m}|\tilde{\theta}) = \text{vm}(\tilde{m}; \tilde{\theta}, \kappa_i) \ , \tag{2.3}$$

where κ_i represents the sensory noise magnitude. The likelihood function in stimulus space $p(m|\theta)$ can be computed by applying the inverse mapping $\theta = F^{-1}(\tilde{\theta})$. Finally, the posterior over stimulus orientation given the sensory measurement is

$$p(\theta|m) \propto p(m|\theta)p(\theta) \ , \tag{2.4}$$

where $p(\theta)$ is the prior distribution over orientation.

Post-adaptation. After adapting to a single orientation θ_a, the distribution of the square root of Fisher information in the pre-adaptation sensory space becomes

$$\sqrt{\tilde{I}_{\text{adapt}}(\tilde{\theta}; \theta_a)} \propto p_a(\tilde{\theta} - \tilde{\theta_a}) \ , \tag{2.5}$$

where $\tilde{\theta_a} = F(\theta_a)$ is the adaptor orientation in the pre-adaptation sensory space, and p_a is the adaptation kernel. Thus, adaptation reallocates Fisher information by the same adaptation kernel shifted according to the adaptor. Now, the pre-adaptation sensory space is not a uniform space any more. A post-adaptation uniform sensory space can be obtained by applying another transformation $\theta^* = F_a(\tilde{\theta}) \propto \int \sqrt{\tilde{I}_{\text{adapt}}(\tilde{\theta})}\mathrm{d}\tilde{\theta}$. The likelihood in the post-adaptation sensory space is homogeneous von Mises function with the same the same internal noise parameter as the likelihood function in the pre-adaptation sensory space before adaptation (Eq. 2.3), so the total coding resource does not

change after adaptation. The likelihood function in the stimulus space can be obtained by applying the inverse transformation $\tilde{\theta} = F_a^{-1}(\tilde{\theta^*})$ and $\theta = F^{-1}(\tilde{\theta})$, and the posterior over stimulus orientation given the sensory measurement is

$$p(\theta|m) \propto p(m|\theta)p_{\text{adapt}}(\theta;\theta_a) \ , \tag{2.6}$$

where $p_{\text{adapt}}(\theta;\theta_a)$ is the prior distribution after adapting to θ_a.

In the 2-peak models, we assume that the adaptation kernel $p_a(\tilde{\theta})$ is the weighted sum of two von Mises distributions and a uniform distribution:

$$p_a(\tilde{\theta}) = k_1 \ \text{vm}(\tilde{\theta};0,\kappa_1) + k_2 \ \text{vm}(\tilde{\theta};\pi,\kappa_2) + \frac{1-k_1-k_2}{2\pi} \ , \tag{2.7}$$

where k_1 and k_2 represent the intensities of the two peaks, and κ_1 and κ_2 represent the widths of the two peaks. In the 1-peak model, we assume that $p_a(\tilde{\theta})$ is the weighted sum of one von Mises distribution and a uniform distribution:

$$p_a(\tilde{\theta}) = k_1 \ \text{vm}(\tilde{\theta};0,\kappa_1) + \frac{1-k_1}{2\pi} \ , \tag{2.8}$$

where k_1 and κ_1 represents the intensity and the width of the peak respectively. In the "2-peak + Fisher" model, the total Fisher information is allowed to change after adaptation, so the width of the von Mises likelihood in the post-adaptation sensory space κ_i^a is allowed to be different from the the width of the von Mises likelihood in the pre-adaptation sensory space κ_i. In the "2-peak + kernel" model, the adaptation kernel $p_a(\tilde{\theta})$ for 45 deg and 22.5 deg adaptor are allowed to be different.

Discrimination decision and response distribution. Let θ_t and θ_r be the orientation of the test and reference stimulus, and m_t and m_r their sensory measurements respectively. The posterior of each stimulus can be computed according to Eq. 2.4 or Eq. 2.6. The probability of the reference

orientation being more clockwise than the test orientation is

$$p(\theta_t - \pi < \theta_r < \theta_t | m_t, m_r) = \int_0^{2\pi} p(\theta_t|m_t) \int_{\theta_t - \pi}^{\theta_t} p(\theta_r|m_r)\, \mathrm{d}\theta_r\, \mathrm{d}\theta_t \ . \tag{2.9}$$

If the probability is larger than 0.5, the observer responds the reference orientation being more clockwise; otherwise, the subject responds the test orientation being more clockwise.

In our experiment, because the test and reference stimulus were always under the same adaptation state and had the same spatial frequency, contrast and presentation duration, we assume that they had the same prior distribution and sensory noise. So the decision rule above simplifies to directly comparing the measurements of the two stimuli: if m_r is more clockwise than m_t, the subjects respond the reference orientation being more clockwise, and vice versa. So the probability of responding the reference stimulus being more clockwise is:

$$p(\text{“}\theta_\mathrm{r} \text{ more CW”} | \theta_t, \theta_r) = p(m_t - \pi < m_r < m_t | m_t, m_r) = \int_0^{2\pi} p(m_t|\theta_t) \int_{m_t - \pi}^{m_t} p(m_r|\theta_r)\, \mathrm{d}m_r\, \mathrm{d}m_t \ . \tag{2.10}$$

2.5.4. Model fitting

We fit the model to the data by maximizing the likelihood of the data given the model:

$$p(D|\rho) = \prod_{j=1}^{n} p(D^j|\rho) = \prod_{j=1}^{n} p(r^j|\rho, \theta_t^j, \theta_r^j) \ , \tag{2.11}$$

where D is the data, ρ represents the parameters of the model, θ_t^j and θ_r^j are the test and reference orientation and r^j is the response in trial j, and n is the total number of trials.

We first fit the model to the control adaptor condition with the following free parameters:

- κ_i for sensory noise;
- k for the intensity of the prior;
- κ for the width of the von Mises prior.

Then we fix these parameters and fit the model to the oblique adaptor condition. The 2-peak model has four free parameters for the adaptation kernel:

- k_1 and k_2 for the strength of the two peaks;
- κ_1 and κ_2 for the width of the two von Mises.

The 2-peak + Fisher model has an additional parameter κ_i^a for sensory noise after adapting to an oblique adaptor. The 2-peak + kernel model has four parameters for the adaptation kernel of each oblique adaptor, eight in total. The 1-peak model has only two free parameters for the strength and width of the adaptation kernel.

2.5.5. Natural scene statistics

Dataset. The videos were filmed with a head-mounted camera by participants freely walking in the woods (24 fps, 1290*960 pixels resolution, spanning 60*46 deg visual angle). Eye movement was recorded (120 samples per second). We included videos from 9 participants, with a total length of 12 minutes. Videos were converted to grayscale for analysis.

Data analysis. We looked at a 6*6 deg area centered at the fixation in each frame of the videos. We extracted the orientation at each position within the area using the steerable pyramid (Simoncelli and Freeman, 1995). We rotated and applied the 1st-order steerable filter to find the orientation with the strongest response as the orientation of each position. We computed the mean and circular variance of orientation in 3s (72 frames) at each position, then computed the difference between the orientation in the next frame and the mean orientation in the previous 3s. We included the results from level 2 to 4 of the steerable pyramid. In Fig. 2.7, we included time and positions where the history variance is smaller than 0.1 and combined the data from three levels.

2.5.6. PredNet

PredNet is a recurrent neural network that predicts the next frame of a video. We used the PredNet pretrained by Lotter et al. (2016) using the KITTI dataset (Geiger et al., 2013) in our experiment.

Stimuli. The stimuli were images of filtered white noise patterns with the size of 128*160 pixels. The control adaptors were filtered by the orientationally averaged spatial frequency spectrum extracted from the training dataset within the range of 8-12 cycles per image. The spatial frequency filter was obtained by taking the average of the 2D Fourier transformation of the image across all frames and averaging across orientation for each spatial frequency, with a cutoff at 8 and 12 cycles per image. The oblique adaptors and test stimuli were further filtered by an orientation filter with a symmetric wrapped Laplace profile centered at the desired orientation with a standard deviation of 1.4 deg, in addition to the same spatial frequency filter as the control adaptor. Stimuli had 100% contrast. The noise pattern was embedded in a circular aperture at the center of the image; the contrast of the noise pattern fades linearly from 100% to 0 as the distance from the center goes from 48 to 60 pixels.

Procedure. Each test sequence consists of an adapting sequence of four different adapting frames followed by a testing frame. We input the test sequence to PredNet and extract the activation of the first representational layer in response to the testing frame. In each adaptation condition, we tested 200 test sequences; the testing frame of each test sequence was rotated and tested in all orientations from 0 deg to 359 deg in 1 deg interval.

Calculating Fisher information. Assuming independent Gaussian noise, Fisher information can be calculated for each test sequence according to Eq. 2.1. For orientation, we combined 0 – 179 deg and 180 – 359 deg. Because we focus on the distribution of coding resources, we normalized the square root of Fisher information across orientations. The ratio of the square root of Fisher information was computed by taking the average of the normalized square root of Fisher information across test sequences within each adapting condition, then taking the ratio between two conditions. Confidence intervals were obtained by bootstrapping the test sequences and computing the corresponding Fisher information distributions and their ratios in each bootstrap sample.

Parameter	Subj 1	Subj 2	Subj 3	Subj 4	Subj 5
Control					
κ_i: sensory noise	191.04	128.63	25.44	60.54	128.46
k: prior intensity	0.17	0.24	0.12	0.25	0.15
κ: prior width	14.73	22.47	15.49	14.69	16.83
2-peak					
k_1: kernel strength	0.05	0.08	0.09	0.11	0.08
κ_1: kernel width	74.8	77.2	39.7	31.8	105.7
k_2: kernel strength	0.13	0.11	49.47	0.25	49.49
κ_2: kernel width	3.17	2.32	0.0046	2.17	0.0071
1-peak					
k_1: kernel strength	0.04	0.07	0.07	0.07	0.06
κ_1: kernel width	127.2	93.4	64.2	77.0	182.8
2-peak + Fisher					
κ_i^a: sensory noise (adapt)	185.08	143.58	25.43	48.33	108.98
k_1: kernel strength	0.05	0.08	0.09	0.12	0.08
κ_1: kernel width	74.4	78.4	39.7	30.4	108.9
k_2: kernel strength	0.12	0.17	47.23	0.18	1.15
κ_2: kernel width	3.53	1.52	0.0048	2.45	0.27
2-peak + kernel					
k_1^{45}: 45° kernel strength	0.05	0.09	0.07	0.09	0.08
κ_1^{45}: 45° kernel width	134.6	75.2	92.0	49.2	92.3
k_2^{45}: 45° kernel strength	0.13	0.01	0.09	0.21	0.39
κ_2^{45}: 45° kernel width	5.21	700.0	7.53	7.52	1.22
$k_1^{22.5}$: 22.5° kernel strength	0.06	0.07	0.12	0.09	0.07
$\kappa_1^{22.5}$: 22.5° kernel width	35.8	79.1	22.9	46.4	191.9
$k_2^{22.5}$: 22.5° kernel strength	0.12	0.19	415	0.35	381
$\kappa_2^{22.5}$: 22.5° kernel width	1.96	1.27	5.7e-4	0.71	7.6e-4

Table 2.1: Best-fitting model parameters for individual subjects.

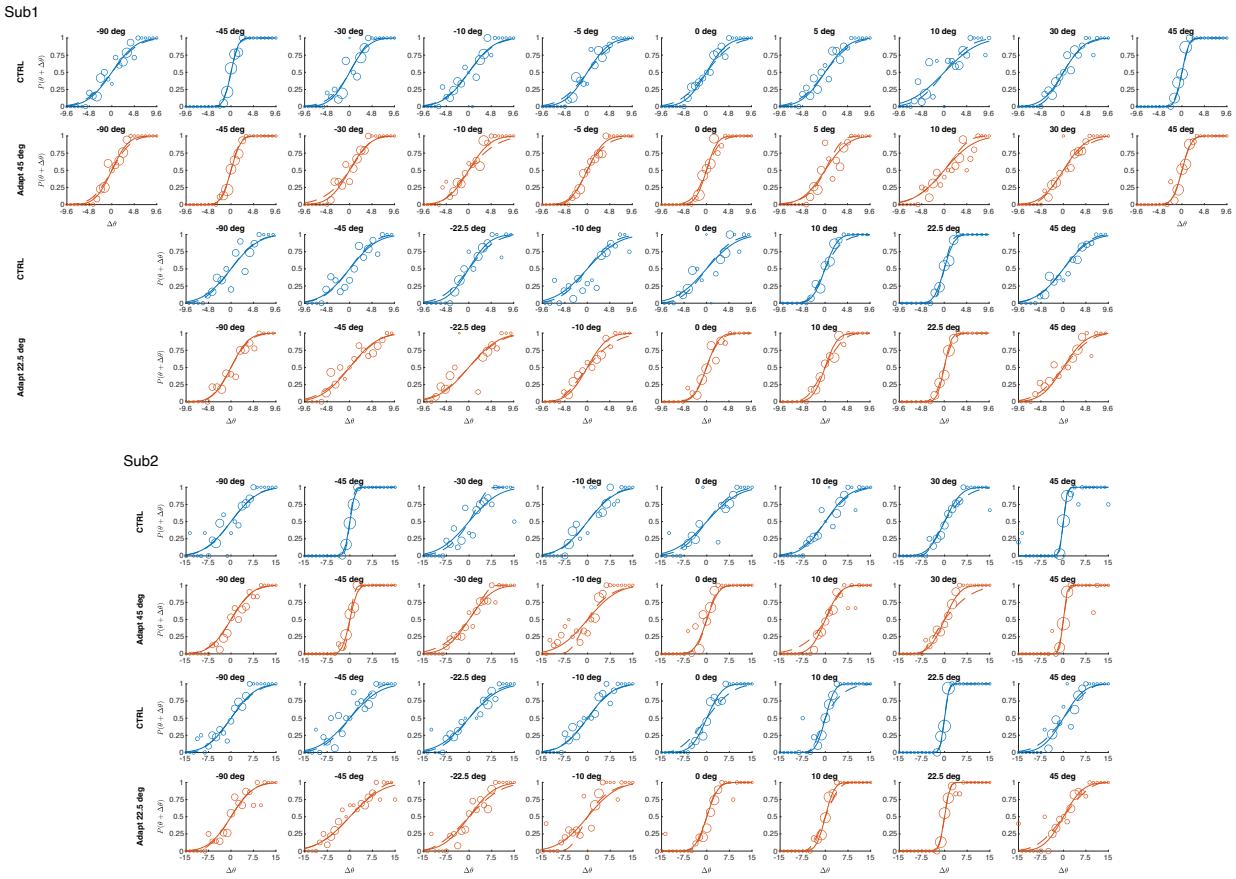

Figure 2.11: 2AFC response data of subject 1 & 2 and psychometric curves fitted by a cumulative Gaussian distribution (solid line) or the reallocation model (dashed line). Size of the data point represents the number of trials. The title above each subplot indicates the difference between the test and the adaptor orientation.

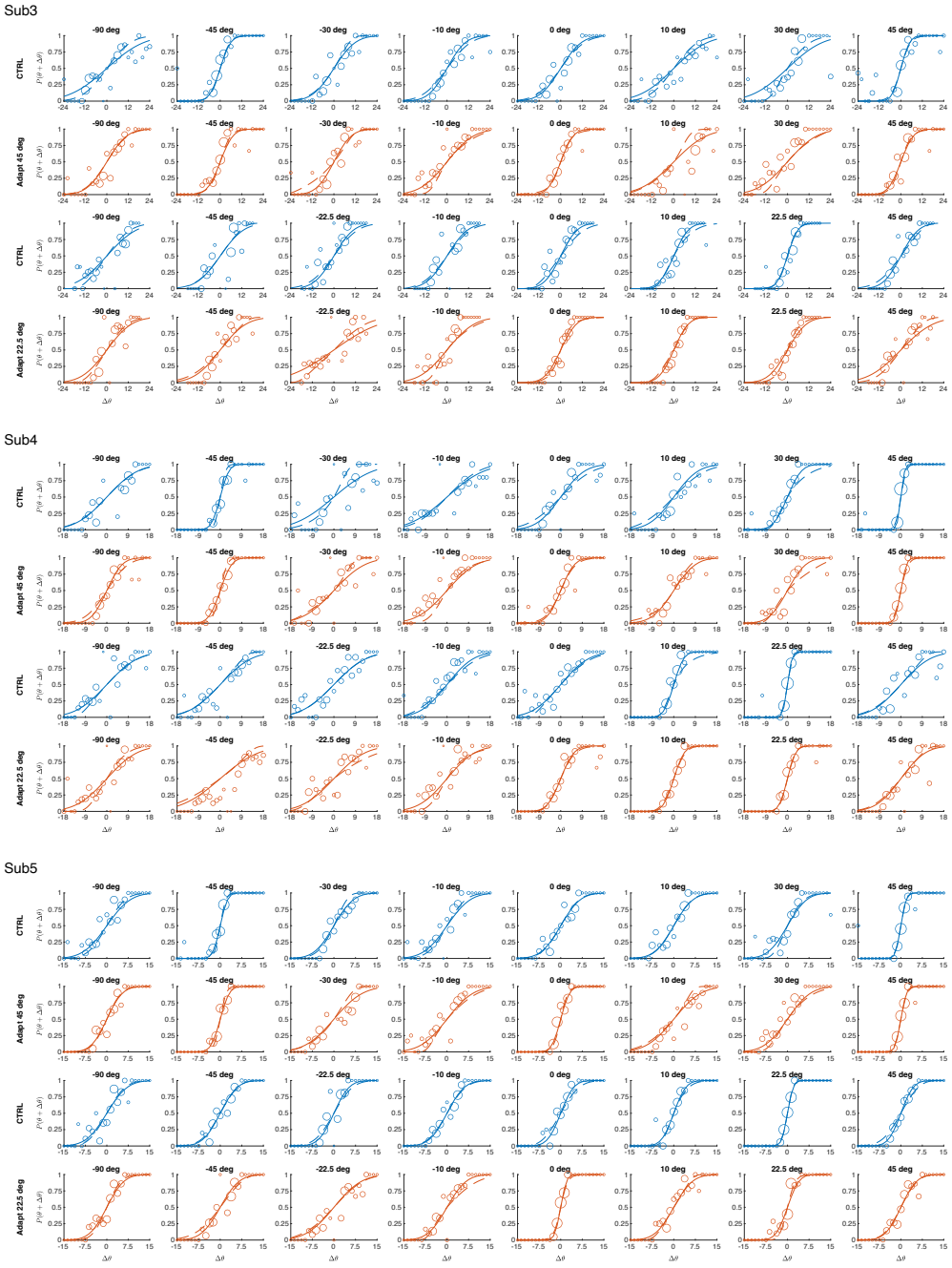

Figure 2.12: 2AFC response data of subject 3, 4 & 5 and psychometric curves fitted by a cumulative Gaussian distribution (solid line) or the reallocation model (dashed line). Size of the data point represents the number of trials. The title above each subplot indicates the difference between the test and the adaptor orientation.

CHAPTER 3

SENSORY PERCEPTION IS A HOLISTIC HIERARCHICAL INFERENCE PROCESS

3.1. Abstract

Perception of stimulus features such as orientation is widely considered a Bayesian inference process. In contrast to previous Bayesian observer models, we propose that perception is a *holistic* inference process that simultaneously operates at all levels of the representational hierarchy. We test this hypothesis in the context of a typical psychophysical matching task in which subjects are asked to estimate the perceived orientation of a test stimulus by adjusting a probe stimulus (method-of-adjustment). We present a holistic matching model that assumes that subjects' responses reflect an optimal match between the test and the probe stimulus, both in terms of their inferred feature (orientation) but also their higher-level representations (category). Validation against multiple existing psychophysical datasets and data from a new psychophysical experiment demonstrates that compared to previous models, our model provides a quantitatively accurate and detailed description of subjects' response behavior, which includes data that previous models have failed to even qualitatively account for. We also show that the model generalizes to other feature domains and thus offers a universal explanation for categorical influences in low-level sensory perception.

3.2. Introduction

Perception is considered an inference process that optimally combines noisy sensory signals with prior knowledge about the statistical regularities of the world. Countless studies have argued that models of perceptual inference can be parsimoniously expressed within the probabilistic framework of Bayesian estimation (Knill and Richards, 1996). Perception in this framework equates to finding an optimal estimate of a stimulus feature given noisy sensory observations. A characteristic prediction of Bayesian estimation is that perception is biased towards the peak of the prior density distribution, which has been validated by the results of many perceptual and sensorimotor studies (e.g., Körding and Wolpert, 2004; Stocker and Simoncelli, 2006; Jazayeri and Shadlen, 2010; Kim and Burge, 2018).

A quantitative validation of the Bayesian estimation framework crucially depends on an accurate

specification of the prior distribution. Visual orientation is one of the few stimulus features for which the prior distribution is well specified in form of the local orientation statistics in natural visual scenes. These statistics have been repeatedly measured and show robust peaks at cardinal orientations (Coppola et al., 1998; Girshick et al., 2011; Wang et al., 2016). Perceived stimulus orientation is typically biased away from cardinal orientations (De Gardelle et al., 2010; Noel et al., 2021), i.e., it is seemingly "anti-Bayesian" for this natural prior distribution. Recent work has demonstrated, however, that the efficient coding hypothesis (Attneave, 1954; Barlow et al., 1961) provides a powerful constraint on sensory uncertainty to resolve this apparent paradox, leading to a consistent Bayesian interpretation of visual orientation perception (Wei and Stocker, 2012, 2015, 2017). Since then, the Bayesian estimation model constrained by efficient coding (in the following simply referred to as the "Efficient Bayesian estimator") has demonstrated to offer a unifying account for human behavior in a wide variety of perceptual and working memory tasks (e.g., Taylor and Bays, 2018; Polania et al., 2019; Fritsche et al., 2020; Langlois et al., 2021; Prat-Carrabin and Woodford, 2021).

Despite its promise, however, there are several reasons to question this model's ability to provide a unifying account of orientation perception. First, a full quantitative validation of the model by comparing the entire predicted response distributions against data is still outstanding; previous studies mainly focused on summary statistics such as estimation bias (but see Taylor and Bays (2018)). Furthermore, there are psychophysical data that are difficult to reconcile with the model. Specifically, Tomassini et al. (2010) reported results of a typical orientation matching experiment where subjects were asked to estimate the orientation of a test stimulus by adjusting a probe stimulus. In half of the trials the stimuli used as test and probe were interchanged. If subjects estimated the orientations of test and probe stimulus independently, one would expect the sign of their estimation biases to flip in those trials; a pair of matched stimuli would yield opposite estimation errors when the assignment of test and probe is interchanged. Tomassini et al. (2010) found, however, that the sign of the bias pattern did not flip in those trials – there was a repulsive bias away from cardinal orientations under both conditions. This result cannot be explained with any model that independently estimates the orientations of the test and the probe stimulus. Last but not least,

there is the long-standing notion that higher-level, categorical representations influence perception at the feature level. Several studies have suggested that the perception of visual orientation is affected by a cardinal/oblique category distinction (Rosielle and Cooper, 2001; Wakita, 2004). The efficient Bayesian estimator does not provide the possibility to incorporate potential categorical effects unless the orientation prior implicitly reflects the categories, i.e., has peaks at orientations that correspond to the category centers (e.g., Bae et al., 2015). Allowing such an orientation prior, however, generally violates the fundamental Bayesian assumption that the prior distribution reflects the statistical distribution of visual orientations.

Here, we introduce a hierarchical inference model of perception that resolves these issues. What fundamentally separates our proposal from previous models is that we describe perception as a *holistic inference* process, where the percept of a stimulus is jointly represented by the inference outcomes (i.e., the posteriors) at every level of a representational hierarchy. Specifically with regard to orientation perception, we assume that perceived orientation of a stimulus is characterized by inference at both the feature (orientation) as well as potential higher level representations (orientation categories). Our hypothesis reflects the holistic experience we typically associate with perception. Furthermore, the model assumes that cognitive processes downstream of perception (e.g. a decision stage) operate on these holistic perceptual representations of sensory information. We tested our model in the context of a typical psychophysical matching task in which subjects are asked to estimate the perceived orientation of a test stimulus by adjusting a probe stimulus. We show that our model provides a highly accurate account of several existing datasets: the model not only correctly predicts the non-inverted biases when test and probe stimuli are switched (see above; Tomassini et al., 2010) – something that previous models can not – but also provides a superior quantitative account for the full error distributions of subjects' perceptual estimates reported in other experimental studies (De Gardelle et al., 2010; Noel et al., 2021). Our model also predicts repulsive bias when test and probe stimuli have the same noise, which is qualitatively different from what previous models predict and is verified by our new experiment. Finally, we demonstrate that the model generalizes to other stimulus domains by providing an accurate account of human behavioral data in a color matching experiment based on measured color categories (Bae et al., 2015).

3.3. Results

3.3.1. Holistic perceptual matching

Perception is commonly assessed with a psychophysical matching task often referred to as "the method of adjustment". In this task, a subject is asked, for example, to adjust the orientation of a probe stimulus in order to match the perceived orientation of a test stimulus (Fig. 3.1a). Typically, the probe stimulus is unambiguous and noise-free leading to the general assumption that the probe orientation is a direct reflection of the subject's perceptual estimate of the test stimulus orientation, aside from some potential motor noise. Under this noise-free probe condition, the efficient Bayesian estimator can provide a qualitative accurate account of the repulsive perceptual bias and variance patterns and their dependence on stimulus uncertainty observed in these matching experiments (Wei and Stocker, 2015, 2017; Taylor and Bays, 2018). The model assumes that the perceived test orientation depends on a noisy orientation measurement m and a prior distribution over orientation $p(\theta)$. Bayesian inference then results in a posterior distribution $p(\theta|m)$, based on which the perceived orientation (i.e., the optimal estimate) is determined according to a loss function L_θ.

In contrast, the proposed holistic matching model assumes a hierarchical generative process where each stimulus orientation θ is associated with a category C distinguishing cardinal and oblique orientations (Fig. 3.1b). Furthermore, it assumes that perceptual inference is performed at both levels of the hierarchy. The outcome is thus a holistic representation of the perceived orientation stimulus, jointly represented by the posteriors at both the orientation and the category level, $p(\theta|m)$ and $p(C|m)$, respectively. The key innovation is that the matching stage operates on these holistic perceptual representations. That is, the model assumes that the observer aims to adjust the probe orientation θ_p until the percepts of the probe and the test stimulus optimally match at both representational levels (Fig. 3.1c). We express this as finding the probe orientation that minimizes a weighted average of the expected mismatch at the orientation (L_θ) and the category level (L_c), thus

$$L_{tot}(\theta, \theta_p, C, C_p) = L_\theta(\theta, \theta_p) + wL_c(C, C_p) \ , \tag{3.1}$$

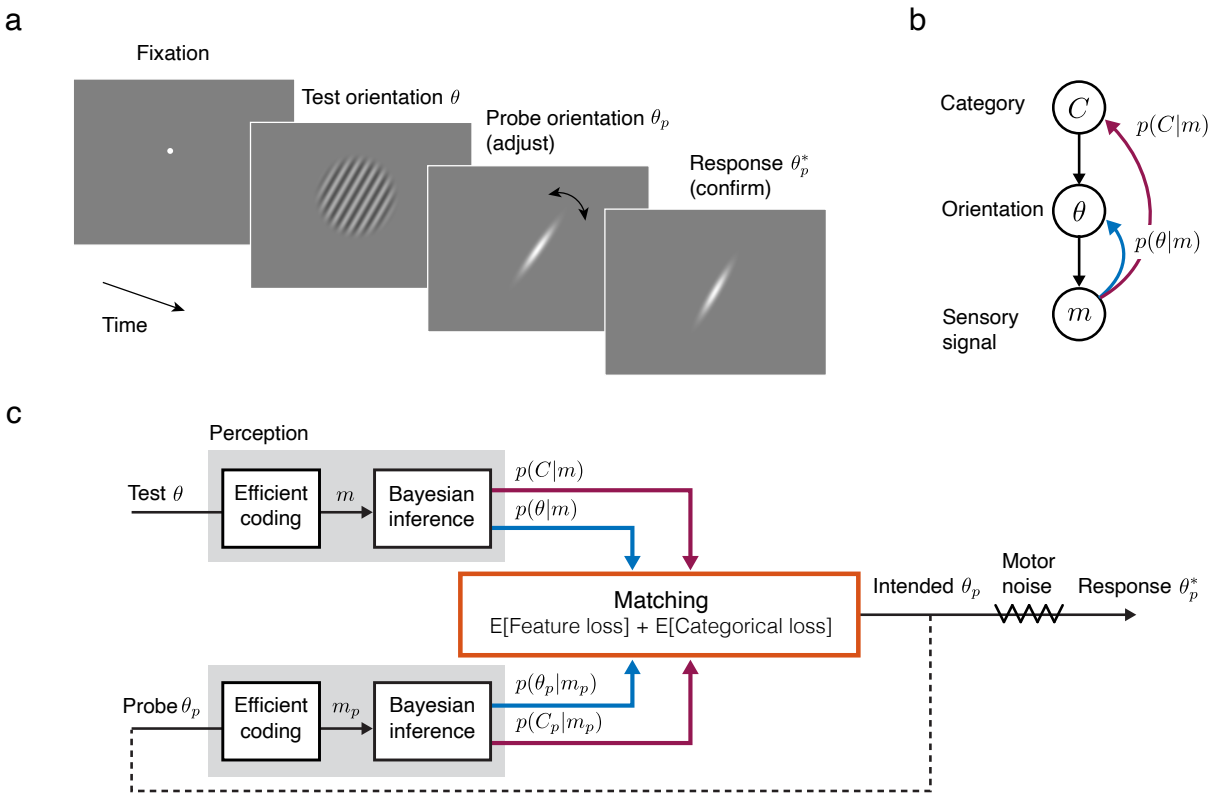

Figure 3.1: Holistic perceptual matching. (a) Typical psychophysical matching task to characterize visual orientation perception. Subjects are presented with a test stimulus with orientation θ. Then they are asked to adjust the orientation θ_p of a probe stimulus such that it best matches the perceived test orientation. Subjects typically press a button to confirm their choices, at which time their response θ_p^* is recorded. (b) Graphical model representing the hierarchical generative process by which a stimulus with orientation θ and a higher-level, categorical identity C (cardinal/oblique) generates a noisy sensory signal m. Our key assumption is that perceptual inference is holistic and consists of computing both the posteriors over orientation $p(\theta|m)$ (blue arrow) and category identity $p(C|m)$ (purple arrow). (c) Holistic matching model. The model assumes that both the test θ and the probe θ_p orientation are efficiently encoded according to the orientation prior $p(\theta)$ (Wei and Stocker, 2015), resulting in sensory measurements m and m_p, respectively. As illustrated in (b), perceptual (Bayesian) inference results in posteriors $p(\theta|m)$, $p(C|m)$ and $p(\theta_p|m_p)$, $p(C_p|m_p)$, respectively. By minimizing a combined objective that quantifies mismatch at both the feature and category level (Eq. (3.1)), the model computes the probe orientation that optimally matches the test orientation. Note, that a non-holistic version of the proposed model (i.e., removing the categorical inference pathway - purple arrows) is equivalent to the efficient Bayesian estimator when the probe stimulus is noise-free.

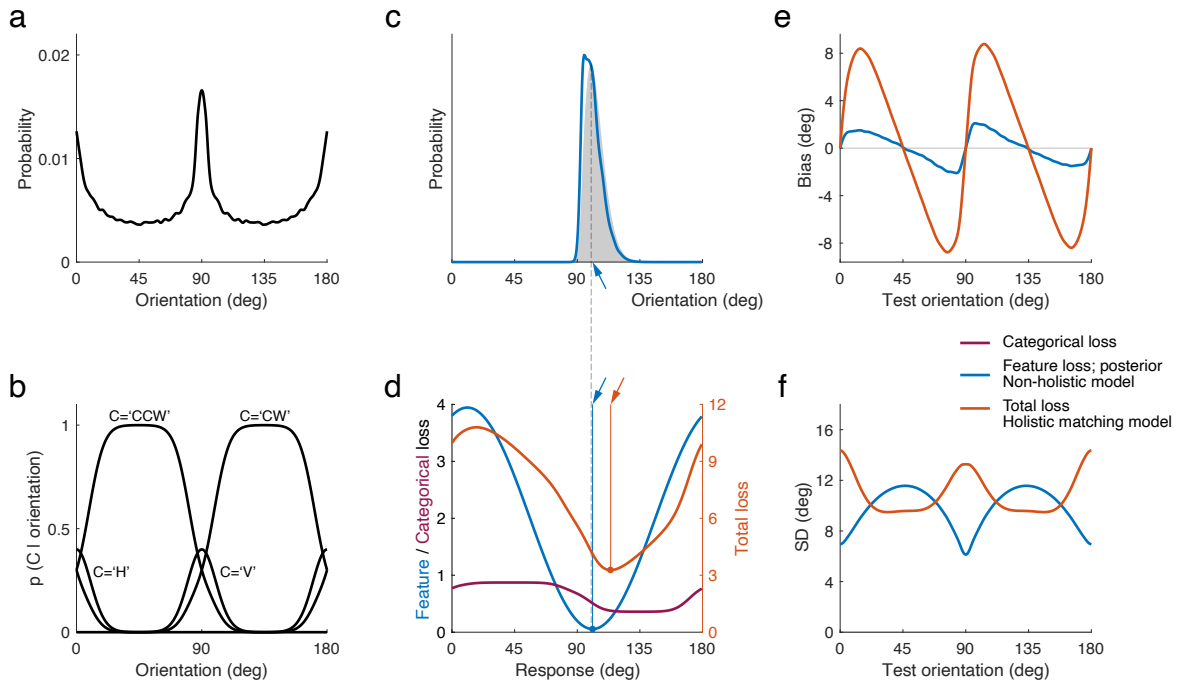

Figure 3.2: Model simulations for matching task with noiseless probe (typical condition). (a) Prior distribution used for model simulations and fits throughout this paper. It reflects the average statistics of local visual orientations in natural indoor and outdoor scenes measured by Coppola et al. (1998). (b) Category structure assumed by the holistic matching model. (c) Likelihood (shaded area) and posterior (blue curve) of test orientation for a given sensory measurement m (dashed line). The blue arrow marks the mean of the posterior, which is the optimal estimate predicted by the efficient Bayesian estimator. (d) Expected feature (blue), categorical (purple), and total loss (orange) given m (dashed line), and the optimal match predicted by the efficient Bayesian estimator (blue arrow) and the holistic matching model (orange arrow). (e) Bias pattern and (f) standard deviation in probe responses predicted by the two models.
Supplementary Figure 3.17. Outdoor and indoor natural scene statistics.

where $w > 0$ is the relative contribution of the categorical mismatch. In the current implementation of the model, we define mismatch at the orientation level L_θ as the cosine difference between the test and the probe orientation, while the mismatch at the categorical level L_c is assumed to come with a constant penalty if test and probe belong to different orientation categories and zero otherwise. Finally, we assume that subjects' responses θ_p^* represent noisy samples of the optimally matching probe orientation θ_p due to additive, constant motor noise (*Methods*).

Simulations shown in Fig. 3.2 illustrate how and why the predictions of the holistic matching model differ from those of the efficient Bayesian estimator. We constrain the prior distribution of visual

orientation $p(\theta)$ to reflect the orientation statistics in natural scenes. Previous studies have shown that these statistics are relatively robust with regard to the specific methods they were measured with and the image content of the natural scenes they were computed for, showing characteristic peaks at both cardinal orientations (Coppola et al., 1998; Girshick et al., 2011; Wang et al., 2016). However, outdoor scenes containing fewer man-made objects typically show less pronounced peaks at the cardinals compared to indoor scenes (Coppola et al., 1998; Straub and Rothkopf, 2021) (see Supplementary Fig. 3.17). We computed the average of previously measured distributions across both indoor and outdoor scenes (Coppola et al., 1998), and used this distribution (shown in Fig. 3.2a) as the fixed orientation prior $p(\theta)$ for all simulations and fits presented in the paper. Furthermore, we consider four natural categories for orientation: vertical ('V'), horizontal ('H'), clockwise('CW') or counterclockwise ('CCW') relative to vertical (Fig. 3.2b). We assume that there is some uncertainty associated with the categorical representation expressed in overlapping categorical distributions as well as in noisy centers of the two cardinal categories that may vary trial by trial. Note that assuming a categorical structure that only distinguishes two categories ('CW' and 'CCW' relative to the vertical meridian) does not significantly change the model behavior (see Supplementary Figs. 3.18 and 3.20).

Efficient encoding predicts likelihood functions that have long tails away from the nearest cardinal orientation for the assumed orientation prior (Fig. 3.2a). Figure 3.2c shows the likelihood function and the posterior distribution for a sensory measurement m of the test stimulus close to vertical (90 deg). Although the posterior is shifted towards vertical due to the prior, it inherits the long tail from the likelihood function. The long-tailed posterior distribution in combination with the loss function L_θ is ultimately responsible for the predicted repulsive bias away from vertical of the efficient Bayesian estimator (Wei and Stocker, 2015). Figure 3.2d illustrates this by plotting the feature loss L_θ for the same measurement m where the matching percept represents the point of minimal loss (arrow). This point of minimal loss, however, changes when considering the combined loss L_{tot} of the holistic matching model. Because of category uncertainty, the minimum of the category loss L_c does not coincide with the minimum of the feature loss L_θ. Rather, it is shifted towards the center of the most probable category of the test stimulus resulting in larger repulsive biases (Fig. 3.2e).

The precise bias pattern of the holistic matching model depends on the relative levels of feature (test and probe stimuli) and category uncertainty. Compared to the efficient Bayesian observer, the holistic matching model predicts larger repulsive biases for the same level of stimulus uncertainty.

3.3.2. Matching experiment with noiseless probe stimulus

We first validated the holistic matching model in the typical, noiseless probe condition using an extensive dataset from a previous orientation matching experiment (De Gardelle et al., 2010). In the experiment, subjects were asked to estimate the orientation of a briefly presented test stimulus by adjusting the orientation of a probe stimulus. Sensory noise of the test stimulus was modulated by varying the presentation duration (*Methods*). Figure 3.3 shows the full error distributions of the combined human subject data for each of the four presentation durations. The distributions exhibit the characteristic repulsive bias away from cardinal orientations, and show no apparent asymmetry between the two cardinal orientations. Bias and variability increase with decreasing stimulus presentation duration, which is a fundamental characteristic of Bayesian perception (Stocker and Simoncelli, 2006).

We fit the holistic matching model as well as its non-holistic variant (i.e., the efficient Bayesian estimator) to the response distribution data. For comparison, we also included a standard Bayesian estimator with homogeneous sensory encoding (*Methods*). We assumed the probe stimulus to be noise-free as it consisted of a Gabor patch with one visible strip that was continuously present until subjects confirmed their choice. Note, that all models use the same formulation of the feature loss L_θ and the natural orientation prior shown in Fig. 3.2a. The holistic matching model fully captures the entire shape of the error distributions for all noise conditions, which is not the case for the two Bayesian estimation models (Fig. 3.3). Their predicted error distributions are mostly centered around zero and show much larger variability for oblique than for cardinal orientations. Furthermore, as expected, the standard Bayesian estimator predicts attractive bias near cardinal orientations. In general, the estimation models exhibit distribution patterns that are substantially different from the data. The difference is also evident when comparing bias and standard deviation of the data with the predictions of the models (Fig. 3.4). Human subjects exhibit repulsive bias

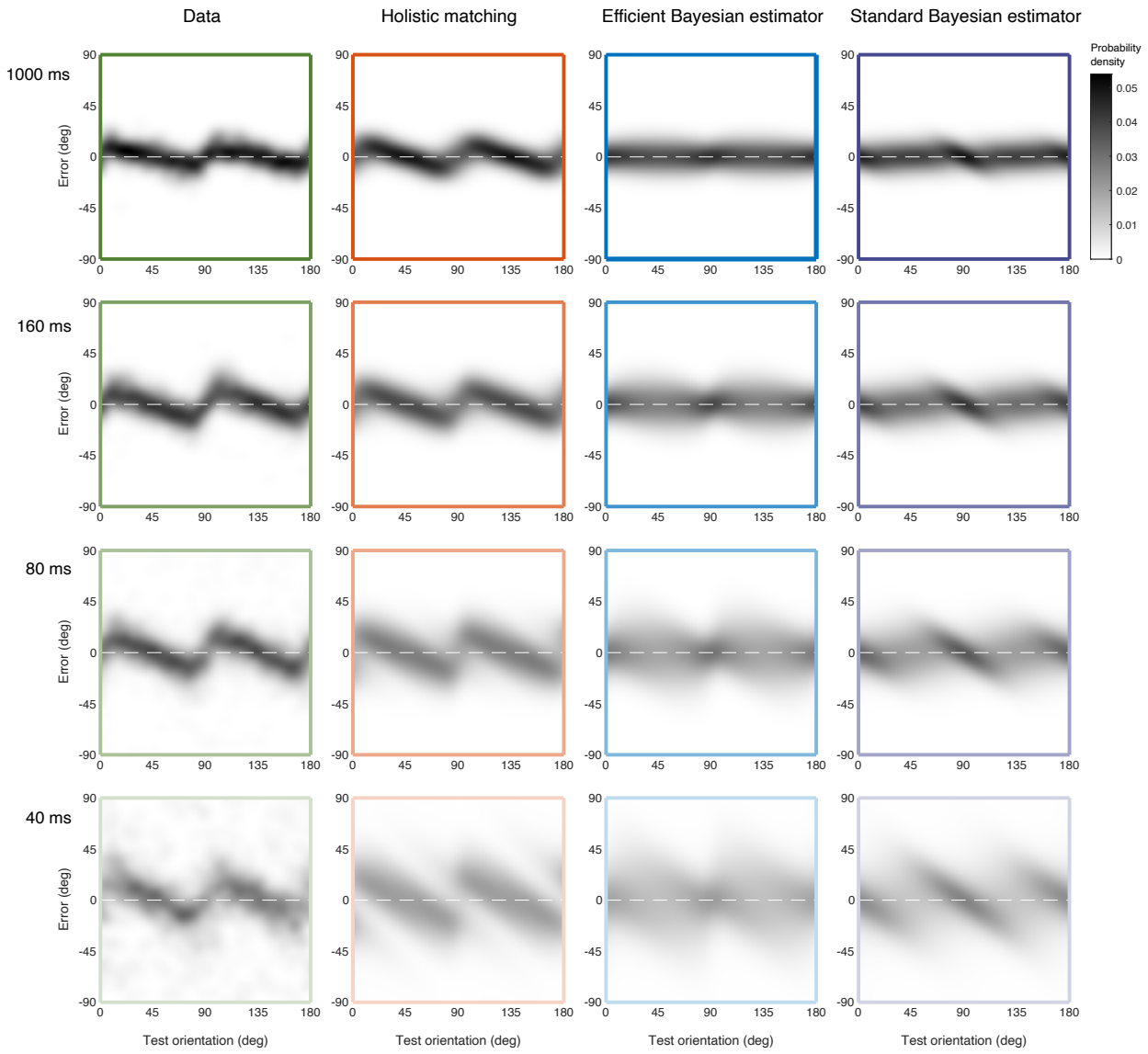

Figure 3.3: Data and model fits for matching task with noiseless probe. Shown are the error distributions of the matching responses for different presentation durations of the test stimulus (rows). Columns show the data (De Gardelle et al., 2010) and the corresponding best-fit model predictions, respectively. Data distributions show clear repulsive biases away from the cardinal orientations. Bias and variability increase with decreasing presentation duration. The overall pattern of the distribution is well captured by the holistic matching model across all conditions. While the efficient Bayesian estimator correctly predicts repulsive biases, the overall shape of the predicted error distributions does not match the data. The standard Bayesian estimator (homogeneous encoding) predicts attractive biases. See *Methods* for details about the data and the models.
Supplementary Table 3.1. Fit parameter values of the holistic matching model.
Supplementary Figure 3.15. Fit category structure of the holistic matching model.
Supplementary Figure 3.18. Model fit assuming two orientation categories.

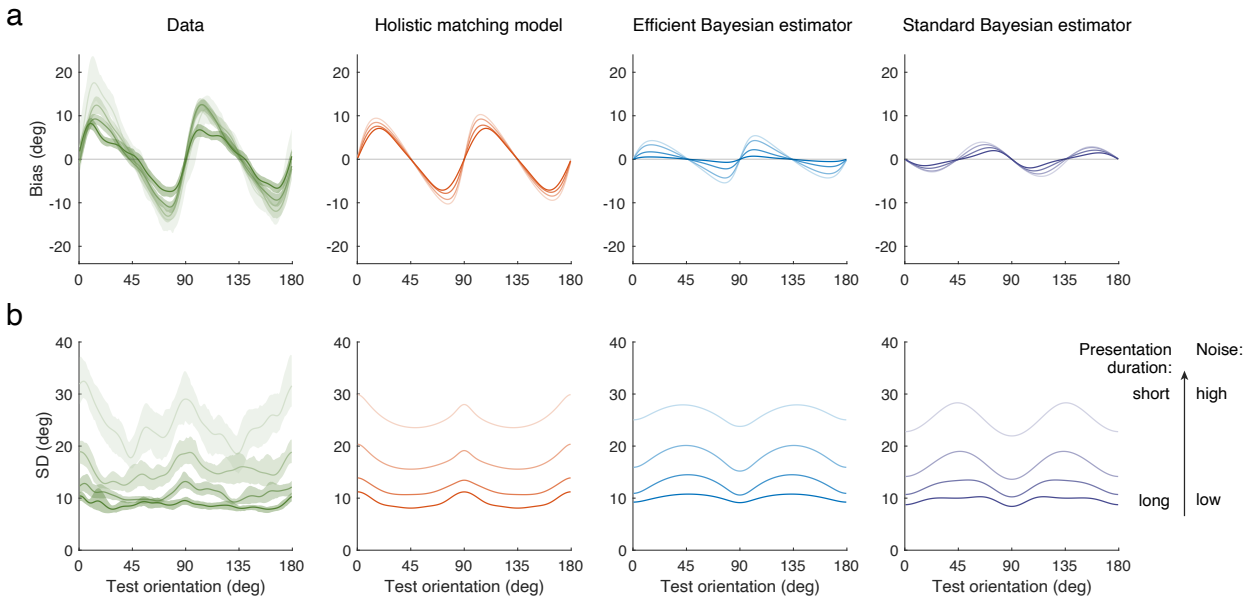

Figure 3.4: Data and model fits for matching task with noiseless probe: bias and standard deviation. (a) Subjects exhibit increasing repulsive bias with decreasing presentation duration. The holistic matching model fits both the pattern and the magnitude of the bias. The bias magnitude predicted by the efficient Bayesian model, however, is too small compared to the data. In contrast, the standard Bayesian model predicts attractive bias. (b) Subjects' variability is higher around cardinal orientations, which is well captured by the holistic matching model. The standard and the efficient Bayesian model predict the opposite. Data is re-analyzed from De Gardelle et al. (2010). Shaded areas represent 95% confidence intervals from 100 bootstrap runs.

Supplementary Figure 3.17. Predictions of the holistic matching model using either the outdoor or the indoor scene statistics as the prior $p(\theta)$.

Supplementary Figure 3.18. Model fit assuming two orientation categories.

away from cardinal orientations with increasing amplitude for increasing sensory noise (i.e., shorter presentation duration), while the standard deviation of their response distribution is higher at cardinal compared to oblique orientations. Predictions of the standard Bayesian estimator are the exact opposite. Although the efficient Bayesian estimator qualitatively captures the repulsive bias pattern in the data and its dependency on sensory noise, the predicted overall bias magnitudes are too small. Like the standard Bayesian estimator, it also incorrectly predicts higher standard deviation at oblique compared to cardinal orientations. In contrast, the holistic matching model predicts bias and standard deviations that not only qualitatively but also quantitatively match the data. Note, that the predicted, higher standard deviation at cardinal orientations is caused by the fact that for test orientations close to the categorical boundaries, small differences in sensory measurements across trials can lead to large differences in probe responses due to the categorical matching process. This additional, categorical bias offsets the increased sensory accuracy at cardinal orientations due to efficient coding.

We used cross-validation to quantitatively compare the performance of the different models. Cross-validation intrinsically corrects for differences in model complexity (i.e. number of parameters); overly complex models that overfit the training data score typically low in accounting for the test data. It favors those model that have just "the right" level of complexity to account for the data. We included an "omniscient" observer model in this comparison, which is an empirical model that directly transforms the training data distribution into a prediction probability of the error distribution using kernel density estimation (see *Methods*). The omniscient observer serves as reference and indicates the best possible statistical prediction of the test set given the training set. As shown in Fig. 3.5, the holistic matching model predicts the data substantially better than the efficient and standard Bayesian model, and its performance is almost at the level of the omniscient model. Cross-validation demonstrates that the holistic matching model provides an excellent account of orientation estimation behavior with a model complexity that does not lead to over-fitting of the data.

Figure 3.6 shows validation against data from another recent study, investigating the differences

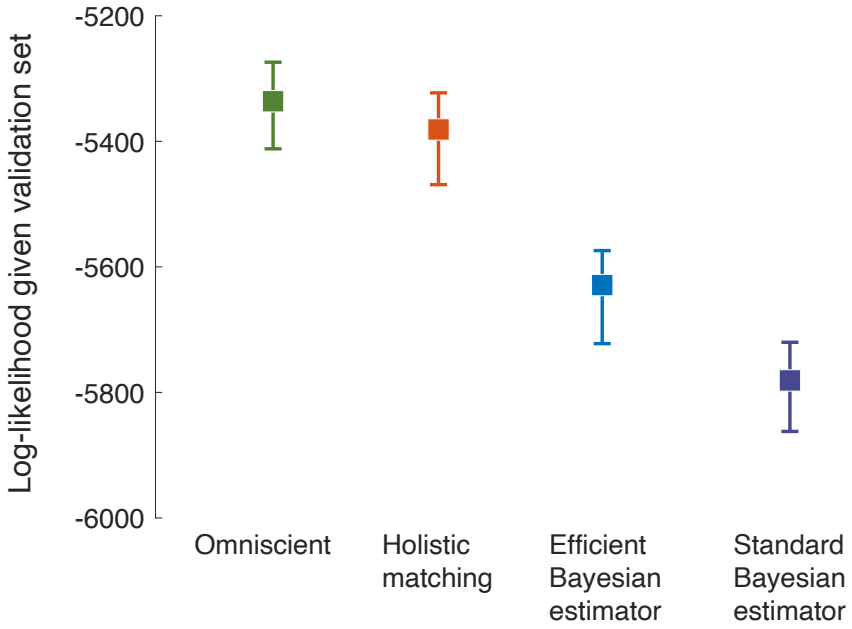

Figure 3.5: Cross-validation. Log-likelihood values of the model fit to the training set (80% of the data; randomly sampled), given the validation set (remaining 20% of the data). Squares represent the median and error bars indicate 95% confidence intervals over 100 repetitions. The holistic matching model performs significantly better than the efficient and the standard Bayesian observer model. The "omniscient" model is an empirical model that uses the data distribution in the training set as predictor of the validation data using optimal kernel density estimation (*Methods*). The error bar of the holistic matching model and the omniscient model largely overlap, indicating that the holistic matching model explains the data as good as statistically possible.
Supplementary Figure 3.19. Cross-validation of the omniscient model with different kernel sizes.

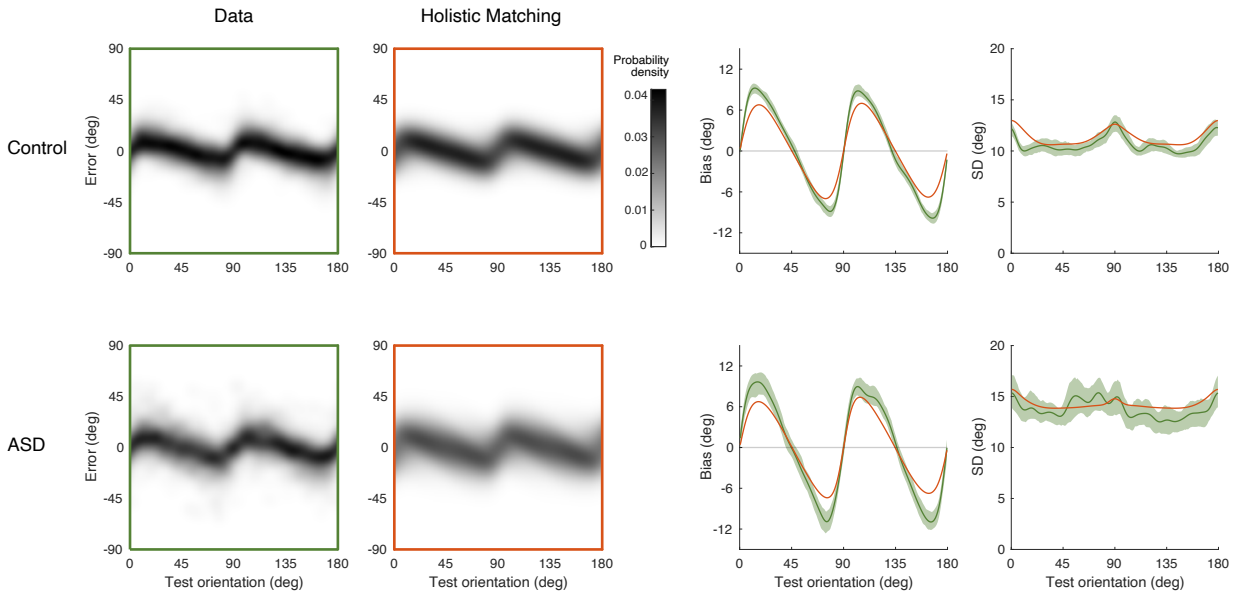

Figure 3.6: Data and model fits for matching task with noiseless probe in neurotypical and autistic (ASD) subjects. Data are re-analyzed from Noel et al. (2021) and are obtained from the control experiment (no feedback), for which both subject groups have been identified to have similar prior expectations that match the assumed prior distribution $p(\theta)$. Shaded areas represent 95% confidence intervals from 100 bootstrap runs.

Supplementary Table 3.3. Fit parameter values.

in orientation perception between individuals with autism spectrum disorder (ASD) and a control group (Noel et al., 2021). The study used a similar experimental design as in De Gardelle et al. (2010), reporting similar behavior signatures of holistic hierarchical inference such as the shape of the error distribution, the relatively large bias magnitude, and the higher standard deviation at cardinal orientations compared to oblique orientations. As shown in Fig. 3.6, the model can well account for the data from both subjects groups. Interestingly, a comparison of the fit model parameters indicates that the difference between the two groups is mainly limited to differences in sensory and motor noise (ASD: higher sensory and lower motor noise), while the expectations about the categorical structure seem identical (see Supplementary Table 3.3).

3.3.3. Efficient sensory encoding

Is the categorical inference component of the holistic matching model sufficient to explain the repulsive bias pattern? To answer this question, we compared the predictions of the fit holistic matching model with and without efficient sensory coding. As illustrated in Fig. 3.7, without efficient sensory encoding (i.e., assuming uniform sensory accuracy) the model predicts decreasing bias magnitudes with increasing stimulus presentation time, which is opposite to the pattern seen in the data (Fig. 3.4a).

Bias is modulated by sensory noise via three different processes of the model: sensory encoding, inference at the feature level, and inference at the categorical level. As sensory noise increases, the posterior distribution of the test orientation is more attracted to the peak of the prior distribution. At the same time, the difference in category posterior probability of the test stimulus decreases, leading to a flatter categorical loss curve (see Fig. 3.2d). Both effects lead to less repulsive bias as the sensory noise increases, which is the outcome shown in Figure 3.7 (dashed line). Efficient coding introduces repulsive biases by skewing the likelihood function away from the peak of the prior at cardinal orientations (Wei and Stocker, 2015). However, larger sensory noise leads to more skewness in the likelihood function and therefore larger repulsive biases. Thus, efficient coding is the only component of the holistic matching model that causes larger repulsive biases with higher sensory noise. As such, efficient sensory encoding is an indispensable assumption of the holistic matching

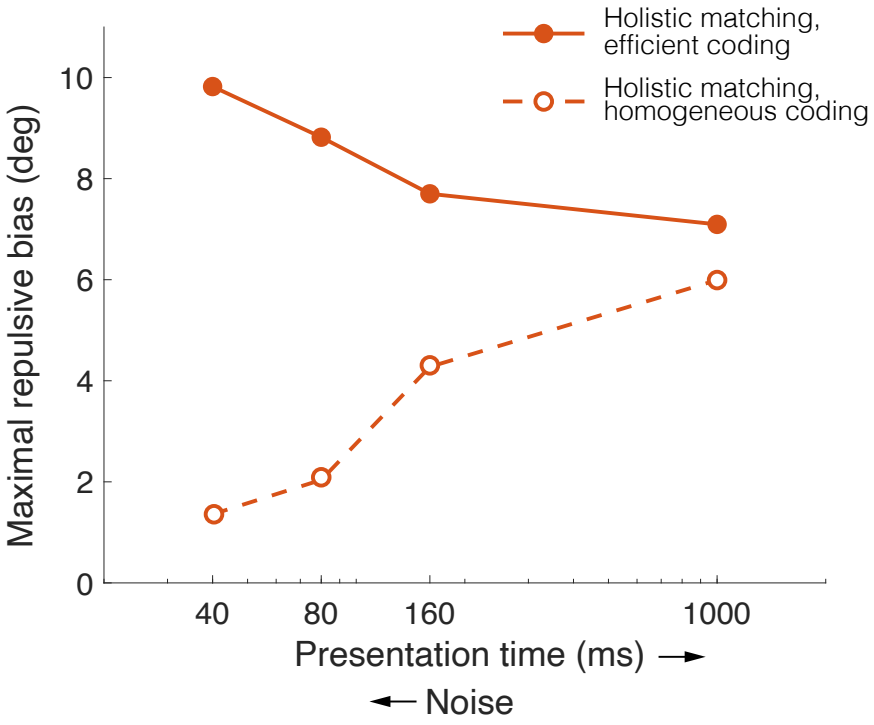

Figure 3.7: Effect of efficient sensory encoding. Maximum bias predicted by the holistic matching model with and without the efficient coding constraint. With efficient coding, the holistic matching model predicts decreasing bias magnitudes with increasing presentation times (i.e., decreasing sensory noise) consistent with the data. With homogeneous coding and all else equal, however, it predicts the opposite pattern.

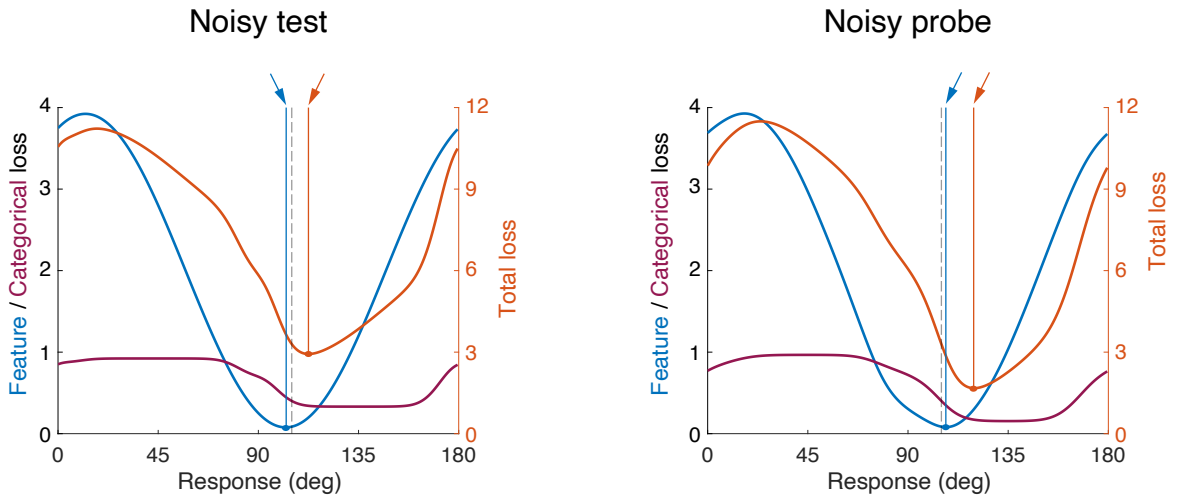

Figure 3.8: Model predictions for interchanging test and probe stimuli. When the test and the probe stimulus are interchanged in the matching experiment and thus stimulus uncertainties are reversed, the optimal response of the non-holistic model (blue arrow) flips to the other side of the measurement of the test (gray dashed line). The optimal response according to the holistic matching model (orange arrow), however, remains on the same side because of the influence of the categorical loss. The illustration is shown for a single pair of sensory measures.

model in order to accurately explain the data.

3.3.4. Matching experiment with noisy probe stimulus

In most perceptual matching experiments the probe stimulus is unambiguous and noiseless, and thus its percept can be considered veridical. However, the holistic matching model explicitly models the perception of the probe orientation and thus can make predictions for more general experimental conditions (Fig. 3.1b). Here, we specifically consider the case where stimulus uncertainties in the test and probe stimulus are reversed.

Any model that compares the two stimuli at the feature level will predict a reversal of the bias pattern when reversing the roles of the test and probe stimulus. However, the holistic matching model makes a qualitatively different prediction. Because the probe stimulus is adjusted to match the test also at the category level, adjusting the probe orientation towards the center of the most probably category of the test stimulus always reduces the expected categorical loss L_c. This leads to a stable repulsive bias pattern whether the test and probe stimulus are interchanged or not.

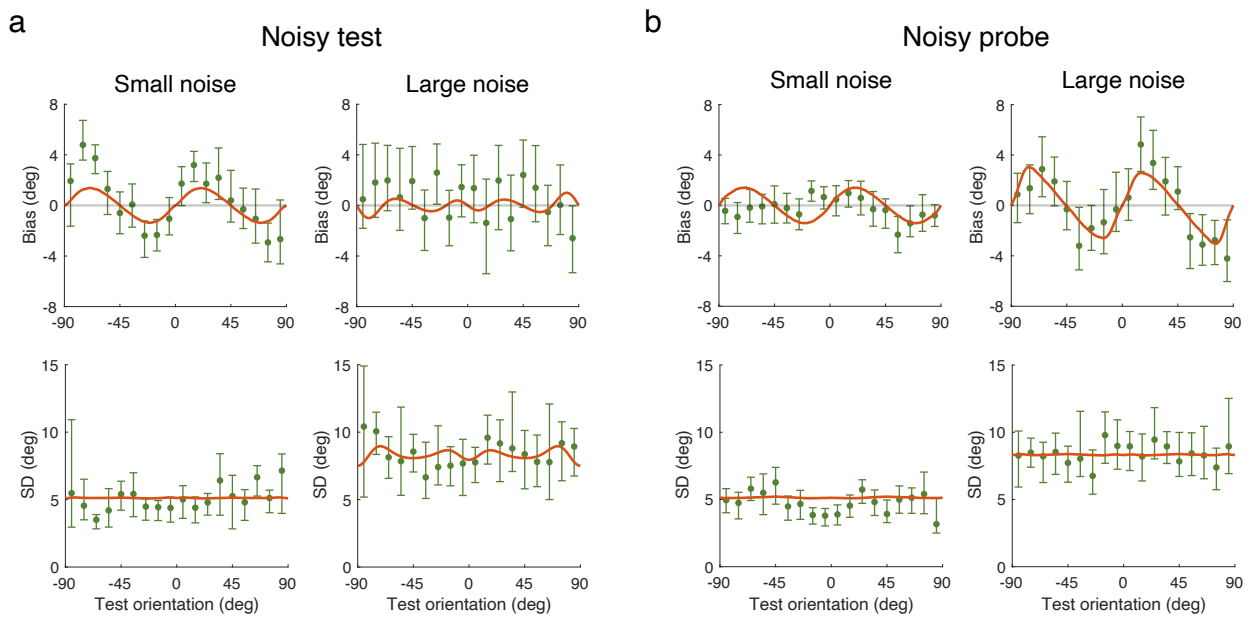

Figure 3.9: Data and model fits for interchanging test and probe stimuli for two different stimulus noise levels (combined subject). (a) Bias and standard deviation of subjects' matching response data when the test stimulus is noisy and the probe is noiseless. (b) Same as (a) but probe and test stimuli are interchanged. The sign of the biases is not inverted. Biases are always repulsive or close to zero, depending on the level of stimulus noise. Solid lines represent the joint fit of the holistic matching model across all conditions. Data is re-analyzed from Tomassini et al. (2010). Error bars represent 95% confidence intervals from 100 bootstrap samples of the data.
Supplementary Table 3.1. Fit parameter values of the holistic matching model.
Supplementary Figure 3.15. Fit categories of the holistic matching model.
Supplementary Figure 3.18. Model fit assuming two orientation categories.

Figure 3.8 illustrates this qualitative difference between the model predictions.

Tomassini et al. (2010) performed an orientation matching experiment where test and probe stimuli were interchanged. During the first half of the experiment, participants were shown an array of Gabor patches (noisy test) and were asked to adjust the orientation indicated by two dots (noiseless probe) to estimate the mean orientation of the Gabor patches. During the second half of the experiment, the roles of the stimuli were reversed; subjects were asked to rotate the array of Gabor patches (noisy probe) until the array orientation matched the orientation indicated by the two dots (noiseless test). Biases and standard deviations of subjects' matching responses are shown in Fig. 3.9. As predicted by the holistic matching model, the biases are indeed repulsive under both conditions. We preformed a joint model fit to the data across all conditions. The model well accounts for the

observed repulsive biases in the small stimulus noise condition when the test stimulus is noisy and in the large stimulus noise condition when the probe stimulus is noisy. Likewise, when the test stimulus is noisy the predicted bias is close to zero for large stimulus noise but does not have a clear repulsive or attractive pattern, which matches the data (Fig. 3.9a). When the probe stimulus is noisy but the test stimulus is not, the bias is smaller in the small noise condition than in the large noise condition, which is the same pattern as in the data (Fig. 3.9b). The standard deviation predicted by the model is for most part uniform with a magnitude that again is consistent with the data. The matching experiment by Tomassini et al. (2010) revealed human matching behavior that is well accounted for by the proposed holistic matching model, yet is difficult to even qualitatively reconcile with any non-holistic estimation model.

3.3.5. Validation of model prediction with same-noise test and probe stimuli

Another prediction made by our model that is qualitatively different from other estimator models is when the test and the probe stimuli are the same, or have the same noise profile. Any model that compares the two stimuli at the feature level will predict zero bias, because they are essentially the same stimuli. However, because the matching process in the holistic matching model also tries to reduce the expected categorical loss L_c, again there will be a stable repulsive bias pattern even when the test and the probe have the same noise (Fig 3.10d).

In order to validate the prediction, we ran a new orientation matching experiment with same-noise test and probe stimuli (Fig 3.10a, see *Methods* for details). Both the test and the probe stimuli are filtered white noise patches. Stimulus noise was manipulated by changing the variance of the orientation filter applied to the white noise. Both the test and the probe can have low (L) or high (H) stimulus noise, resulting in 4 conditions in total, indicated by "test-probe" noise level pairs: L-L, L-H, H-L and H-H. We adopted response-terminated display of the test stimulus, so the test and the probe stimuli both had small and the same sensory noise. Therefore, the test and the probe stimuli have the same noise in the L-L and H-H noise conditions, and the test and the probe are switched relative to each other in the L-H and H-L noise conditions similar to Tomassini et al. (2010). A non-holistic model that matches the orientation estimate of the test and the probe predicts no bias in

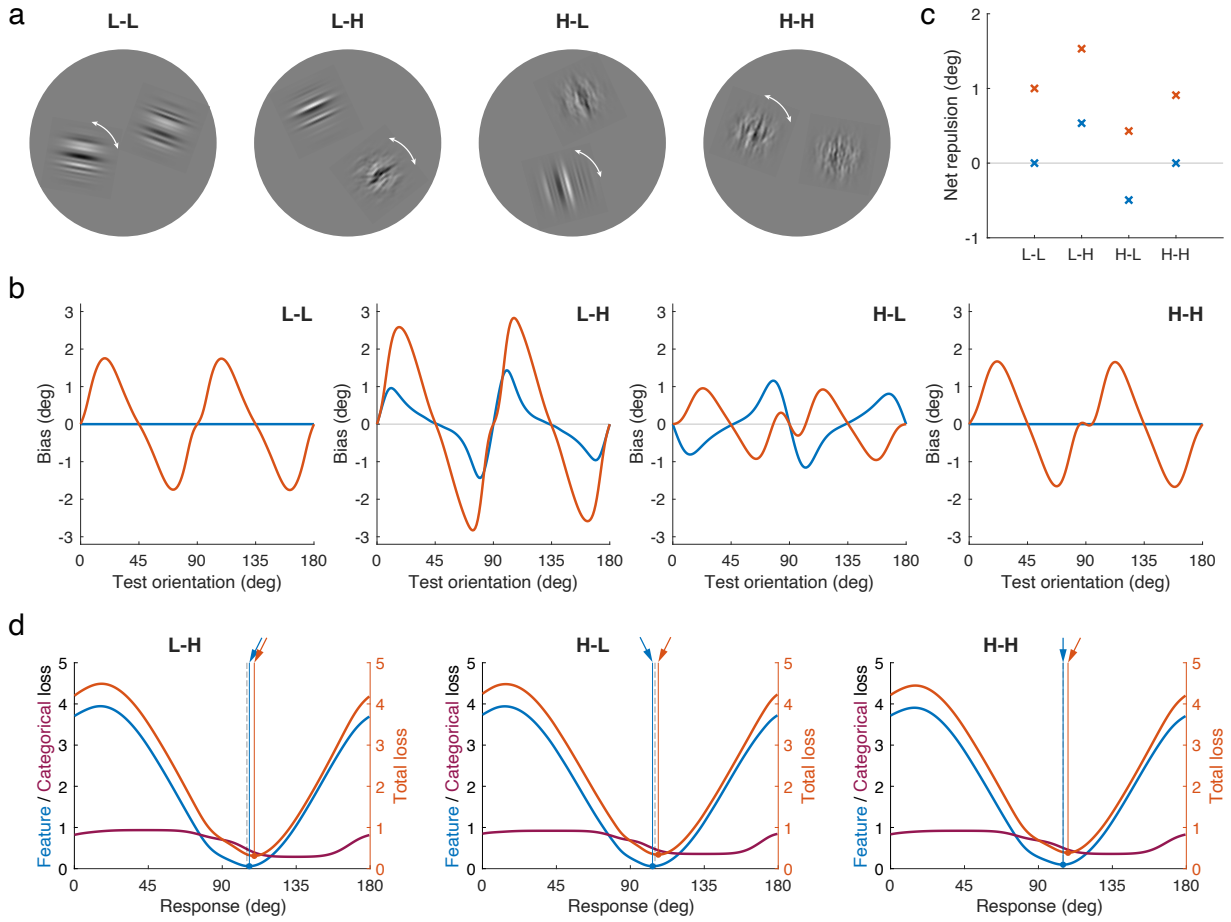

Figure 3.10: Experiment design and model predictions for matching test and probe with stimulus noise. (a) Experiment design. For illustration purpose, stimuli are not drawn to scale, and the arrows indicating the probe are not shown in the experiment. See *Methods* for details. (b) The non-holistic model (blue) predicts no bias in the two same-noise conditions, and reversed bias in the L-H and H-L conditions. The holistic matching model (orange) predicts repulsive bias in the same noise conditions, and the bias in the H-L condition may not be reversed relative to L-H and still be repulsive with a large enough categorical effect. (c) Net repulsion from cardinal orientations. The categorical effect in the holistic matching model leads to an upward shift in all conditions relative to the non-holistic model. (d) When the stimulus uncertainties in the test and the probe are reversed (L-H and H-L), the predictions are similar to those in Figure 3.8. When the stimulus uncertainties in the test and the probe are the same (H-H), the optimal response of the non-holistic model (blue arrow) equals the measurement of the test, while the optimal response according to the holistic matching model (orange arrow) is repulsed away from the nearest cardinal orientation. The illustration is shown for a single pair of sensory measures.

the two same-noise conditions, and reversed bias in the L-H and H-L conditions, while the holistic matching model predicts repulsive bias in the same noise conditions, and uninverted bias in the H-L relative to the L-H condition with a large enough categorical weight (Fig 3.10b). We further calculate a net repulsion metric to quantify the amount of repulsion from cardinal orientations, which is the average of the bias for test orientation from 0 - 45 deg and 90 - 135 deg and the negative bias for test orientation from 45 - 90 deg and 135 - 180 deg (Fig 3.10c). The non-holistic model predicts no net repulsion in the same noise conditions, and opposite net repulsion in the flipped noise conditions, while net repulsion predicted by the holistic matching model shifts upwards relative to the non-holistic model in all four conditions.

When the probe has higher noise than the test stimulus (L-H), subjects consistently showed a repulsive bias (Fig. 3.11); in the opposite noise condition (H-L), the bias is close to zero when combining all subjects. When the test and the probe have the same noise (L-L and H-H), although there is individual variability (see Supplementary Fig. 3.21, 3.22, 3.23), the average and the combined subjects exhibited significant net repulsion, and most subjects had repulsive bias in at least one of the same noise conditions. The net repulsion in the same noise conditions can only be explained by the holistic matching model and not the non-holistic matching model or estimator models. The fit by the holistic matching model captures the repulsive bias in the L-H condition and the same-noise conditions very well. The standard deviation is lower at cardinal orientations, which is also well captured by the model fit.

This experiment replicated the uninverted bias with switched test and probe stimuli showed by Tomassini et al. (2010). It further validated the holistic matching model by demonstrating a net repulsive bias when the test and the probe stimuli have the same noise, which cannot be accounted for by any non-holistic estimation model.

3.3.6. When to expect behavioral signatures of holistic inference

As the above data analysis and model comparison have shown, the holistic nature of the proposed matching model is responsible for increased (repulsive) response bias towards the category centers, different patterns in response variability, and the surprising finding that the bias pattern is not

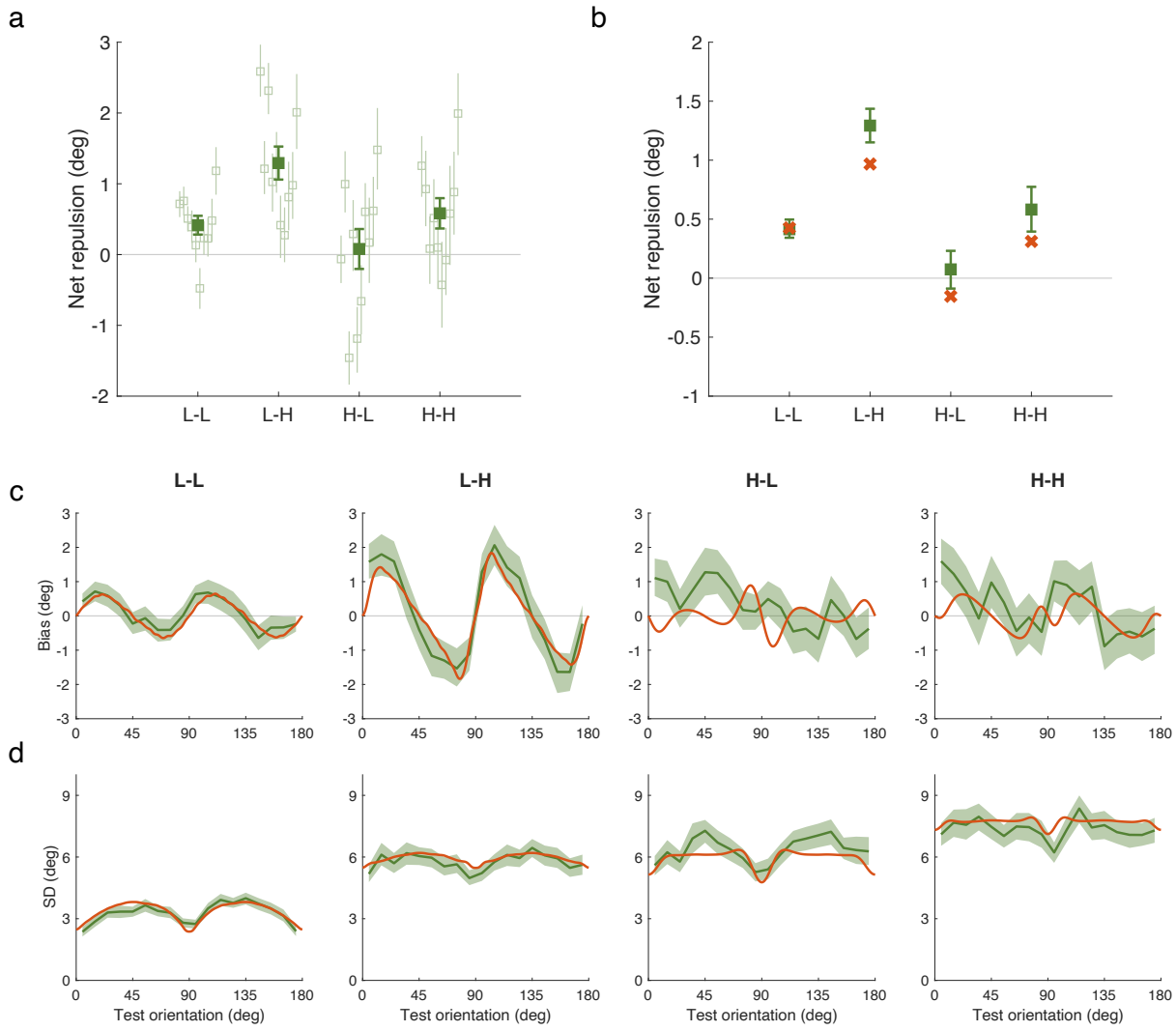

Figure 3.11: Experiment results and model fit. (a) Net repulsion of individual subjects (light) and averaged across subjects (dark). Although there is individual variability, most subjects and all subjects on average have repulsive bias in the same-noise conditions. (b-d) Data (green) and model fit (orange) of the combined subject. The same-noise conditions have repulsive bias; the L-H condition has larger repulsive bias and the H-L condition has close to zero repulsion, which is captured by the model fit. For average net repulsion across subjects, error bars represent SEM. For individual and combined subjects' data, error bars represent 95% confidence intervals from 1000 bootstrap samples of the data.

Supplementary Table 3.4. Fit parameter values of the holistic matching model.

Supplementary Figure 3.16. Fit categories of the holistic matching model.

Supplementary Figure 3.21, 3.22, 3.23. Data from individual subjects.

inverted when interchanging test and probe stimuli. However, the holistic matching model formally subsumes the efficient Bayesian estimator. Thus, we can clearly predict when we expect subjects' behavior to show signatures of holistic inference, and when not because the models are equivalent.

The impact of the category level inference on behavior is determined by how gradually the expected categorical loss L_c decreases towards the center of the most probable category of the test stimulus (i.e., the slope of the categorical loss). Only if there is little category uncertainty and there is no uncertainty associated with the probe stimulus, then L_c approaches a step function with zero slope anywhere except at the boundary (Fig. 3.12b, top left). In this case, the expected categorical loss is constant and independent of the probe orientation, and thus the expected response will be identical with or without holistic inference (Fig. 3.12c, top). Otherwise, if there is category uncertainty (Fig. 3.12a, right) the probability of the probe category gradually changes with probe orientation, and thus the expected categorical loss also gradually decreases towards the center of the most likely category resulting in increased repulsive biases compared to the non-hierarchical model (Fig 3.12b and c, top). Likewise, if there is uncertainty associated with the probe stimulus then the expected category loss also gradually decreases for probe orientations towards the center of the most probably test category, which again results in increased repulsive biases (Fig. 3.12b and c, bottom).

We predict that signatures of holistic hierarchical inference are, to various degree, present in most psychophysical matching data. However, data from experiments that are designed such that they minimize subjects' uncertainty about the category boundaries and use probe stimuli without uncertainty are expected to be equally well accounted for with the efficient Bayesian estimator.

3.3.7. Generalization to color perception

To test the generality of the model we extend validation to color perception, widely considered to be influenced by color categories (Witzel and Gegenfurtner, 2013; Cibelli et al., 2016; Hardman et al., 2017). Specifically, we considered the data from a recent color study (Bae et al., 2015). In the study, the authors conducted two color matching experiments where subjects were asked to estimate the color of either a previously (delayed condition) or a simultaneously presented test color by selecting the matching color on a color wheel.

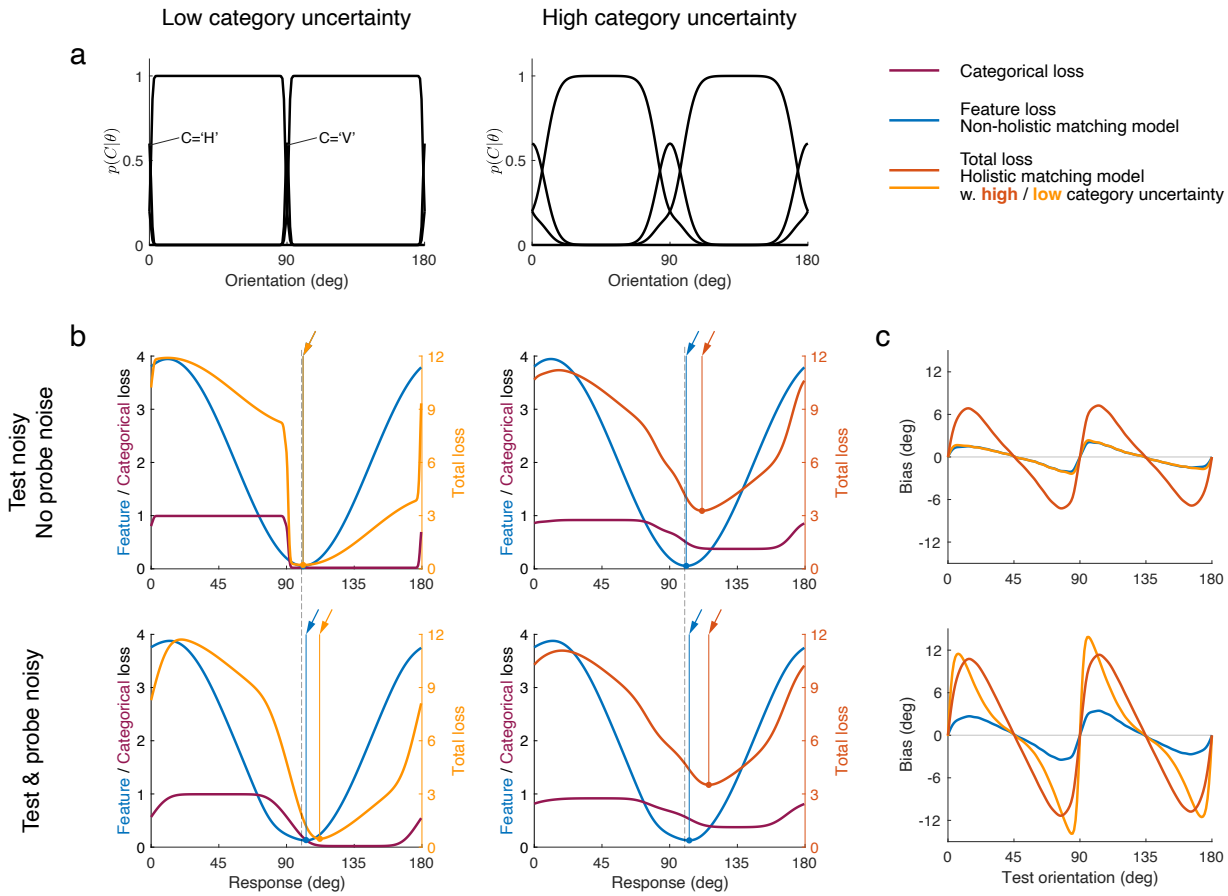

Figure 3.12: Model simulation with and without category uncertainty and with and without stimulus uncertainty in the probe stimulus. (a) Orientation categories $p(C|\theta)$ when there is low (left) or high (right) category uncertainty. (b) Expected loss for a certain measurement of test orientation (dashed line) when there is low (left column) or high category uncertainty (right column) and when the probe stimulus is (bottom row) or is not prone to stimulus uncertainty (top row). Arrows mark the optimal responses predicted by the non-holistic matching model (blue) or the holistic matching model (orange). (c) Bias predicted by the non-holistic (blue) and the holistic matching model when there is low (orange) or high (dark orange) category uncertainty, and when the probe stimulus is (bottom) or is not (top) prone to stimulus uncertainty.

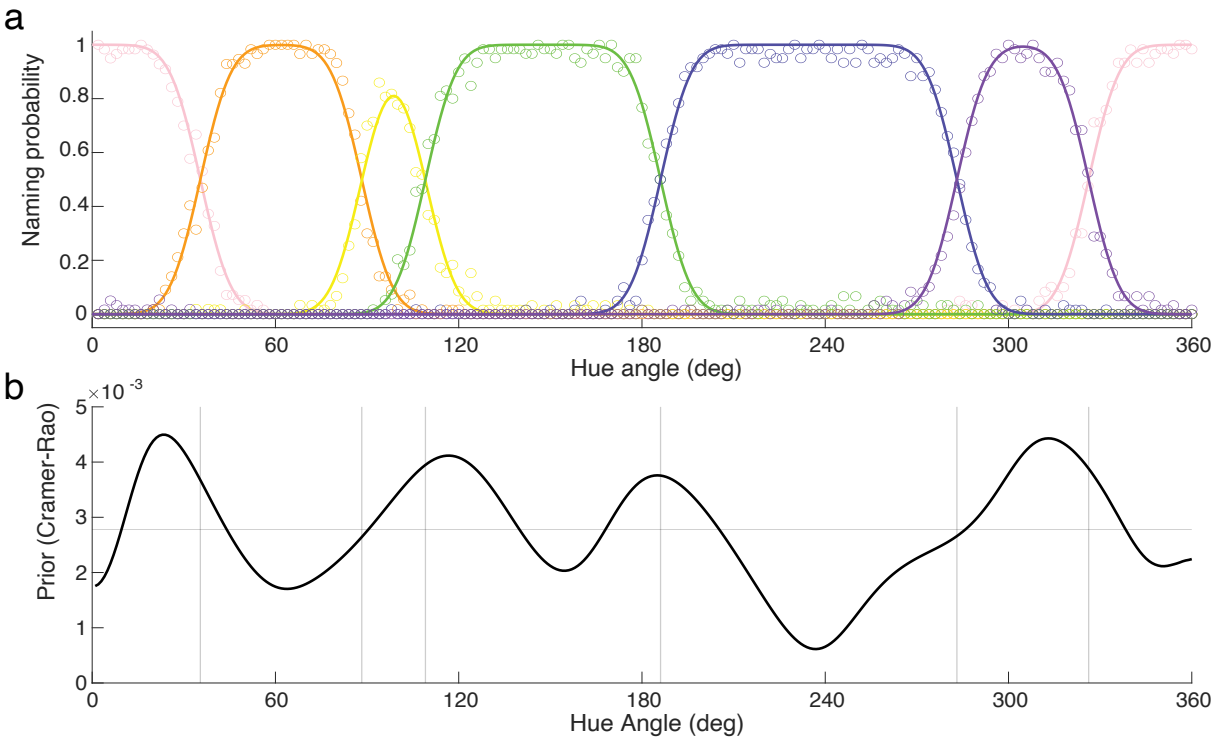

Figure 3.13: Categorical structure and prior of color. (a) Data from the color naming experiment in Bae et al. (2015) and smooth approximations using cumulative von Mises distributions (solid lines). These naming probabilities served as proxies for the underlying categorical structure $p(C|\theta)$. (b) Prior extracted from the bias and standard deviations of participants' response in the color matching experiment, based on the Cramer-Rao bound and the assumption that sensory encoding is efficient (Wei and Stocker, 2017; Noel et al., 2021). See *Methods* for details.

In order to validate the model, we first aim to specify and constrain the color category structure and color prior. In their study, Bae et al. (2015) also ran a color naming experiment where subjects were asked to select the color name that best described the test color out of a range of basic color names. Figure 3.13a shows subjects' probability of choosing each color name given a test color. We fit cumulative von Mises distributions to the boundaries of each pair of adjacent categories and use the results to constrain the categorical structure $p(C|\theta)$ and category uncertainty of our holistic matching model (see *Methods*). One of the advantages of validating our model against data of orientation perception is the availability of reliable measurements of the natural statistics of location visual orientations. For color spectra such measures are technically much more difficult to obtain with regard to different color spaces. As a result, we extracted an approximation of the hue prior directly from the color matching data in Bae et al. (2015) for the used CIELAB color space. This approximation is based on theoretical assumptions about how bias and standard deviation of an estimator are mutually dependent, and how they are connected to the input statistics for an efficient encoder (Wei and Stocker, 2017; Noel et al., 2021) (see *Methods*). The reconstructed hue prior $p(\theta)$ is shown in Fig. 3.13b.

Having extracted the categorical structure and hue prior, we then fit the data of the color matching experiments with both the hierarchical matching model and, for comparison, the efficient Bayesian estimator. Data for each experimental condition and the corresponding model fits are shown in Fig. 3.14. Similar to the orientation matching data, the hierarchical matching model remarkably well captures the entire shape of the error distributions for both conditions, especially the shifts of the distributions at category boundaries (Fig. 3.14a). In contrast, although the efficient Bayesian estimator qualitatively captures the repulsive bias pattern in the data and its dependency on sensory noise, predicted bias is generally too small similar to the orientation matching data (Fig. 3.4). A direct comparison of biases and variances predicted by the two models makes this more explicit (Fig. 3.14b,c). The comparison also demonstrates that even though the prior was approximated and extracted from the matching data, which introduces an element of circularity in the modeling process, such prior alone does not guarantee a good fit of the data without considering the categorical structure.

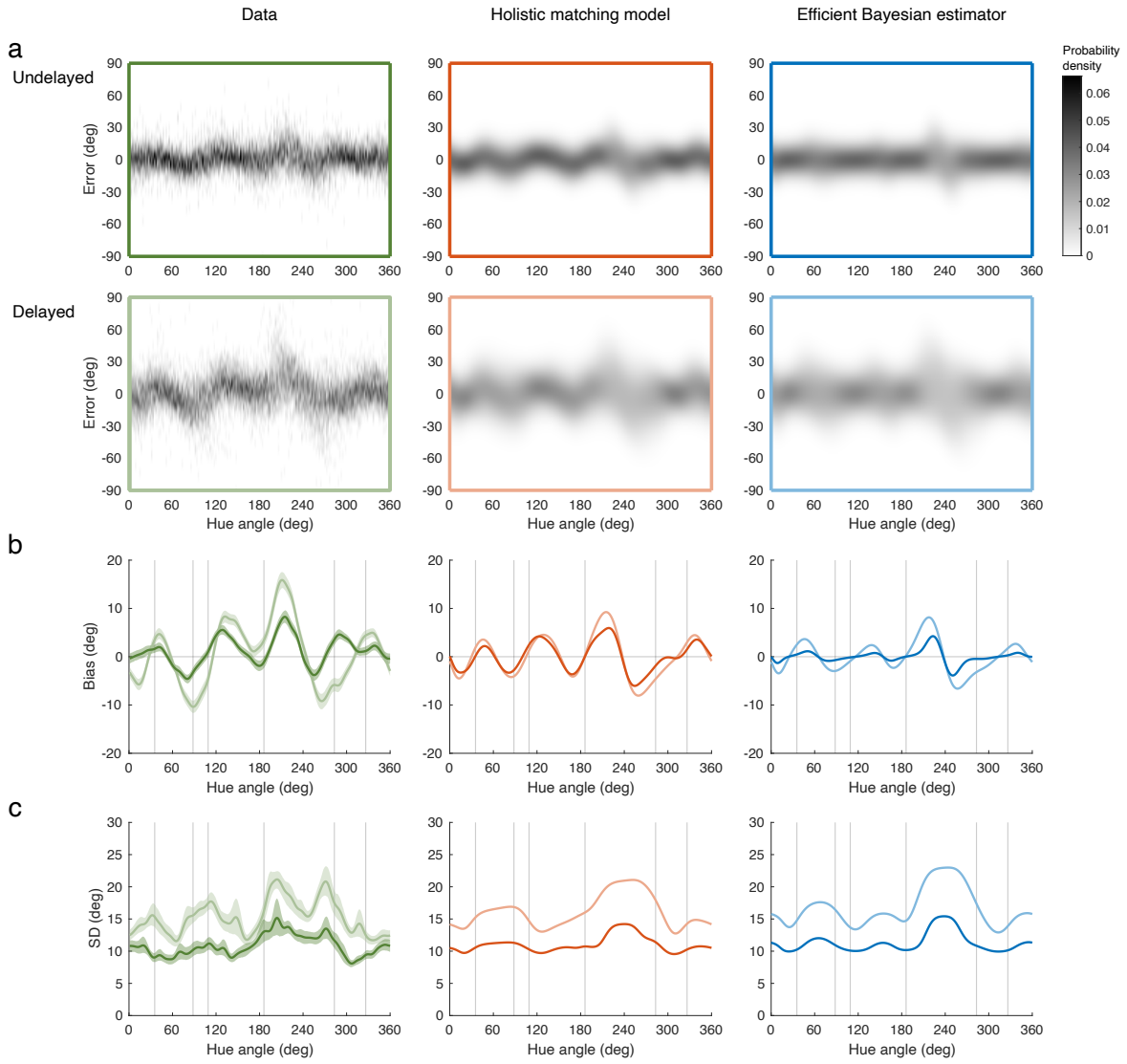

Figure 3.14: Data and model fits for the color matching experiment: (a) error distributions, (b) bias and (c) standard deviation. Columns show the data and the corresponding best-fit model predictions. Vertical lines show categorical boundaries. Data distributions show clear shifts with the biases generally repulsed away from categorical boundaries. Bias and variability are larger in the delayed condition than the undelayed condition. The overall pattern of the error distribution, the bias, and the standard deviation are well captured by the hierarchical matching model across conditions. While the efficient Bayesian estimator correctly predicts repulsive biases, the overall shape of the predicted error distributions does not well match the data, and the bias magnitude is much smaller. Data is re-analyzed from Bae et al. (2015). Shaded areas represent 95% confidence intervals from 1000 bootstrap runs. See *Methods* for details.
Supplementary Table 3.5. Fit parameter values of the holistic matching model.
Supplementary Figure 3.25. Fit category structure of the holistic matching model.
Supplementary Figure 3.26. Model comparison.

Our proposed holistic matching model also provides a better account of the data than the 'CATMET' model suggested by Bae et al. (2015). This model assumes that color perception is a conditioned inference process where an observer first picks the most likely color category of the stimulus and then infers its hue value according to the hue prior of the chosen color category (Stocker and Simoncelli, 2007). Such "self-consistent" inference model has previously been shown to account for choice-induced categorical effects in orientation perception (Luu and Stocker, 2018, 2021) (see also *Discussion*). However, model comparison shows that the holistic hierarchical matching model is superior to both the efficient Bayesian estimator as well the as CATMET model (Supplementary Fig. 3.26). More so, however, it suggests that the holistic matching model is general and provides a quantitative accurate account for human matching behavior for different perceptual modalities that are likely subject to categorical influences.

3.4. Discussion

We presented empirical and theoretical evidence that human sensory perception is a holistic inference process operating across a hierarchy of sensory representations. We introduced a holistic matching model to account for human behavior in a typical perceptual matching task for stimulus orientation. The model assumes that a subject's response in this task represents an optimal match of the probe and the test stimulus in terms of a combined objective function considering both their differences in orientation and category identity. We validated the model against three different existing psychophysical datasets probing human orientation perception. We showed that, in addition to an efficient sensory encoding of the stimulus orientation, a holistic inference process is necessary in order to provide an accurate account of subjects' full response distributions. The fact that subjects' response bias is not inverted when switching the role of the test and the probe stimulus in the matching experiment and that subjects still have repulsive bias when the test and the probe stimulus have the same bias is unique evidence for the holistic nature of the matching process, because any model operating only at the feature level would predict the opposite behavior. Furthermore, validation against data from color matching experiments confirmed the generality of the proposed model framework. The significance of our work lies in formulating and validating a novel, holistic inference theory for perception that explains why and how perception of a low-level

visual feature (e.g., orientation and color) is dependent on its high-level structural representation.

Its holistic nature fundamentally separates our framework from existing Bayesian observer models that have been proposed to account for categorical effects in perception (Feldman et al., 2009; Bae et al., 2015; Kronrod et al., 2016; Landy et al., 2017; Bill et al., 2020a; Gifford et al., 2014). While these models share a similar hierarchical generative process (Fig. 3.1b), inference in these models is limited to the feature level (i.e., orientation) by marginalizing over the entire generative hierarchy (i.e., categories). Marginalization effectively collapses the hierarchy, thereby reducing the inference process to a non-hierarchical Bayesian estimator with a heterogeneous prior determined by the weighted sum of the stimulus prior given each category. Thus the predictions of these models are identical to those of the non-holistic model considered in our study (e.g., Fig. 3.3). Other studies have proposed that inference over these hierarchical generative models is a sequential, top-down process where the category of the stimulus is inferred first before computing the posterior at the feature level conditioned on the inferred category (Stocker and Simoncelli, 2007; Bae et al., 2015; Ding et al., 2017; Luu and Stocker, 2018; Qiu et al., 2020) or the updated category belief (Lange et al., 2021), respectively. Although these "self-consistent" inference models predict increased perceptual biases away from categorical boundaries toward the center of the more likely stimulus category (i.e., confirmation biases), inference again is ultimately limited to an estimate at the feature level. Thus these models too can not explain why biases do not flip their sign when interchanging the probe and the test stimuli in a matching task(Tomassini et al. (2010); Fig. 3.9). Also, in contrast to these "self-consistent" inference models, the proposed holistic matching model is optimal and provides a rational, normative explanation for why and how categorical structures affect perceptual behavior. Sims et al. (2016) proposed a rate-distortion theory (RDT) based model using an objective function combining a cost at the feature and the category level similar to our approach. While the study showed how such optimal mapping can account for the estimation biases in color perception with regard to color categories, RDT is intrinsically an estimation model that would be rather difficult to adapt to modeling the matching process between test and probe stimulus under more general task conditions (i.e., with noisy probe stimuli). As a result, it too cannot account for the data by Tomassini et al. (2010).

It is worth highlighting the specific strengths of the proposed model, as well as its current limitations. First, our model makes detailed predictions of subjects' behavior by specifying the entire response distributions, which permits a stringent and fine-grained model validation. This is in contrast to the many models that limit validation to some summary statistics such as the average response and its biases (e.g., Huttenlocher et al., 1991). Similarly, the model makes individual predictions for meaningful parameters such as sensory noise levels, or the subjective uncertainty in the categorical structure of the stimulus. These are parameters that can be experimentally manipulated, allowing for selective empirical tests of the model. Second, despite its complexity due to the hierarchical structure, the model is relatively well constrained. In particular for visual orientation, we used a fixed prior distribution over stimulus orientation that reflects the measured statistics of visual orientation in natural scenes. This constraint likely prevents an even better quantitative account of the data yet demonstrates the robustness of our model (Supplementary Fig. 3.17). Furthermore, the model assumes that perceptual inference operates on efficient sensory representations, thus incorporating and extending previous work showing that human perception ubiquitously exhibits lawful hallmarks of efficient coding in combination with optimal Bayesian inference (Wei and Stocker, 2017). Thus, aside from the specification of the noise levels, the free model choices are essentially limited to the details specification of the categorical structure of orientation. Somewhat surprisingly, little is known about the natural categorical structure of human orientation perception. Thus our choice of 'cardinal' and 'oblique' discrimination reflects an intuitive assumption that, however, is shared with previous studies (e.g., Rosielle and Cooper, 2001; Wakita, 2004). However, it is reassuring that assuming a categorical structure that only distinguishes between clockwise and counter-clockwise orientations across the vertical meridian does not significantly change the model behavior (see Supplementary Fig. 3.18 and 3.20 for the 2-category model fit to both datasets). Future experiments are necessary to better constrain the categorical representations of visual orientation in human observers. Another caveat of the model is that the categorical weights fitted from different datasets vary substantially. Besides the possible differences in experiment setup, it is likely that there is trade-off between the feature loss and the categorical loss depending on the relative reliability of the evidence on the feature level and the categorical level that is not accounted for in the current

model. The feature loss in the current model is a circular variable version of the L2 loss for linear variables, which does not scale proportionally to the variance of stimulus uncertainty. Adopting a loss function that scales more prominently with feature uncertainty such as an information based feature loss may solve this problem. An ideal model would be to remove the arbitrary categorical weight parameter and to fully express the trade-off with the loss functions.

Our model makes clear predictions when we expect signatures of holistic inference to emerge in human response behavior, and when not (Fig. 3.12). A recent study showed that data from three similar orientation matching experiments (Van den Berg et al., 2012; Bays, 2014; Pratte et al., 2017) can be well accounted for by the efficient Bayesian estimation model (Taylor and Bays, 2018). However, in line with our predictions, the differences in experimental design compared to the study by De Gardelle et al. (2010) explain why the Bayesian estimator was sufficient to fit the data. In particular, presenting the test stimulus at fixed equi-distant stimulus locations and using an unambiguous probe stimulus reduces both categorical and probe uncertainty, thereby reducing the influence of categorical inference on the matching response (Fig. 3.12). Future experiments that will systematically manipulate categorical uncertainty and noise in the probe stimulus will help to validate these predictions in more detail. Finally, as low-level perception has been shown to follow common characteristics of sensory inference (e.g., Wei and Stocker, 2017), there is good reason to believe that our model generalizes to stimulus domains other than visual orientation or color. For example, many perceptual domains that involve circular variables, such as motion direction (Rauber and Treue, 1998), pointing direction (Smyrnis et al., 2014), visual and vestibular heading direction (Cuturi and MacNeilage, 2013), exhibit repulsive bias away from cardinal directions just like orientation, and have also been shown to have better discrimination at cardinal directions (Gros et al., 1998). Similarly, studies in visuospatial memory distortion have found biases towards landmarks, which has been explained by the efficient Bayesian estimation model (Langlois et al., 2021). It will be interesting to investigate the degree to which a full quantitative account of these effects requires to consider not just efficient sensory representations but also a holistic hierarchical inference process as proposed here.

Conclusions Bayesian estimation models have been successful in accounting for many well-known distortions in perceptual behavior. In particular in combination with efficiency constraints on the sensory representations, they provide meaningful (normative) explanations for many of the characteristic bias and variability pattern observed in perceptual estimation tasks in terms of prior expectations and sensory uncertainty. Our results suggest, however, that it is time to augment these models to address the holistic nature of perception, where inferences at all levels of the representational hierarchy are combined to generate perceptual behavior even in simple low-level perceptual tasks. The novel, holistic matching model is a first step in this direction, providing a normative and intuitive explanation for how category representations affect perceptual behavior in a frequently used psychophyiscal task.

3.5. Methods

Experimental methods

Subjects. 10 subjects participated in the experiment, with 9 naive subjects and one of the authors. All subjects had corrected-to-normal vision.

Apparatus. The experiment was run using Matlab and PsychToolbox (Brainard and Vision, 1997). Participants sat in a dark room and viewed stimuli on a VPixx3D screen (1920*1080 pixels resolution, 120 Hz refresh rate) at a 89 cm distance. A circular aperture (26 cm diameter) was placed on the screen to occlude the edges of the screen, removing cardinal orientation cues.

Stimuli. All stimuli were presented on a gray background. The test and probe stimuli were white noise with 100% contrast, filtered by a 1/f frequency filter within the range of 1-6 cpd, and an orientation filter with a symmetric wrapped Laplace profile centered at the desired orientation with a standard deviation of 1.4 deg or 35.6 deg for low- or high-noise stimuli, respectively. The stimulus had a Gaussian envelope (SD = 0.5 deg). In each trial, the test and the probe stimuli were presented at random locations 4.5 deg from the center of the screen in opposite directions (Fig.).

Procedure. In each trial, the test and the probe were presented on the screen. Subjects were instructed to push a joystick to continuously rotate the probe to match its orientation to the test. They could also finely adjust the probe orientation in steps of 0.5 deg by pressing one of two buttons.

The assignment of test and probe was not indicated, but could be easily inferred through an initial rotation by the subject. Both the test and the probe remained on the screen until the subjects confirmed their response by pressing a confirmation button.

There were four test-probe noise conditions: low-low, low-high, high-low, high-high. The test orientations were 5, 15, 25, ..., 175 deg, 18 orientations in total. The initial probe orientation was randomized. Subjects completed 40 trials for each test orientation in each noise condition. The noise conditions and test orientations were randomly interleaved across trials.

Data analysis

For each subject, each noise condition, and each test orientation, trials where the response was 4 standard deviations away from the mean were eliminated from the analysis and model fit. The net repulsion (Fig.) was computed by taking the average of the bias for test orientation in the 0 – 45 deg and 90 – 135 deg range and the negative bias for test orientation in the 45 – 90 deg and 135 – 180 deg range.

3.5.1. Existing psychophysical data

Dataset by De Gardelle et al. (2010). Each trial began with a background noise texture, then a test stimulus (Gabor patch) was presented for a variable duration at a random location 6.5 degs away from fixation. After the presentation of a mask and a blank interval, a randomly oriented probe stimulus (blue Gabor patch with only one visible strip) appeared at the test position. Participants were instructed to adjust the orientation of the probe using the mouse in order to reproduce the test orientation. Finally, they were also asked to report the visibility of the test stimulus on a continuous scale from 0 ("nothing seen") to 1 ("fully visible"). Presentation times of the test Gabor were [1000, 160, 80, 40, 20] ms and 0 ms (no stimulus presented), randomly intermixed. 46 subjects participated in the experiment, divided into five groups. Four groups were presented with random test orientations in 2/3 of the trials and one particular orientation (vertical, horizontal, right or left oblique) in the remaining trials. The fifth group always received random test orientations. Each subject completed two to four blocks of 120 trials each. For our analysis, we combined the data of all five groups of participants but only included trials in which the test orientations were randomly

selected. Furthermore, we excluded trials with presentation durations 20 ms (because data of those trials were too noisy to be reasonably analyzed) and 0 ms (because no test stimulus was shown). We also excluded trials for which the visibility rating was smaller than 0.01. After exclusion, the dataset contained 1103, 2187, 2140, and 1383 trials for each presentation duration, respectively. Illustrations of the data distributions (Fig. 3.3) and bias and standard deviation (Fig. 3.4) are based on smoothing the raw trial data with a symmetric Gaussian kernel centered at each data point. Kernel size (standard deviation) was chosen to provide the most accurate density estimation based on cross-validation (5 degs; see Supplementary Fig. 3.19). Distributions are normalized to indicate the conditional probability of response for each test orientation. In order to allow for a fair visual comparison between models and data, we applied the same smoothing procedure for the model predictions shown in Figs. 3.3 and 3.4.

Dataset by Noel et al. (2021). In each trial, a Gabor was presented at fixation for 120 ms. Then after the presentation of a mask and a blank interval, a randomly oriented probe stimulus (white Gabor patch with only one visible strip) appeared. Participants were instructed to adjust the orientation of the probe by button press to reproduce the test orientation. The experiment consisted of 3 blocks of 200 trials; the first block was without feedback; in the second and third blocks, participants were given feedback. 25 neurotypical individuals and 17 individuals diagnosed as within the ASD participated in the experiment. For our analysis, we combined the data across subjects within each group, but only included trials in the no-feedback block. We also excluded trials in which the responses were 3 standard deviations away from the mean response.

Dataset by Tomassini et al. (2010). In the main experiment, subjects viewed an array of Gabor patches and adjusted the implied orientation of two dots, placed on opposite sides of the fixation mark, such that it matched the average orientation of the Gabor patches. In the control experiment, the test and probe stimuli were interchanged: subjects adjusted the orientation of the Gabor array to match the orientation indicated by the two dots. Adjustments were done by pressing two keys on the keyboard. The orientation of each Gabor patch in the array was randomly selected from a Gaussian distribution centered at the test orientation with two different standard deviations, resulting in two different stimulus noise conditions (Fig. 3.9). The orientation of the test stimulus

was randomly selected from 18 orientations each 10 degs apart. For the main experiment, conditions with different fixed and response-terminated test presentation durations were measured in separate blocks. The control experiment only consisted of response-terminated presentations. Five subjects participated in the main experiment, each completing 8 trials per test orientation, presentation time, and stimulus noise level. Four subjects participated in the control experiment, each completing 16 trials per test orientation and stimulus noise level. Three subjects participated in both experiments. For our analysis, we combined the data across all subjects in both experiments, but only included the trials with response-terminated presentations. We also excluded trials in which the responses were 3 standard deviations away from the mean response.

Dataset by Bae et al. (2015). In the color naming experiment, subjects viewed a colored square and selected out of eight basic color terms the term that most closely described the test color. In the matching experiments, subjects viewed a colored square and and chose the color that best matched the test color by clicking on a color wheel. In the undelayed estimation, the colored square remained on the screen until the subject responded. In the delayed estimation, there was a delay period after the colored square disappeared before the color wheel was presented. 10 subjects participated in the color naming experiment, each completed 6 trials for each test color. 8 subjects participated in the undelayed estimation experiment, each completing 16 trials per test color for half of the test colors, resulting in 64 trials per test color from all subjects. 3 subjects participated in the delayed estimation experiment, each completing 20 trials per test color, resulting in 60 trials per test color from all subjects. For our analysis, we combined the data across all subjects in each experiments. We excluded trials in which the responses were 5 standard deviations away from the mean response in the estimation experiments.

Illustrations of the estimation error distributions (Fig. 3.14a) are based on smoothing the raw trial data with a 1D Gaussian kernel along the error axis centered at each data point with a SD of 3 deg. Illustrations of the bias and standard deviation (Fig. 3.14c,d) are based on smoothing the raw trial data with a running Gaussian window with a SD of 5 deg.

3.5.2. The holistic matching model

Efficient coding and feature inference. The following derivation follows Wei and Stocker (2015). Let θ be the orientation of the test stimulus and m its sensory measurement in a given trial. We assume that sensory encoding maximizes the mutual information between stimulus orientation and the sensory measurement (approximated by Fisher information (Wei and Stocker, 2016)), given that the total mutual information is limited. As a result, the prior distribution of the stimulus $p(\theta)$ and the Fisher information $J(\theta)$ of the sensory representation satisfy the efficient coding constraint

$$p(\theta) \propto \sqrt{J(\theta)} \ . \tag{3.2}$$

Sensory noise: consider a sensory space in which Fisher information is uniform. The optimal mapping $\tilde{\theta} = F(\theta)$ from stimulus to this sensory space is the cumulative of the stimulus distribution, thus $F(\theta) = \int p(\theta) d\theta$. The likelihood function in stimulus space $p(m|\theta)$ can be computed by applying the inverse mapping $\theta = F^{-1}(\tilde{\theta})$ to the homogeneous likelihood function in sensory space $p(\tilde{m}|\tilde{\theta})$ obtained by assuming uniform sensory noise according to a von Mises distribution

$$p(\tilde{m}|\tilde{\theta}) = \mathrm{vm}(\tilde{m}; \tilde{\theta}, \kappa_i) \ , \tag{3.3}$$

with κ_i representing the sensory noise magnitude.

Stimulus noise: for the test stimulus in Tomassini et al. (2010) (array of Gabor patches) we assume uniform noise in stimulus space with stimulus samples θ' on each trial for a given stimulus θ following a von Mises distribution

$$p(\theta'|\theta) = \mathrm{vm}(\theta'; \theta, \kappa_e) \ , \tag{3.4}$$

where κ_e represents the stimulus noise magnitude. The stimulus sample θ' corresponds to $\tilde{\theta'} = F(\theta')$ in sensory space and elicits a noisy sensory measurement $\tilde{m}$ according to Eq. (3.3), hence

$$p(\tilde{m}|\tilde{\theta'}) = \mathrm{vm}(\tilde{m}; \tilde{\theta'}, \kappa_i) \ . \tag{3.5}$$

The distribution of the sensory measurement m in stimulus space is

$$p(m|\theta') = p(\tilde{m}|\tilde{\theta'})F'(m) \ , \tag{3.6}$$

where $\tilde{m} = F(m)$. The likelihood function that takes both stimulus noise and sensory noise into account is

$$p(m|\theta) = \int p(m|\theta')p(\theta'|\theta)\,\mathrm{d}\theta' \ . \tag{3.7}$$

Finally, based on the generative model (Fig. 3.1b) the posterior over stimulus orientation given the sensory measurement is

$$p(\theta|m) \propto p(m|\theta)\sum_i p(\mu|C_i)p(C_i) \propto p(m|\theta)p(\theta) \ , \tag{3.8}$$

where the orientation prior $p(\theta)$ is represents the natural orientation statistics (Fig. 3.2a).

Categorical inference. *Orientation categories:* We assume four natural categories for orientation: vertical ('V'), horizontal ('H'), clockwise('CW') or counter-clockwise ('CCW') oblique relative to vertical ($C \in \mathbb{C} = \{\text{'H'}, \text{'V'}, \text{'CW'}, \text{'CCW'}\}$). The horizontal category is defined by the von Mises distribution

$$p(C = \text{'H'}|\theta; \mu_H) = \alpha\frac{\mathrm{vm}(\theta; \mu_H, \kappa_c)}{\mathrm{vm}(\mu_H; \mu_H, \kappa_c)} \ , \tag{3.9}$$

where α is the probability of the horizontal category at μ_H, κ_c represents the uncertainty in the categorical boundaries, and μ_H represents a noisy signal of the horizontal orientation that may stochastically vary across trials according to

$$p(\mu_H) = \mathrm{vm}(\mu_H; 0\,\mathrm{deg}, \kappa_c) \ . \tag{3.10}$$

The vertical category is similarly defined with its center μ_V always 90 degrees away from μ_H (Eq. (3.10)). The oblique categories are the orientations in between the cardinal categories with a

smooth transition given by the cumulative von Mises distributions

$$p(\text{`CCW'}|\theta;\mu_H) = (\text{cum vm}(\theta;\mu_H,\kappa_c) - \text{cum vm}(\theta;\mu_V,\kappa_c))(1 - p(\text{`H'}|\theta) - p(\text{`V'}|\theta)) \tag{3.11}$$

and

$$p(\text{`CW'}|\theta;\mu_H) = (\text{cum vm}(\theta;\mu_V,\kappa_c) - \text{cum vm}(\theta;\mu_H,\kappa_c))(1 - p(\text{`H'}|\theta) - p(\text{`V'}|\theta)) \ , \tag{3.12}$$

respectively. For simplicity, we assume a single parameter κ_c to represent the uncertainty in the cardinal orientations and the uncertainty in the categorical boundaries.

Finally, with the generative model (Fig. 3.1b) the posterior probability over category C can be computed as

$$\begin{aligned} p(C|m;\mu_H) = &\ \tfrac{1}{p(m)} \int_\theta p(m|\theta)p(\theta|C;\mu_H)p(C) \\ = &\ \tfrac{1}{p(m)} \int_\theta p(m|\theta)p(\theta)p(C|\theta;\mu_H) \\ = &\ \int_\theta p(C|\theta;\mu_H)p(\theta|m) \ . \end{aligned} \tag{3.13}$$

Color categories: We extract the category structure from the color naming experiment in Bae et al. (2015). Following the original study, we assume six color categories ($C \in \mathbb{C} = \{C_1, C_2, ..., C_6\}$). We assume that due to the uncertainty in the boundary position, every boundary μ_j jitters around its respective mean position b_j by the same deviation $\Delta\mu$ in each trial

$$\mu_j = b_j + \Delta\mu \ (j = 1, 2, ..., 6) \ , \tag{3.14}$$

and the deviation follows a von Mises distribution

$$p(\Delta\mu) = \text{vm}(\Delta\mu; 0, \kappa_b) \ , \tag{3.15}$$

where κ_b is the uncertainty in the categorical boundary. Because in the color naming experiment,

the test color was presented on the screen until observers responded, sensory noise is small, so the uncertainty in the responses is predominantly caused by the uncertainty in the category boundaries. The probability of choosing category C_j for a noiseless stimulus with hue angle θ is

$$p(\text{answer } C_j|\theta) = \text{cum vm}(\theta; b_j, \kappa_b) - \text{cum vm}(\theta; b_{j+1}, \kappa_b) \ . \tag{3.16}$$

We fit this probability to the color naming data to obtain κ_b and b_j $(j = 1, 2, ..., 6)$.

Same as with orientation, we assume that color categories overlap according to cumulative von Mises distributions

$$p(C_j|\theta; \Delta\mu) = \text{cum vm}(\theta; \mu_j, \kappa_c) - \text{cum vm}(\theta; \mu_{j+1}, \kappa_c) \ , \tag{3.17}$$

where κ_c specifies the overlap between neighboring categories. Finally, the posterior probabilities of each category is computed according to Eqs. (3.13).

Matching. We assume that participants adjust the probe stimulus while obtaining continuous visual feedback about the probe orientation (space or color). Let θ_p be the orientation of the probe, m_p the sensory measurement of the probe, and C_p the category of the probe. For simplicity, we assume that motor noise is additive, induced only after the probe is optimally adjusted (Fig. 3.1b). Note that we have considered more elaborate visuomotor control models that take motor noise into consideration when computing the optimal match, but found that the predictions do not significantly differ for the typical noise levels observed in orientation matching tasks.

The categorical loss is defined as whether the category of the probe is different from the category of the test orientation C:

$$L_c(C, C_p) = \begin{cases} 0 & \text{if } C_p = C \\ 1 & \text{otherwise} \ . \end{cases} \tag{3.18}$$

The feature loss is assumed to be the cosine of the difference between the probe and the test

orientation, thus

$$L_\theta(\theta, \theta_p) = 2(1 - \cos(2(\theta - \theta_p))) \ . \tag{3.19}$$

The cosine loss for circular variables is equivalent to the L_2 loss for linear variables in the sense that the optimal estimate is defined by the (circular) mean of the posterior distribution.

The total loss (Eq.3.1) is the weighted sum of the feature loss and the categorical loss. Given the sensory measurements m and m_p, the expected total loss is

$$\begin{aligned} E[L_{tot}|m, m_p; \mu_H] = &\iint L_\theta(\theta, \theta_p) p(\theta|m) p(\theta_p|m_p) \, \mathrm{d}\theta \, \mathrm{d}\theta_p \\ &+ w \sum_{C_0 \in \mathbb{C}} p(C = C_0|m; \mu_H)(1 - p(C_p = C_0|m_p; \mu_H)) \ . \end{aligned} \tag{3.20}$$

When there is no noise in the probe ($m_p = \theta_p$) the expected total loss simplifies to

$$\begin{aligned} E[L_{tot}|m, \theta_p; \mu_H] = &\int L_\theta(\theta, \theta_p) p(\theta|m) \, \mathrm{d}\theta \\ &+ w \sum_{C_0 \in \mathbb{C}} p(C = C_0|m; \mu_H)(1 - p(C_p = C_0|\theta_p; \mu_H)) \ . \end{aligned} \tag{3.21}$$

The optimal probe orientation $\hat{\theta_p}$ that minimizes the expected loss is

$$\hat{\theta_p}(m; \mu_H) = \arg\min_{\theta_p} E[L_{tot}|m, \theta_p; \mu_H] \ . \tag{3.22}$$

When there is noise in the probe, the observer has to minimize the expected loss based on the sensory measurement m_p. For simplicity, we omit a description of how the observer adjusts the probe using visuomotor feedback. We simply assume that the observer adjusts the probe until they detect a probe measurement $\hat{m_p}$ that minimizes the expected total loss, hence

$$\hat{m_p}(m; \mu_H) = \arg\min_{m_p} E[L_{tot}|m, m_p; \mu_H] \ . \tag{3.23}$$

Predicted response distribution. Above is a description of the perceptual decision process of determining the optimal probe response, expressed by Eqs. (3.22) and (3.23), respectively. Now we

calculate the predicted response distribution that follows from this process.

When there is noise in the perception of the probe stimulus, then there are different probe orientations $\hat{\theta_p}$ that could have generated the optimal probe measurement $\hat{m_p}$. Since the probe orientation is generated by the observer and not the natural environment, we simply assume that the probability of the adjusted probe orientation given the optimal probe measurement $\hat{m_p}$ is proportional to the likelihood function as defined by Eq. (3.7) and is not affected by any non-uniform prior assumption, hence

$$p(\hat{\theta_p}|\hat{m_p}(m;\mu_H)) \propto p(\hat{m_p}(m;\mu_H)|\hat{\theta_p}) \ . \tag{3.24}$$

Because $\hat{m_p}(m;\mu_H)$ is a deterministic function (Eq. (3.23)) we can rewrite the probability distribution as

$$p(\hat{\theta_p}|m;\mu_H) = p(\hat{\theta_p}|\hat{m_p}(m;\mu_H)) \ . \tag{3.25}$$

When the probe stimulus is noise-free, Eq. (3.25) turns into a Dirac delta distribution at the optimal $\hat{\theta_p}$ (Eq. (3.22)).

Finally, we assume that when the observer confirms the intended probe orientation θ_p (e.g., with a button press), additive motor noise corrupts the answer leading to a noisy response $\hat{\theta_p}^*$ according to

$$p(\hat{\theta_p}^*|\hat{\theta_p}) = \mathrm{vm}(\hat{\theta_p}^*;\hat{\theta_p},\kappa_m) \ , \tag{3.26}$$

where κ_m represents the motor noise magnitude.

Taken together, the predicted probability distribution of the matching response $\hat{\theta_p}^*$ to a test orientation θ can be computed as

$$p(\hat{\theta_p}^*|\theta) = \iiint p(\hat{\theta_p}^*|\hat{\theta_p})p(\hat{\theta_p}|m;\mu_H)p(m|\theta)p(\mu_H)\,\mathrm{d}\hat{\theta_p}\,\mathrm{d}m\,\mathrm{d}\mu_H \ , \tag{3.27}$$

with the terms in the integral given by Eqs. (3.26), (3.25), (3.7), and (3.10), respectively. For color, μ_H is replaced by $\Delta\mu$ and $p(\Delta\mu)$ is given by Eq. (3.15).

3.5.3. The non-holistic matching model

The non-holistic matching model shares the same feature inference process as the holistic matching model, but does not consider categorical inference (Fig. 3.1b). The matching process only consists of minimizing the feature mismatch between test θ and probe orientation θ_p. The calculation of the response distributions for different noise conditions is identical to the calculations for the holistic matching model above (Eq. (3.27) except that it is not dependent on category noise μ_H.

If the probe stimulus is noiseless, the non-holistic matching model is equivalent to the *efficient Bayesian estimator* (Wei and Stocker, 2015) with the assumption that the probe orientation θ_p is a direct representation of the optimal estimate $\hat{\theta}$ of the test orientation according to the loss function L_θ (Eq. (3.19)) aside from potential, additive motor noise. The efficient Bayesian estimator shares the same efficient feature encoding as the holistic matching model described above. In contrast, the *standard Bayesian estimator* assumes homogeneous encoding such that the sensory measurements m given the stimulus sample θ' follows the von Mises distribution

$$p(m|\theta') = \mathrm{vm}(m; \theta', \kappa_i) \ , \tag{3.28}$$

where κ_i represents the constant sensory noise magnitude, independent of θ'.

3.5.4. Model fitting

We jointly fit the model to the data of all the conditions in each dataset by maximizing the likelihood of the data given the model:

$$p(D|\rho) = \prod_{j=1}^{n} p(D_j|\rho) = \prod_{j=1}^{n} p(\hat{\theta_j}|\rho, \theta_j) \ , \tag{3.29}$$

where D is the data, ρ represents the parameters of the model, θ_j is the test orientation and $\hat{\theta_j}$ is the measured matching orientation (probe) in trial j, and n is the total number of trials.

We assume a fixed orientation prior for all model fits, representing the average natural orientation statistics extracted from indoor and outdoor scenes images (Coppola et al., 1998). More specifically,

we apply a spline fit to the orientation histograms of indoor and outdoor scenes separately assuming that the distributions are symmetric around the vertical orientation (Supplementary Fig. 3.17a), and then take the average of the two spline fits as the orientation prior $p(\theta)$ (Fig. 3.2a).

For fitting the data by De Gardelle et al. (2010), we assume no stimulus noise and four sensory noise levels corresponding to the four different presentation durations, resulting in a total of 8 free parameters:

- a group of four parameters κ_i for four sensory noise levels;
- κ_c for category uncertainty;
- α for the probability of cardinal category;
- w for the weight of the categorical loss;
- κ_m for motor noise.

For fitting the data by Noel et al. (2021), we assume no stimulus noise and we fit data from the neurotypical and ASD subjects separately, resulting in 5 free parameters for each subject group:

- κ_i for sensory noise;
- κ_c for category uncertainty;
- α for the probability of cardinal category;
- w for the weight of the categorical loss;
- κ_m for motor noise.

For fitting the data by Tomassini et al. (2010), we assume one sensory noise level across all the conditions, and two stimulus noise levels corresponding to the two different standard deviations of the Gabor orientations in the stimulus array. So the holistic matching model fit contains 7 free parameters:

- κ_i for sensory noise;
- a group of two parameters κ_e for two stimulus noise levels;
- κ_c for category uncertainty;
- α for the for the probability of cardinal category;
- w for the weight of the categorical loss;
- κ_m for motor noise.

For fitting the data from the new orientation matching experiment (Fig. 3.11), we assume one sensory noise for both the test and the probe across all the conditions, and two stimulus noise levels corresponding to the low and high noise stimulus for the test and the probe. So the holistic matching model fit contains 7 free parameters:

- κ_i for sensory noise;
- a group of two parameters κ_e for two stimulus noise levels;
- κ_c for category uncertainty;
- α for the for the probability of cardinal category;
- w for the weight of the categorical loss;
- κ_m for motor noise.

For fitting the color data by Bae et al. (2015), we first extract the categorical structure by fitting the color naming probabilities to the color naming data according to the parameterization described above (Eq. (3.16)), with 7 free parameters for the mean boundary positions b_j $(j = 1, 2, ..., 6)$ and the uncertainty in boundary positions κ_b.

Furthermore, we reconstruct the prior $p(\theta)$ on hue orientation from the bias $b(\theta)$ and standard

deviation $\sigma(\theta)$ of the participants' response using the Cramer-Rao bound (Wei and Stocker, 2017; Noel et al., 2021)

$$\sqrt{J(\theta)} \propto \frac{|1 + b'(\theta)|}{\sigma(\theta)} \tag{3.30}$$

and the efficient coding constraint (Eq. (3.2)). For reconstruction, we use the data from the delayed condition because bias and standard deviation are larger, and thus the deviations from the the Cramer-Rao bound due to motor and other late noise is smaller. This results in a more accurate, closer reconstruction of the underlying prior distribution. Reconstruction is based on a polynomial fit (degree 20) to the measured bias and standard deviation, respectively (Supplementary Fig. 3.24).

For fitting the matching data (Fig. 3.14), we assume no stimulus noise and two sensory noise levels corresponding to the undelayed and delayed condition, resulting in a total of 5 free parameters:

- a group of two parameters κ_i for two sensory noise levels;
- κ_c for the overlap between categories;
- w for the weight of the categorical loss;
- κ_m for motor noise.

The efficient Bayesian estimator has free parameters for sensory noise, stimulus noise, and motor noise; thus it has five free parameters for the data by De Gardelle et al. (2010), four for the data by Tomassini et al. (2010), and three for the data by Bae et al. (2015) (no stimulus noise).

The standard Bayesian estimator has the same free parameters as the efficient Bayesian estimator, except that for the comparison with the data by De Gardelle et al. (2010), we fixed the motor noise to be the same value obtained from the fit with the efficient Bayesian estimator (including the fit to the training set in each cross-validation run).

3.5.5. Cross-validation

In each run of cross-validation, we randomly partition the data into a training set containing 80% of the trials and a validation set consisting of the remaining 20% of the trials. The partition is done

separately for each noise level. We fit the model to the training set, then compute the likelihood of the fit model given the validation data. This likelihood represents the degree to which the fit model is supported by the validation data. We repeat this process 100 times.

The "omniscient" observer model. The omniscient model is an empirical model that serves as a reference for cross-validation. It directly considers the data in the training set as a prediction of the error distribution using kernel density estimation. Each data point in the training set is transformed into a symmetric 2D Gaussian probability kernel (diagonal covariance matrix). The resulting distribution is then normalized for each test orientation. The performance of the omniscient model on the validation set depends on the width of the Gaussian kernel: if the width is too small, the model over-fits the training set; if the width is too large, the prediction is too general and the model loses predictive power. We cross-validated the omniscient model with different standard deviations and found that a standard deviation of 5 degrees leads to the best performance (Supplementary Fig. 3.19).

Parameter	Value
De Gardelle et al. (2010)	
κ_i: sensory noise	[356.94, 15.79, 4.58, 2.10]
κ_c: category uncertainty	8.29
α: cardinal probability	0.60
w: categorical weight	10.75
κ_m: motor noise	34.63
Tomassini et al. (2010)	
κ_i: sensory noise	698.50
κ_e: stimulus noise	[651.39, 19.14]
κ_c: category uncertainty	4.79
α: cardinal probability	0.52
w: categorical weight	0.52
κ_m: motor noise	38.20

Table 3.1: Best-fitting model parameters for data in De Gardelle et al. (2010) and Tomassini et al. (2010).

Parameter	Value
De Gardelle et al. (2010)	
κ_i: sensory noise	[211.47, 15.49, 4.59, 1.98]
κ_b: boundary noise	58.61
κ_c: category overlap	1.82
w: categorical weight	6.59
κ_m: motor noise	24.32
Tomassini et al. (2010)	
κ_i: sensory noise	695.69
κ_e: stimulus noise	[699.92, 17.56]
κ_b: boundary noise	9.27
κ_c: category overlap	2.33
w: categorical weight	0.69
κ_m: motor noise	37.29

Table 3.2: Best-fitting parameters of the 2-category holistic matching model for data in De Gardelle et al. (2010) and Tomassini et al. (2010).

Parameter	Value
Noel et al. (2021): Neurotypical subjects	
κ_i: sensory noise	19.17
κ_c: category uncertainty	8.25
α: cardinal probability	0.56
w: categorical weight	6.98
κ_m: motor noise	27.56
Noel et al. (2021): ASD subjects	
κ_i: sensory noise	6.19
κ_c: category uncertainty	8.04
α: cardinal probability	0.60
w: categorical weight	4.93
κ_m: motor noise	120.18

Table 3.3: Best-fitting model parameters for data in Noel et al. (2021).

Parameter	Value
κ_i: sensory noise	349.46
κ_e: stimulus noise	[698.17, 33.14]
κ_c: category uncertainty	6.85
α: cardinal probability	0.64
w: categorical weight	0.18
κ_m: motor noise	398.62

Table 3.4: Best-fitting model parameters for the combined subject from the new orientation matching experiment.

Parameter	Value
Bae et al. (2015): color naming	
b: mean boundary positions	[35.4, 88.3, 109.1, 186.1, 283.1, 326.2]
κ_b: boundary noise	52.42
Bae et al. (2015): color matching	
κ_i: sensory noise	[102.52, 21.63]
κ_c: category overlap	7.70
w: categorical weight	0.42
κ_m: motor noise	40.57

Table 3.5: Best-fitting parameters of the holistic matching model for the color naming and color estimation data in Bae et al. (2015).

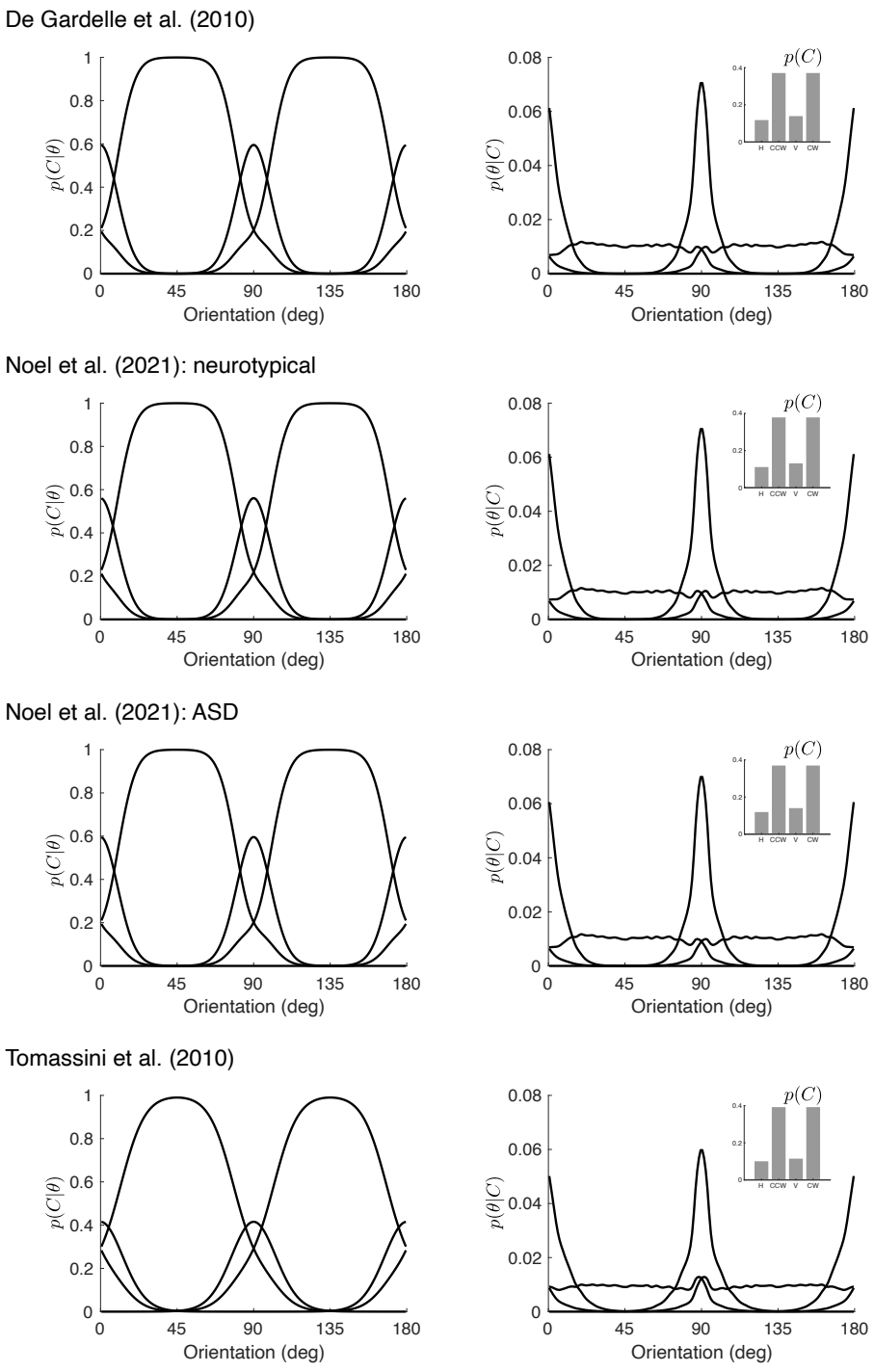

Figure 3.15: Best-fitting categories for the three existing orientation datasets.

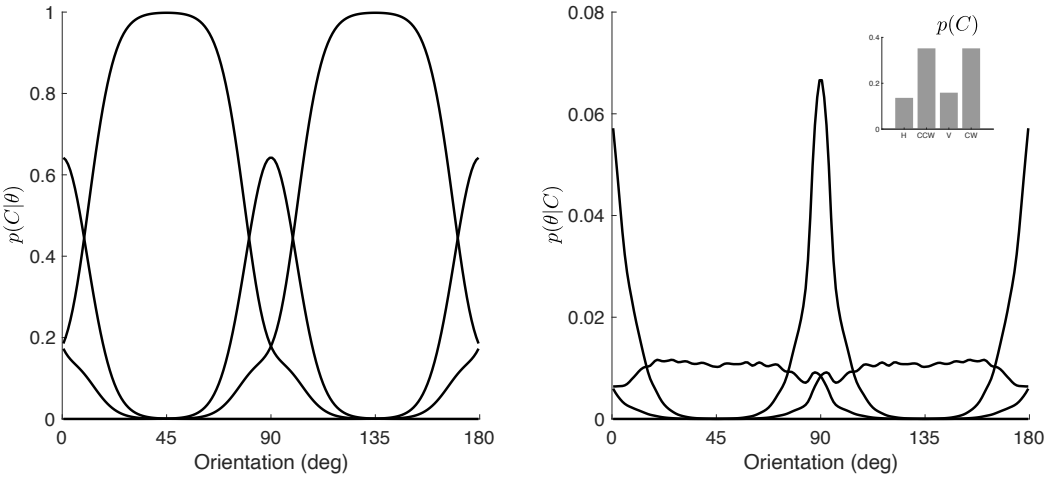

Figure 3.16: Best-fitting categories for the combined subject from the new orientation matching experiment.

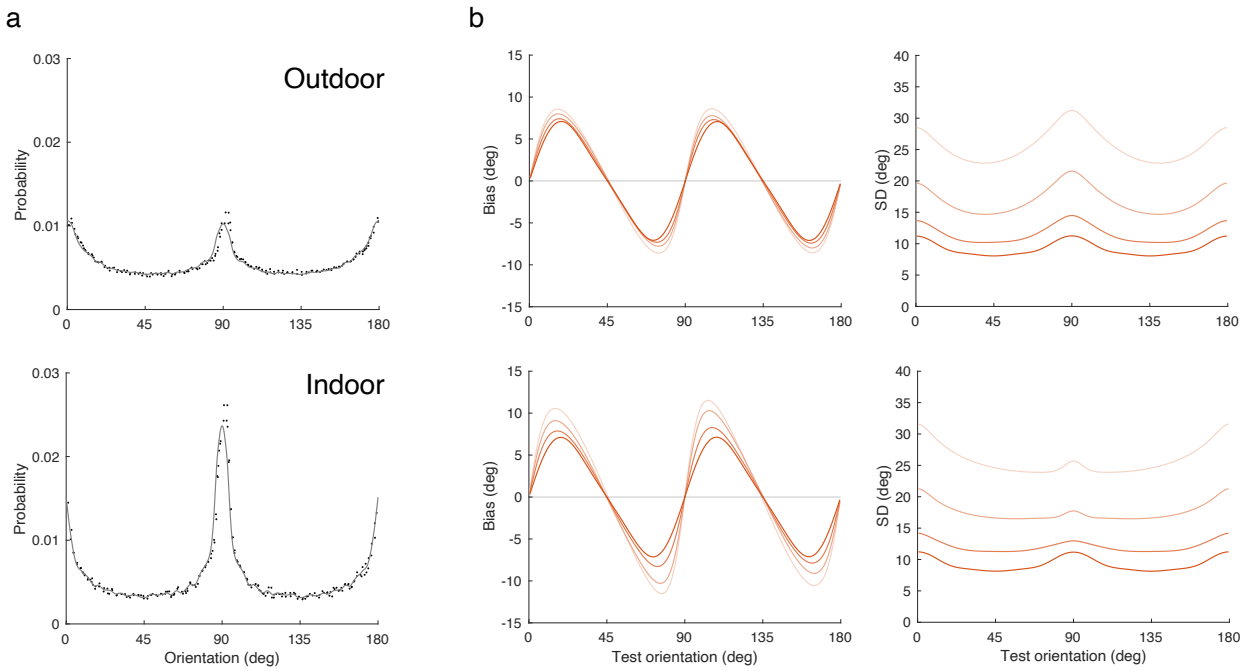

Figure 3.17: Predictions of the holistic matching model using "outdoor" and "indoor" orientation priors, respectively. (a) Image statistics of indoor and outdoor natural scenes (dots) and their smooth spline interpolations representing the corresponding prior distributions (lines). We assume the distributions to be symmetric around vertical. Data reanalyzed from Coppola et al. (1998). (b) Predicted bias and standard deviation of the holistic matching model using the two different prior distributions. All other model parameters are identical to the best-fit values listed in Supplementary Table 3.1. Patterns in bias and standard deviation are qualitatively similar across the two priors. The peakier "indoor" prior leads to larger repulsive biases yet less pronounced differences in standard deviation compared to the "outdoor" prior. Simulations and model fits in the main text all use a fixed prior distribution that represents the average between the "indoor" and "outdoor" prior (Fig. 3.2).

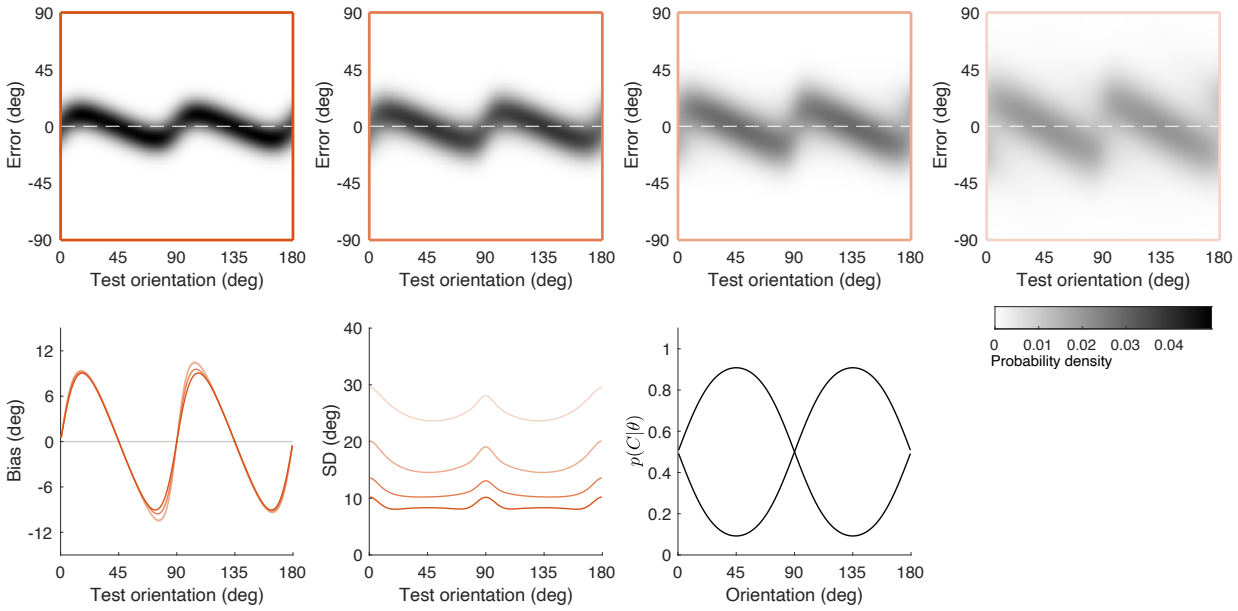

Figure 3.18: Two-category holistic matching model fit for matching task with noiseless probe. The cardinal probability α is set to zero, and the parameters for boundary noise and category overlap are allowed to vary independently. The fitting procedure is otherwise identical to the model with four categories. Fit parameter values are listed in Supplementary Table 3.2.

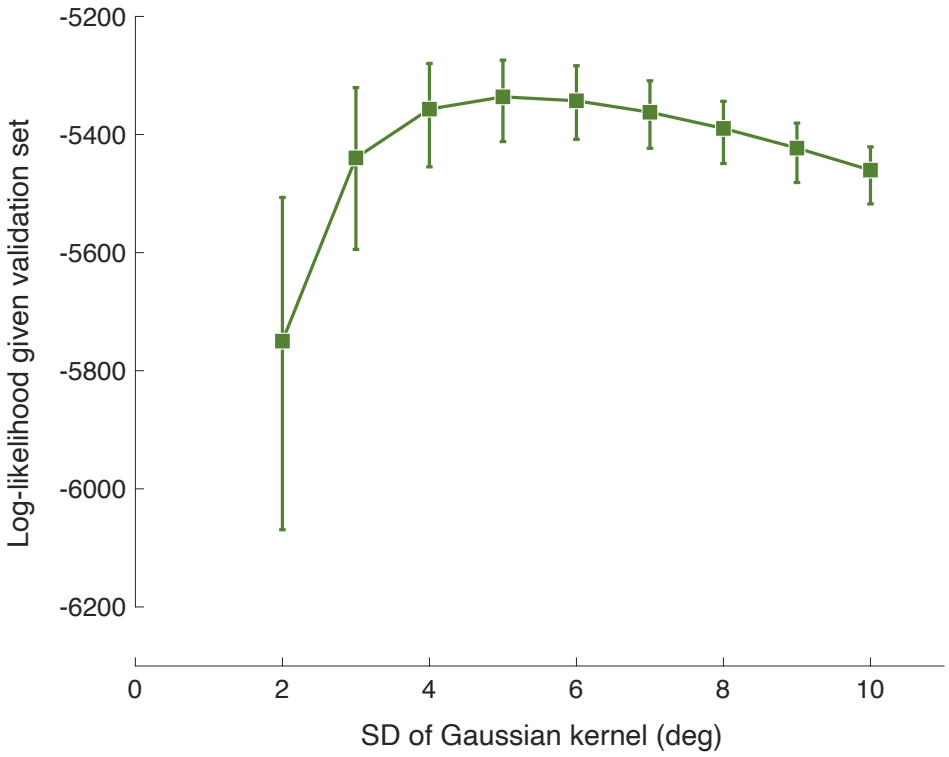

Figure 3.19: Cross-validation of the kernel density estimation accuracy for the omniscient model as a function of different Gaussian kernel size (standard deviation). Squares represent the median and error bars represent 95% confidence intervals of 100 repetitions of a repeated random sub-sampling cross-validation procedure. Accuracy shows a lawful dependency on kernel size with a standard deviation of 5 degrees providing the largest median likelihood value.

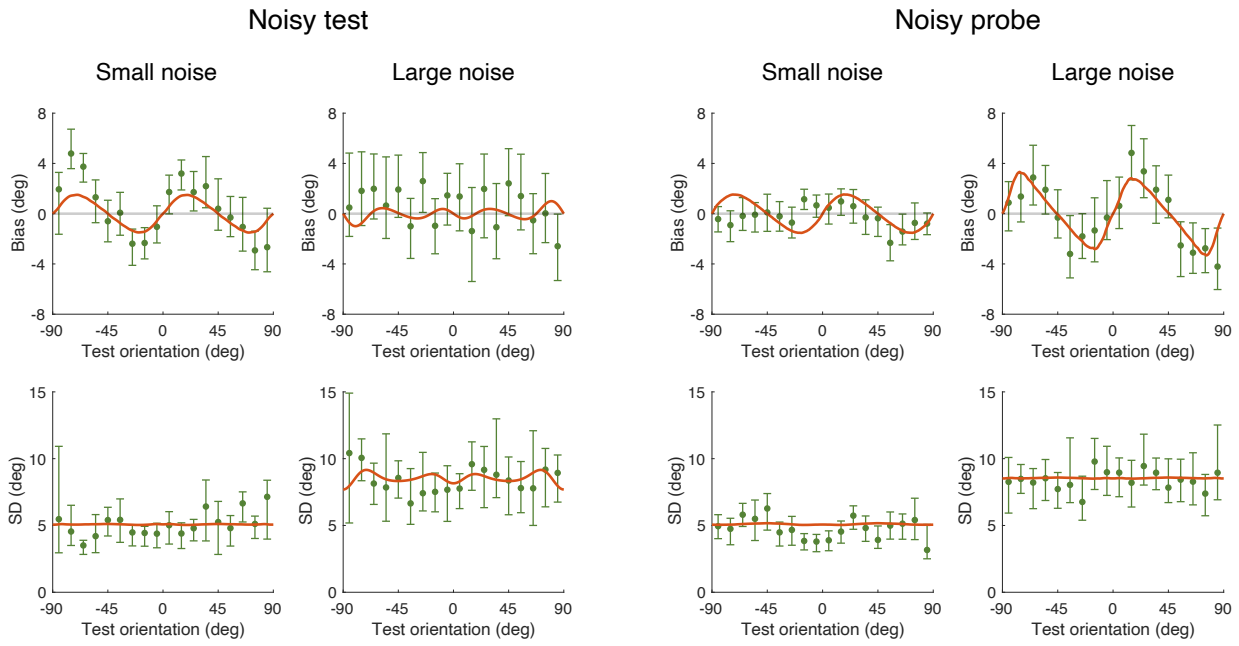

Figure 3.20: Two-category holistic matching model fit for matching task with noisy probe. The cardinal probability α is set to zero, and the parameters for boundary noise and category overlap are allowed to vary independently. The fitting procedure is otherwise identical to the model with four categories. Fit parameter values are listed in Supplementary Table 3.2.

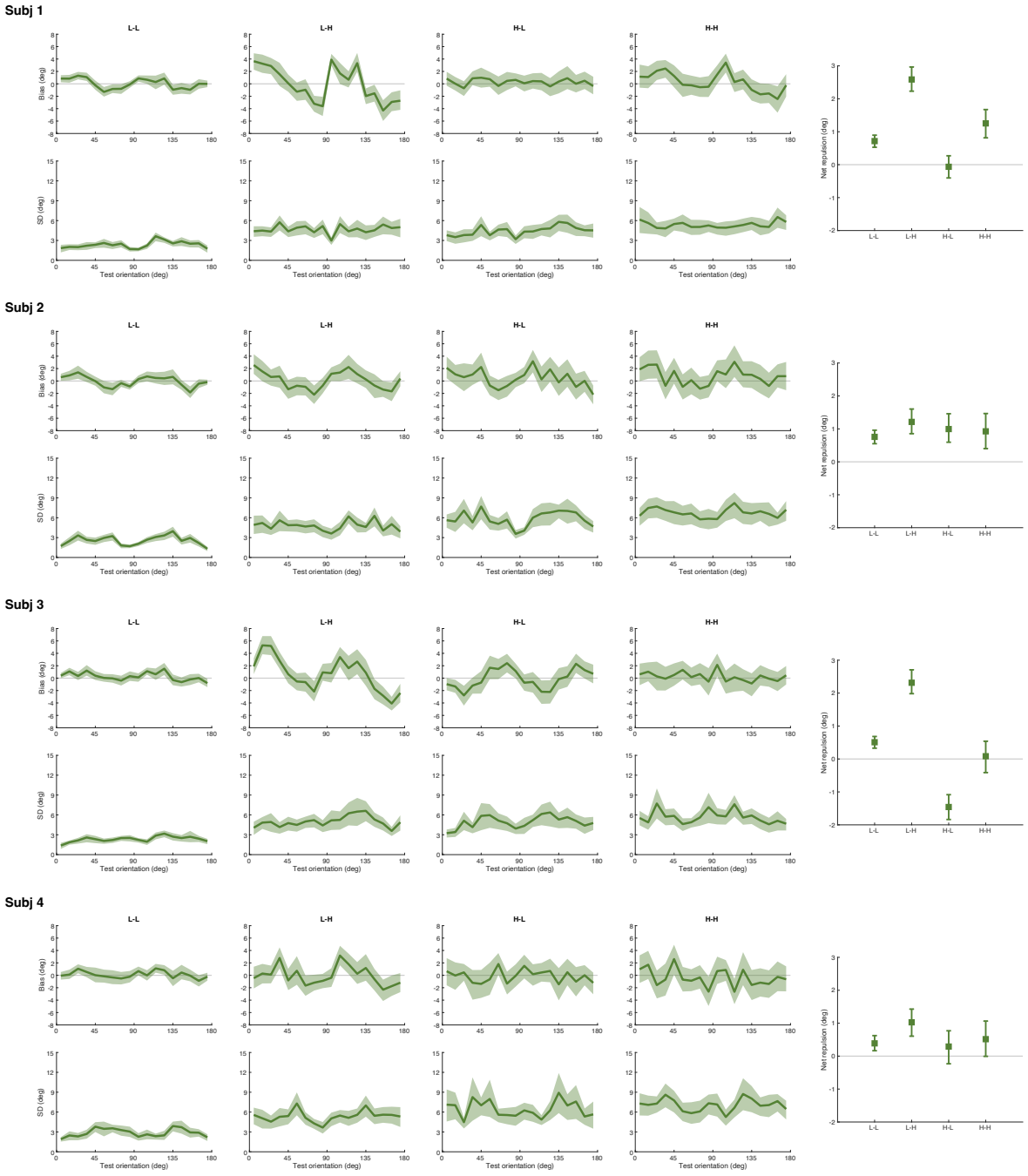

Figure 3.21: Individual subjects' data from the orientation matching experiment: subject 1 – 4. Error bars represent 95% confidence intervals from 1000 bootstrap samples of the data.

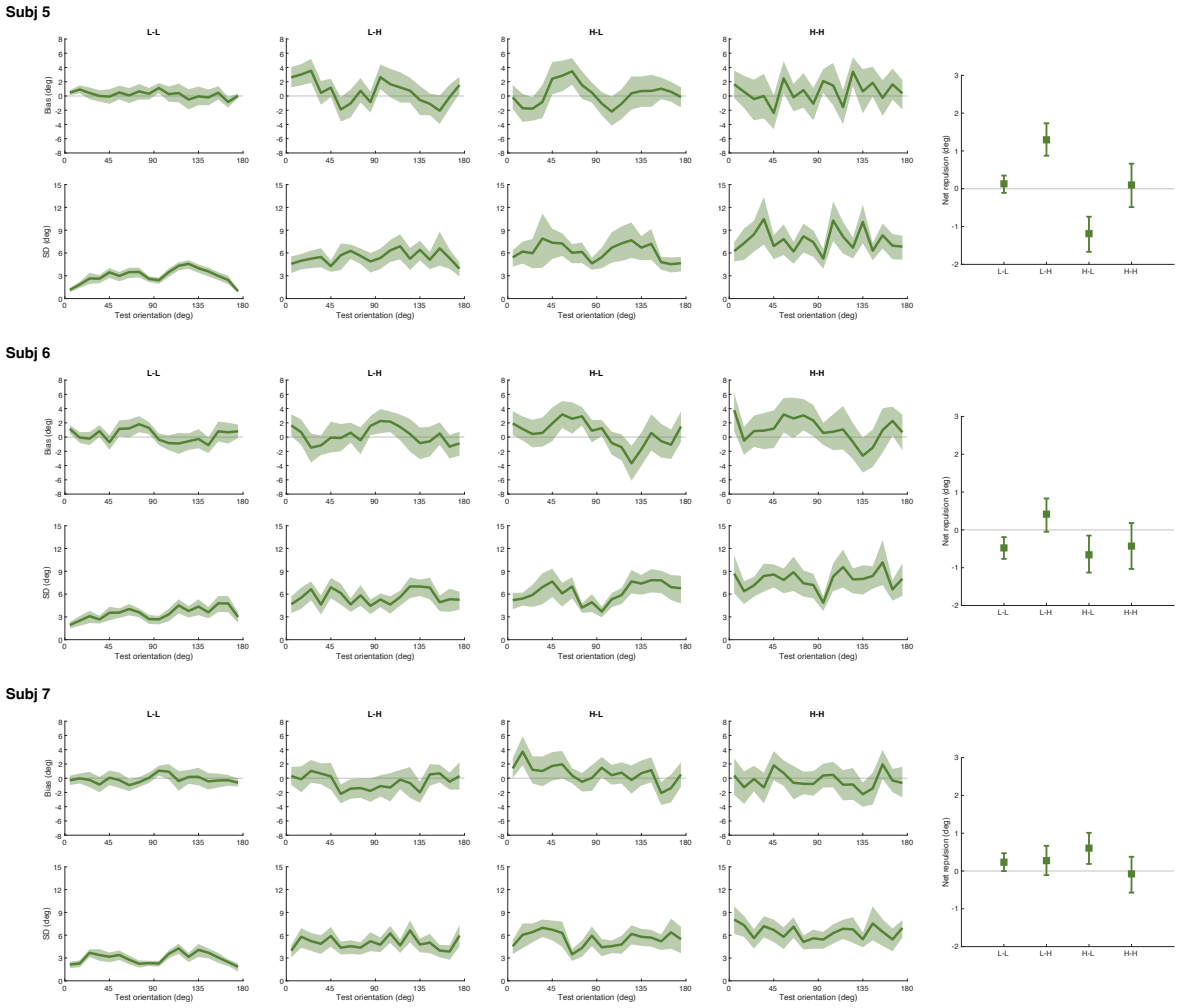

Figure 3.22: Individual subjects' data from the orientation matching experiment: subject 5 – 7. Error bars represent 95% confidence intervals from 1000 bootstrap samples of the data.

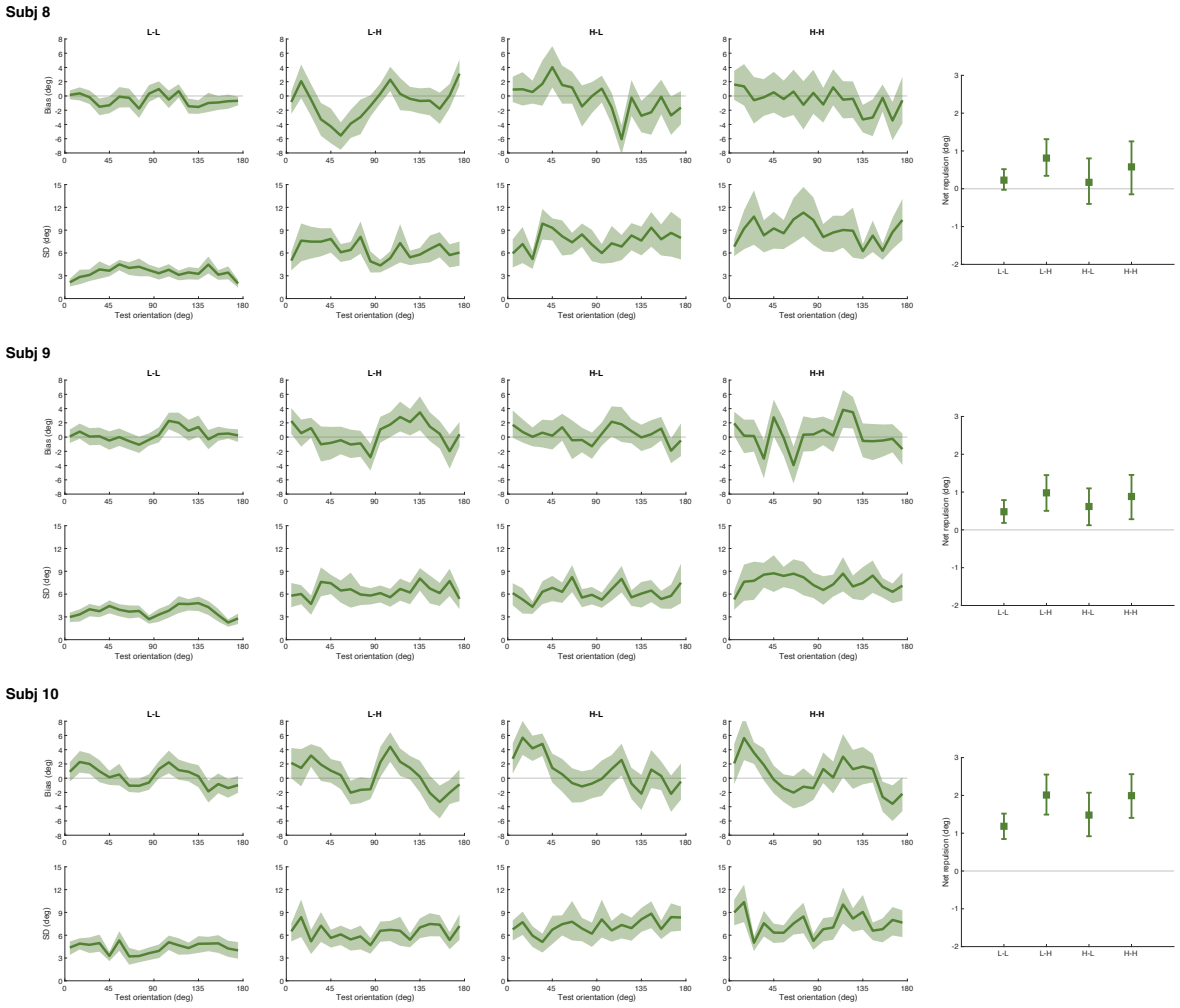

Figure 3.23: Individual subjects' data from the orientation matching experiment: subject 8 – 10. Error bars represent 95% confidence intervals from 1000 bootstrap samples of the data.

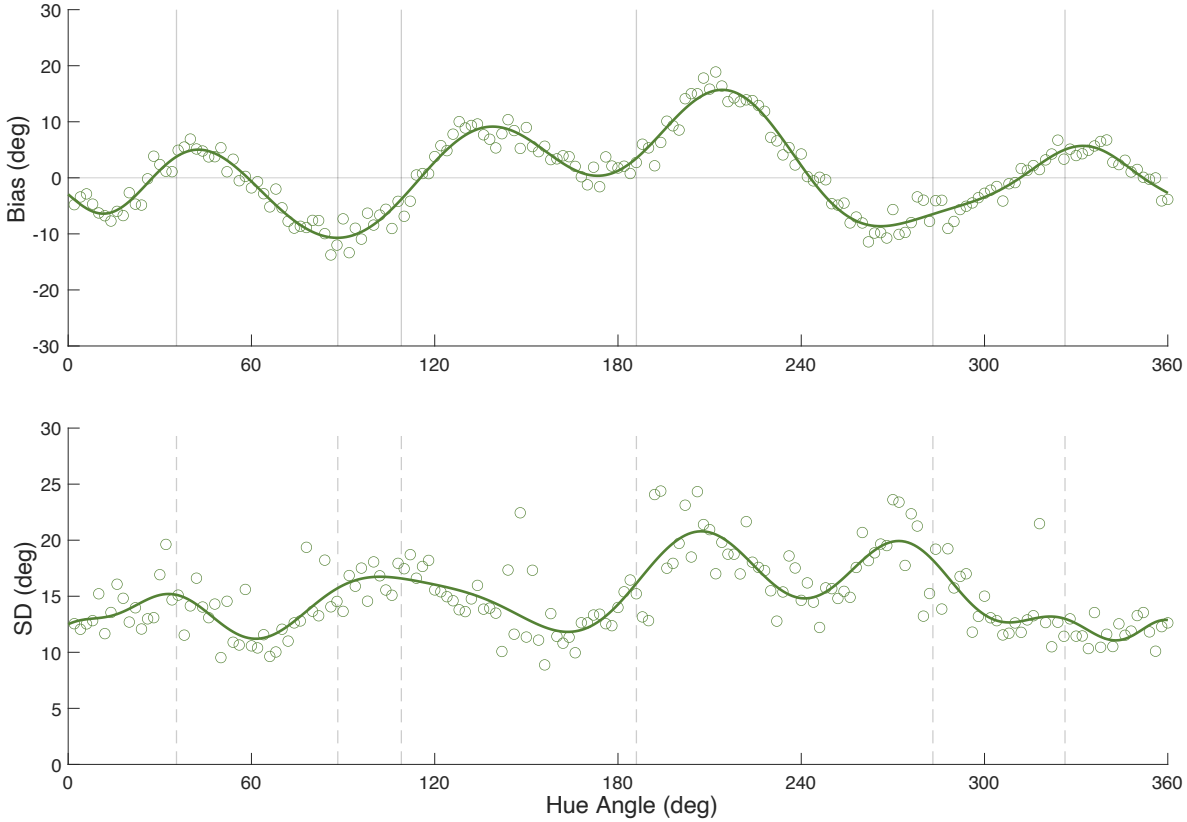

Figure 3.24: Polynomial fit of degree 20 to the bias and standard deviation of subjects' color matching responses in the delayed condition in Bae et al. (2015).

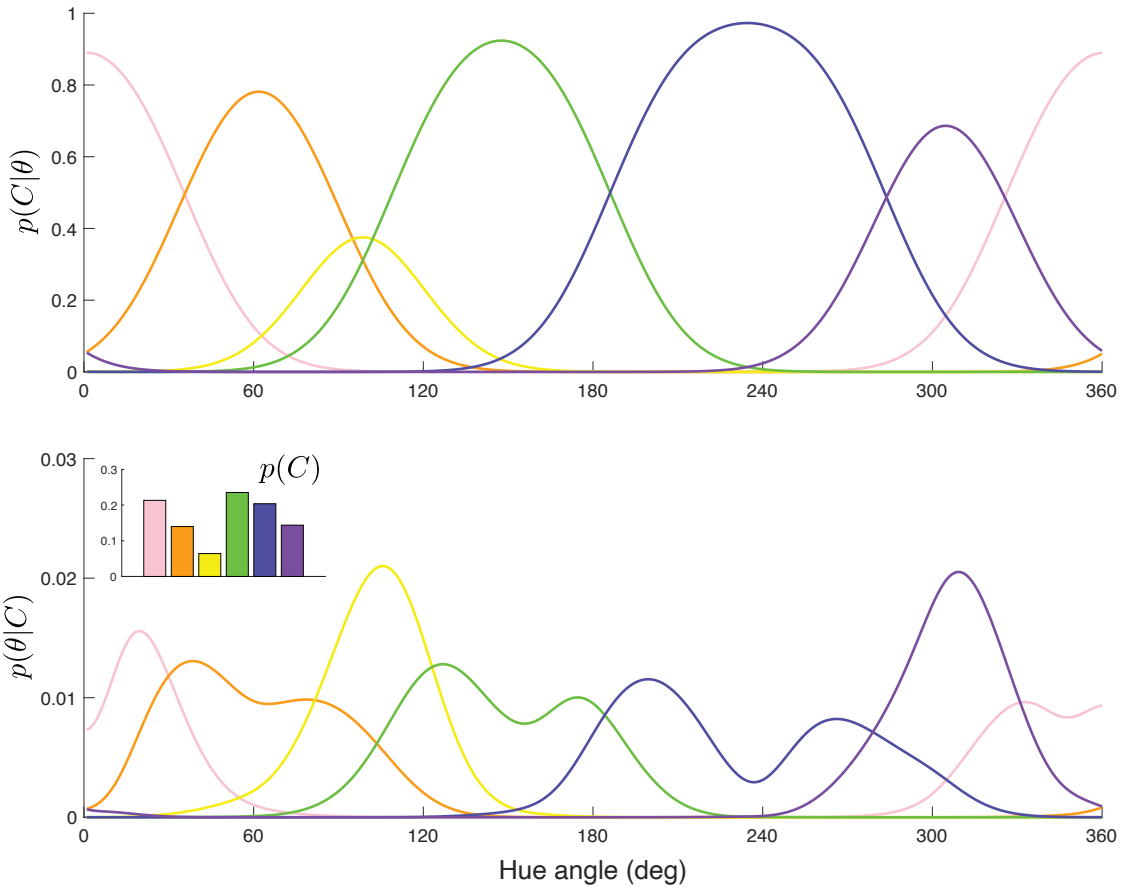

Figure 3.25: Best-fitting color categories.

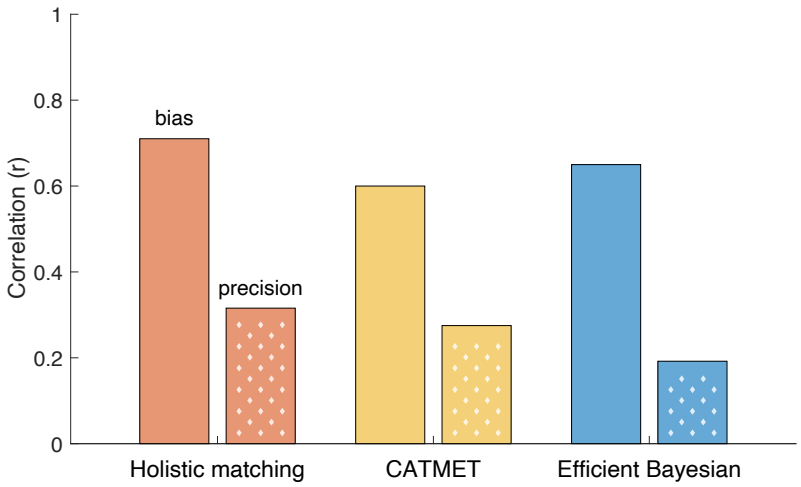

Figure 3.26: Model comparison for color matching data. Correlation values for bias and precision between data and the fit model predictions obtained from the holistic matching model, the four-category CATMET model, and the efficient Bayesian estimator model. Shown are the mean correlation values across the delayed and undelayed conditions. Correlation is used as a measure of model accuracy in order to be able to include the CATMET model in the comparison, using the values indicated in the original paper. Note that we show the values for the best-performing version of the CATMET model that considers only four color categories (Bae et al., 2015). Correlation values are computed as in the original paper.

CHAPTER 4

GENERAL DISCUSSION

In this dissertation, I have investigated how visual perception is influenced by context and the computational principles behind it. In particular, I looked at the perception of orientation under temporal and structural context. The present thesis is rooted in the general hypothesis that the visual system is an ideal system optimized for perception in the natural environment. For temporal context, through a rigorous psychophysical adaptation experiment, analysis of natural scene statistics, and simulation of a recurrent neural network, I demonstrated the dynamic optimality of sensory encoding with temporal context in terms of maximizing the information in the sensory system of the stimulus. For categorical structural context, I developed a holistic matching model that incorporates the categorical context into the optimization goal of the decision process and validated the model with a large body of existing and new psychophysical data, suggesting that structural context might be imposed through holistic processing and active planning. Together, my work provided important insights into the modulation of context on sensory encoding and decoding from a normative point of view.

4.1. Specific contributions

For the temporal context, I ran a psychophysical experiment that cleared up the messy behavioral evidence in the previous literature and characterized the universal adaptation kernel of changes in coding accuracy as described by Fisher information. I tested the efficient coding hypothesis of adaptation by analyzing the natural scene statistics paired with eye-tracking and simulating an adaptation experiment similar to the human psychophysical experiment on a recurrent neural network optimized to predict future video frames. I found that the coding accuracy in human observers increases near and orthogonal to the adaptor and decreases away from the adaptor after adaptation, and the distribution of orientation conditioned on the history mean orientation and the adaptation induced change in Fisher information in the recurrent neural network both support the efficient coding explanation of adaptation.

The efficient coding principle has been a prominent hypothesis in explaining the adaptation effect. However, previous studies have mostly focused on certain predictions derived from the efficient coding hypothesis such as histogram equalization or redundancy reduction, instead of directly verifying the efficient coding hypothesis in terms of maximizing information. One intrinsic difficulty in rigorously testing the efficient coding hypothesis lies in the measurement of neural coding that is comprehensive enough to characterize information distribution. In this work, I bypassed the neural measurement by extracting Fisher information from the behavioral measurement of discrimination threshold. Also, previous studies that measured the adaptation induced changes in orientation discrimination either measured it within a small range surrounding the adaptor orientation, or measured it at one test orientation after adapting to different adaptor orientations. These discrimination measurements are not sufficient to inform us about how adaptation changes the information in the sensory system without the implicit assumption that adaptation effect depends not on the absolute value of the adaptor but only on the relative deviation of the subsequent stimulus. My experiment answered both aspects of this question. I measured the discrimination threshold across the entire range of orientation after adapting to a single orientation in comparison to a properly controlled null-adapted condition, so I was able to extract the changes in coding accuracy across the stimulus range after adaptation. Then I derived a universal parametric description of said coding changes, confirming that the adaptation effect indeed does not depend on the absolute adaptor value.

Artificial neural networks have exploded in popularity in recent years as a means to understand the neural properties and behavioral performance of animals and humans. Artificial system is particularly good for testing hypothesis regarding the building principle of a biological system because one can have full control over how the system is built and build it solely based on the principles under consideration, so any neural properties or behavioral regularities that emerge must be due to these building principles; whereas in real-life pre-existing biological systems, there are always confounding factors or alternative mechanisms that cannot be teased apart. I showed that PredNet, a recurrent neural network built on the predictive coding framework and trained on natural videos to predict the future frames, exhibited similar changes in Fisher information to human observers after adapting to single orientations. This implies that such changes in representations

must be due to the goal of the network to best represent future input combined with the statistical regularities in the training data, which are natural scene videos, supporting the efficient coding hypothesis.

For the structural context, I specifically looked into the categorical context in orientation perception. I proposed a holistic matching model and showed that it can quantitatively explain the response distributions in classic orientation estimation/matching experiments. I also ran a new orientation matching experiment that verified the predictions of the model, which cannot be qualitatively reconciled with non-holistic estimation models. Furthermore, I applied the holistic matching model to color estimation data, showing that this model can be generalized to other features under the influence of categorical effect, or more generally, structural context.

Bayesian ideal observer model has been a successful normative model framework in explaining and predicting perceptual behaviors. However, when the hierarchical context comes into play, the current Bayesian models either collapse the hierarchy and therefore fail to capture the hierarchical effect, or deviate from the normative formulation. By incorporating the categorical loss into the loss function, the holistic matching model both retain a normative formulation and predict the categorical effect. We have shown that this model can explain the bias in not only orientation matching but also color estimation. This model can potentially serve as a general framework for the normative explanation of the categorical effect on perceptual biases.

4.2. Implications on understanding perceptual processes

The efficient coding principle has been a prominent hypothesis in adaptation. My work on the temporal context of perception clearly mapped out the changes in coding accuracy after adaptation, and rigorously confirmed the efficient coding hypothesis from the angle of Fisher information with the analysis of natural scene statistics. The novel approach of using predictive artificial neural network as a model to replicate human behaviors further corroborates the conclusion: adaptation induced changes in coding accuracy emerge from the goal to best represent future sensory input in natural scenes. The results imply that the perceptual system dynamically allocate its coding resources depending on the recent sensory history so that it establishes a representation for future

sensory input that is efficient in the natural scene.

My work on the structural context of perception provide insights into the longstanding discussion of categorical effect. The holistic matching model accurately captures the estimation behavior of multiple features under the influence of categories, answering the question of how the categorical bias occurs. Previously, the biases in estimation or reproduction from memory tasks have been considered a perceptual bias. However, in the holistic matching model, the categorical bias is a result of the categorical loss, which implies that the bias is more of a decision bias or sensorimotor effect rather than a perceptual bias. The post-perceptual nature of the categorical bias is also reflected in the experimental results that repulsive bias persists when the test and the probe have the same noise and when the test and probe stimuli are switched. Our findings provide a fresh angle to reexamine the nature of the categorical bias in features like color and speech sound and a novel model framework of studying the structural context of perception.

My works presented in this thesis are rooted in the popular idea in the field of perception that perception is optimal. My work on temporal and structural context looked into the optimality of encoding and decoding respectively, showing that the encoding process in perception is dynamically efficient depending on recent sensory input history, and the decoding process is a holistic inference process that optimizes for a combined goal across all levels of the hierarchical structure.

4.3. Future directions

One outstanding question in adaptation is how the perceptual bias relates to the encoding changes. Many computational explanations of the perceptual effect of adaptation appeal to "coding catastrophe": downstream decoding mechanism is "unaware" of the changes in the upstream encoding circuit caused by the sensory environment, and simply decode the information the same way as before (Schwartz et al., 2007). However, from a normative point of view, it seems odd that the sensory system should rely on such a suboptimal decoder under ubiquitous sensory adaptation, while the perceptual system in general has been found to be near optimal. The optimal decoder should be an "aware" decoder, which adjusts its computation in accordance with changes in encoding. Now that we have a description of the encoding accuracy before and after adaptation and a model for

orientation estimation under null temporal context both under the same Bayesian framework, it will be straightforward to test the unaware versus aware decoder with an adaptation experiment measuring estimation bias.

The methodology used in this thesis on temporal context can be extended to other contexts, such as spatial context. The spatial contextual effect has many commonalities with the temporal contextual effect (Schwartz et al., 2007). For example, when an oriented grating is surrounded temporally or spatially by a certain orientation, the percept of the grating will be biased in the same way. These effects are called tilt aftereffect (temporal context) and tilt illusion (spatial context). Similar to the temporal context, the efficient coding principle is also a prominent hypothesis in spatial context. Therefore, the efficient coding hypothesis can be tested in the spatial context analogously to the temporal context as in the present research. The influence of surrounding stimuli on coding accuracy can be characterized by the behavioral measurement of discrimination threshold. The efficiency of such spatial influence can be verified in the spatial distribution of stimulus in natural scene (Felsen et al., 2005). Extracting the Fisher information from an artificial neural network trained on natural images might further corroborate the conclusion.

For structural context, the holistic matching model combines efficient coding, holistic Bayesian inference, and a compound loss function. Considering efficient coding with models that describe categorical effects raises an interesting conflict. On the one hand, because discrimination threshold is inversely proportional to coding accuracy, and most categorical features have better discrimination at the categorical boundary than in the middle of a category (Liberman et al., 1957; Kuhl, 1991; Winawer et al., 2007; Etcoff and Magee, 1992), encoding is more accurate at categorical boundary. On the other hand, typical members of a category are usually more common than those near categorical boundaries, leading to an overall stimulus prior that is higher at category centers and lower at boundaries (Feldman et al., 2009; Kronrod et al., 2016; Landy et al., 2017). Thus according to efficient coding, encoding accuracy should also be higher at category centers and lower at boundaries, opposite of what the discrimination performance implies. One way to reconcile this conflict is to take the cost function into account: encoding should be more accurate where the error

is more costly. In the case of categorical perception, with the same amount of feature error, stimulus near the boundary may incur an additional categorical error, so more coding resources should be allocated there. To model the impact of cost function on encoding, we need to come up with a reasonable cost function and develop a model that is optimized for reward as a whole, such as the rate distortion theory based model (Sims et al., 2016) or the neural network model (Schaffner et al., 2023). More future exploration could be done in this direction.

Apart from categorical effect, the holistic matching model might also be generalized to perception under other structural context. For example, the composite face illusion that reflects the holistic processing in face perception can be explained by a weighted integration of information derived from the features and the whole face. When the whole face is different, the judgement about the feature will be biased by the higher-level face representation towards responding "different" while the feature is actually the same.

The interaction between different contexts is also an interesting topic to be explored. When the sensory system adapts to a feature value, how does perception on the higher levels change? Does adaptation also happen on the higher levels? In turn, when we adapt to a high-level feature, how does perception on a lower level change? More generally, how does adaptation on the entire hierarchy interact with each other and influence the percept on each level? For example, face categories have been shown to be influenced by adaptation (Webster et al., 2004). It would be interesting to see whether adaptation on the categorical level influences the lower-level perception of the features of the face. Adapting to a face category may lead to changes in discrimination and bias of faces in that category, but will it also change the discrimination or bias of specific facial features presented on faces in that category as opposed to other category? And if so, how could this process be described by a hierarchical model? These are interesting questions to be answered by future work.

BIBLIOGRAPHY

Laurence Aitchison and Máté Lengyel. With or without you: predictive coding and bayesian inference in the brain. *Current opinion in neurobiology*, 46:219–227, 2017.

DG Albrecht, SB Farrar, and DB Hamilton. Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *The Journal of physiology*, 347(1):713–739, 1984.

Joseph J Atick and A Norman Redlich. Towards a theory of early visual processing. *Neural computation*, 2(3):308–320, 1990.

Fred Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183, 1954.

Stephen A Baccus and Markus Meister. Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36(5):909–919, 2002.

Gi-Yeul Bae, Maria Olkkonen, Sarah R Allred, and Jonathan I Flombaum. Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144(4):744, 2015.

Horace B Barlow et al. Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01), 1961.

Paul M Bays. Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, 34(10):3632–3645, 2014.

James M Beale and Frank C Keil. Categorical effects in the perception of faces. *Cognition*, 57(3): 217–239, 1995.

Ari S Benjamin, Cheng Qiu, Ling-Qi Zhang, Konrad P Kording, and Alan A Stocker. Shared visual illusions between humans and artificial neural networks. In *2019 Conference on Cognitive Computational Neuroscience*, pages 585–588, 2019.

Johannes Bill, Hrag Pailian, Samuel J. Gershman, and Jan Drugowitsch. Hierarchical structure is employed by humans during visual motion perception. *Proceedings of the National Academy of Sciences*, 117(39):24581–24589, 2020a. doi: 10.1073/pnas.2008961117.

Johannes Bill, Hrag Pailian, Samuel J Gershman, and Jan Drugowitsch. Hierarchical structure is employed by humans during visual motion perception. *Proceedings of the National Academy of Sciences*, 117(39):24581–24589, 2020b.

Johannes Bill, Samuel J Gershman, and Jan Drugowitsch. Visual motion perception as online hierarchical inference. *Nature Communications*, 13(1):7403, 2022.

Colin Blakemore and Fergus W Campbell. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of physiology*, 203 (1):237–260, 1969.

Colin Blakemore and Jacob Nachmias. The orientation specificity of two visual after-effects. *The Journal of physiology*, 213(1):157–174, 1971.

Colin Blakemore, Jacob Nachmias, and Peter Sutton. The perceived spatial frequency shift: Evidence for frequency-selective neurones in the human brain. *The Journal of physiology*, 210(3): 727–750, 1970.

David H Brainard and Spatial Vision. The psychophysics toolbox. *Spatial vision*, 10(4):433–436, 1997.

Naama Brenner, William Bialek, and Rob de Ruyter Van Steveninck. Adaptive rescaling maximizes information transmission. *Neuron*, 26(3):695–702, 2000.

Zohar Z Bronfman, Noam Brezis, Rani Moran, Konstantinos Tsetsos, Tobias Donner, and Marius Usher. Decisions reduce sensitivity to subsequent information. *Proceedings of the Royal Society B: Biological Sciences*, 282(1810):20150228, 2015.

Terry Caelli, Hans Brettel, Ingo Rentschler, and Rudi Hilz. Discrimination thresholds in the two-dimensional spatial frequency domain. *Vision research*, 23(2):129–133, 1983.

Andrew J Calder, Andrew W Young, David I Perrett, Nancy L Etcoff, and Duncan Rowland. Categorical perception of morphed facial expressions. *Visual Cognition*, 3(2):81–118, 1996.

Emily Cibelli, Yang Xu, Joseph L Austerweil, Thomas L Griffiths, and Terry Regier. The Sapir-Whorf hypothesis and probabilistic inference: Evidence from the domain of color. *PloS one*, 11 (7):e0158725, 2016.

Colin WG Clifford, Peter Wenderoth, and Branka Spehar. A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267 (1454):1705–1710, 2000.

Colin WG Clifford, Anna Ma Wyatt, Derek H Arnold, Stuart T Smith, and Peter Wenderoth. Orthogonal adaptation improves orientation discrimination. *Vision research*, 41(2):151–159, 2001.

CW Clifford, Derek H Arnold, Stuart T Smith, and Michael Pianta. Opposing views on orthogonal adaptation: a reply to westheimer and gee (2002). *Vision research*, 43(6):717–719, 2003.

David M Coppola, Harriett R Purves, Allison N McCoy, and Dale Purves. The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, 95(7): 4002–4006, 1998.

Tim Crane. The waterfall illusion. *Analysis*, 48(3):142–147, 1988.

Luigi F Cuturi and Paul R MacNeilage. Systematic biases in human heading estimation. *PloS one*, 8(2):e56862, 2013.

Jules Davidoff, Ian Davies, and Debi Roberson. Colour categories in a stone-age tribe. *Nature*, 398 (6724):203–204, 1999.

Vincent De Gardelle, Sid Kouider, and Jerome Sackur. An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. *Journal of Vision*, 10(10):6–6, 2010.

S. Ding, C.J. Cueva, M.V. Tsodyks, and N. Qian. Visual perception as retrospective Bayesian decoding from high- to low-level features. *Proc. National Academies of Sciences U.S.A.*, 114(43): E9115–E9124, 2017. doi: http://dx.doi.org/10.1073/pnas.1706906114.

Valentin Dragoi, Jitendra Sharma, and Mriganka Sur. Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, 28(1):287–298, 2000.

Valentin Dragoi, Jitendra Sharma, Earl K Miller, and Mriganka Sur. Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature neuroscience*, 5(9):883–891, 2002.

Karl Duncker. *A source book of Gestalt psychology*, chapter Induced motion. Kegan Paul, Trench, Trubner & Company, 1938.

Nancy L Etcoff and John J Magee. Categorical perception of facial expressions. *Cognition*, 44(3): 227–240, 1992.

Adrienne L Fairhall, Geoffrey D Lewen, William Bialek, and Robert R de Ruyter van Steveninck. Efficiency and ambiguity in an adaptive neural code. *Nature*, 412(6849):787–792, 2001.

Gustav Theodor Fechner. Elements of psychophysics. *Holt, Rinehart and Winston*, 1966.

Naomi H Feldman, Thomas L Griffiths, and James L Morgan. The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological review*, 116(4):752, 2009.

Gidon Felsen, Jon Touryan, and Yang Dan. Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. *Network: Computation in Neural Systems*, 16(2-3):139–149, 2005.

Matthias Fritsche, Eelke Spaak, and Floris P de Lange. A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife*, 9:e55389, 2020. doi: 10.7554/eLife.55389.

Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The

kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

Samuel J Gershman, Joshua B Tenenbaum, and Frank Jäkel. Discovering hierarchical motion structure. *Vision research*, 126:232–241, 2016.

James J Gibson and Minnie Radner. Adaptation, after-effect and contrast in the perception of tilted lines. i. quantitative studies. *Journal of experimental psychology*, 20(5):453, 1937.

Adam M Gifford, Yale E Cohen, and Alan A Stocker. Characterizing the impact of category uncertainty on human auditory categorization behavior. *PLoS computational biology*, 10(7):e1003715, 2014.

Ahna R Girshick, Michael S Landy, and Eero P Simoncelli. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, 14(7):926–932, 2011.

Valérie Goffaux and Bruno Rossion. Faces are" spatial"–holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human perception and performance*, 32 (4):1023, 2006.

Robert L Goldstone. Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2):178, 1994.

Robert L Goldstone, Yvonne Lippa, and Richard M Shiffrin. Altering object representations through category learning. *Cognition*, 78(1):27–43, 2001.

DiAnne Grieser and Patricia K Kuhl. Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, 25(4):577, 1989.

Bryan L Gros, Randolph Blake, and Eric Hiris. Anisotropies in visual motion perception: a fresh look. *JOSA A*, 15(8):2003–2011, 1998.

Diego A Gutnisky and Valentin Dragoi. Adaptive coding of visual information in neural populations. *Nature*, 452(7184):220–224, 2008.

Stephen T Hammett, Robert J Snowden, and Andrew T Smith. Perceived contrast as a function of adaptation duration. *Vision research*, 34(1):31–40, 1994.

Kyle O Hardman, Evie Vergauwe, and Timothy J Ricker. Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1):30, 2017.

Graham J Hole. Configurational factors in the perception of unfamiliar faces. *Perception*, 23(1): 65–74, 1994.

Janellen Huttenlocher, Larry V Hedges, and Susan Duncan. Categories and particulars: prototype effects in estimating spatial location. *Psychological review*, 98(3):352, 1991.

Mehrdad Jazayeri and J Anthony Movshon. A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446(7138):912–915, 2007.

Mehrdad Jazayeri and Michael N Shadlen. Temporal context calibrates interval timing. *Nature neuroscience*, 13(8):1020, 2010.

David B Kastner, Stephen A Baccus, and Tatyana O Sharpee. Critical and maximally informative encoding between neural populations in the retina. *Proceedings of the National Academy of Sciences*, 112(8):2533–2538, 2015.

Seha Kim and Johannes Burge. The lawful imprecision of human surface tilt estimation in natural scenes. *Elife*, 7:e31448, 2018.

David C Knill and Whitman Richards. *Perception as Bayesian inference*. Cambridge University Press, 1996.

Taisuke Kobayashi, Akiyoshi Kitaoka, Manabu Kosaka, Kenta Tanaka, and Eiji Watanabe. Motion illusion-like patterns extracted from photo and art images using predictive deep neural networks. *Scientific Reports*, 12(1):1–10, 2022.

Adam Kohn. Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of neurophysiology*, 97(5):3155–3164, 2007.

Konrad P Körding and Daniel M Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247, 2004.

Yakov Kronrod, Emily Coppess, and Naomi H Feldman. A unified account of categorical effects in phonetic perception. *Psychonomic bulletin & review*, 23(6):1681–1712, 2016.

Patricia K Kuhl. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, 50(2):93–107, 1991.

David Landy, L Elizabeth Crawford, and Jonathan Corbin. A hierarchical Bayesian model of individual differences in memory for emotional expressions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 39, pages 2518–2523, 2017.

Richard D Lange, Ankani Chattoraj, Jeffrey M Beck, Jacob L Yates, and Ralf M Haefner. A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *PLoS Computational Biology*, 17(11):e1009517, 2021.

Thomas A Langlois, Nori Jacoby, Jordan W Suchow, and Thomas L Griffiths. Serial reproduction

reveals the geometry of visuospatial representations. *Proceedings of the National Academy of Sciences*, 118(13), 2021.

Simon Laughlin. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung c*, 36(9-10):910–912, 1981.

Simon B Laughlin. The role of sensory adaptation in the retina. *Journal of Experimental Biology*, 146(1):39–62, 1989.

Alvin Liberman, Katherine Safford Harris, Peter Eimas, Leigh Lisker, and Jarvis Bastian. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 4(4):175–195, 1961.

Alvin M Liberman, Katherine Safford Harris, Howard S Hoffman, and Belver C Griffith. The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, 54(5):358, 1957.

William Lotter, Gabriel Kreiman, and David Cox. Deep predictive coding networks for video prediction and unsupervised learning. *arXiv preprint arXiv:1605.08104*, 2016.

Long Luu and Alan A Stocker. Post-decision biases reveal a self-consistency principle in perceptual inference. *Elife*, 7:e33334, 2018.

Long Luu and Alan A Stocker. Categorical judgments do not modify sensory representations in working memory. *PLOS Computational Biology*, 17(6):e1008968, 2021.

L Maffei, A Fiorentini, and S Bisti. Neural correlate of perceptual adaptation to gratings. *Science*, 182(4116):1036–1038, 1973.

Donald E Mitchell and Darwin W Muir. Does the tilt after-effect occur in the oblique meridian? *Vision research*, 16(6):609–613, 1976.

Wiktor F Młynarski and Ann M Hermundstad. Adaptive coding for dynamic sensory inference. *Elife*, 7:e32055, 2018.

Wiktor F Młynarski and Ann M Hermundstad. Efficient and adaptive sensory codes. *Nature Neuroscience*, 24(7):998–1009, 2021.

James R Muller, Andrew B Metha, John Krauskopf, and Peter Lennie. Rapid adaptation in visual cortex to the structure of images. *Science*, 285(5432):1405–1408, 1999.

Jean-Paul Noel, Ling-Qi Zhang, Alan A Stocker, and Dora E Angelaki. Individuals with autism spectrum disorder have altered visual encoding capacity. *PLoS biology*, 19(5):e3001215, 2021.

Richard A Normann and I Perlman. The effects of background illumination on the photoresponses

of red and green cones. *The Journal of Physiology*, 286(1):491–507, 1979.

Izumi Ohzawa, Gary Sclar, and RALPH D Freeman. Contrast gain control in the cat's visual system. *Journal of neurophysiology*, 54(3):651–667, 1985.

Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

Brian O'Toole and Peter Wenderoth. The tilt illusion: Repulsion and attraction effects in the oblique meridian. *Vision research*, 17(3):367–374, 1977.

Carlyn A Patterson, Stephanie C Wissig, and Adam Kohn. Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. *Journal of Neuroscience*, 33(2):532–543, 2013.

Raymond E Phinney, Christopher Bowd, and Robert Patterson. Direction-selective coding of stereoscopic (cyclopean) motion. *Vision research*, 37(7):865–869, 1997.

Rafael Polania, Michael Woodford, and Christian C. Ruff. Efficient coding of subjective value. *Nature Neuroscience*, 22(1):134–142, 2019. doi: 10.1038/s41593-018-0292-0.

Arthur Prat-Carrabin and Michael Woodford. Efficient coding of numbers explains decision bias and noise. *bioRxiv*, 2021. doi: 10.1101/2020.02.18.942938.

Michael S Pratte, Young Eun Park, Rosanne L Rademaker, and Frank Tong. Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1):6, 2017.

C. Qiu, L. Luu, and A. A. Stocker. Benefits of commitment in hierarchical inference. *Psychological Review*, 127(4):622–639, 2020. doi: https://doi.org/10.1037/rev0000193.

Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.

Hans-Jürgen Rauber and Stefan Treue. Reference repulsion when judging the direction of visual motion. *Perception*, 27(4):393–402, 1998.

D Regan and KI Beverley. Spatial-frequency discrimination and detection: comparison of postadaptation thresholds. *JOSA*, 73(12):1684–1690, 1983.

D Regan and KI Beverley. Postadaptation orientation discrimination. *JOSA A*, 2(2):147–155, 1985.

Gillian Rhodes, Susan Brake, and Anthony P Atkinson. What's lost in inverted faces? *Cognition*, 47(1):25–57, 1993.

Luke J Rosielle and Eric E Cooper. Categorical perception of relative orientation in visual object recognition. *Memory & Cognition*, 29(1):68–82, 2001.

Bruno Rossion. Picture-plane inversion leads to qualitative changes of face perception. *Acta psychologica*, 128(2):274–289, 2008.

Bruno Rossion. The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, 21(2):139–253, 2013.

Jonathan Schaffner, Sherry Dongqi Bao, Philippe N Tobler, Todd A Hare, and Rafael Polania. Sensory perception relies on fitness-maximizing codes. *Nature Human Behaviour*, pages 1–17, 2023.

Paul R Schrater and Eero P Simoncelli. Local velocity representation: evidence from motion adaptation. *Vision research*, 38(24):3899–3912, 1998.

Odelia Schwartz, Anne Hsu, and Peter Dayan. Space and time in visual context. *Nature Reviews Neuroscience*, 8(7):522–535, 2007.

Peggy Seriès, Alan A Stocker, and Eero P Simoncelli. Is the homunculus "aware" of sensory adaptation? *Neural Computation*, 21(12):3271–3304, 2009.

Tatyana O Sharpee, Adam J Calhoun, and Sreekanth H Chalasani. Information theory of adaptation in neurons, behavior, and mood. *Current opinion in neurobiology*, 25:47–53, 2014.

Eero P Simoncelli and William T Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Proceedings., International Conference on Image Processing*, volume 3, pages 444–447. IEEE, 1995.

Chris R Sims, Zheng Ma, Sarah R Allred, Rachel A Lerch, and Jonathan I Flombaum. Exploring the cost function in color perception and memory: An information-theoretic model of categorical effects in color matching. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 38, pages 2273–2278, 2016.

Yosef Singer, Yayoi Teramoto, Ben DB Willmore, Jan WH Schnupp, Andrew J King, and Nicol S Harper. Sensory cortex is optimized for prediction of future input. *elife*, 7:e31557, 2018.

Nikolaos Smyrnis, Asimakis Mantas, and Ioannis Evdokimidis. Two independent sources of anisotropy in the visual representation of direction in 2-D space. *Experimental brain research*, 232(7):2317–2324, 2014.

A.A. Stocker and E.P. Simoncelli. A Bayesian model of conditioned perception. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems NIPS 20*, pages 1409–1416, Cambridge, MA, December 2007. MIT Press.

Alan A Stocker and Eero Simoncelli. Constraining a bayesian model of human visual speed perception. *Advances in neural information processing systems*, 17, 2004.

Alan A Stocker and Eero P Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4):578–585, 2006.

Alan A Stocker and Eero P Simoncelli. Visual motion aftereffects arise from a cascade of two isomorphic adaptation mechanisms. *Journal of Vision*, 9(9):9–9, 2009.

Dominik Straub and Constantin A. Rothkopf. Looking for image statistics: Active vision with avatars in a naturalistic virtual environment. *Frontiers in Psychology*, 12, 2021. ISSN 1664-1078. doi: 10.3389/fpsyg.2021.641471.

Robert Taylor and Paul M Bays. Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *Journal of Neuroscience*, 38(32):7132–7142, 2018.

Alessandro Tomassini, Michael J Morgan, and Joshua A Solomon. Orientation uncertainty reduces perceived obliquity. *Vision research*, 50(5):541–547, 2010.

Ruben S van Bergen and Janneke FM Jehee. Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *Journal of Neuroscience*, 39(41):8164–8176, 2019.

Ruben S Van Bergen, Wei Ji Ma, Michael S Pratte, and Janneke FM Jehee. Sensory uncertainty decoded from visual cortex predicts behavior. *Nature neuroscience*, 18(12):1728–1730, 2015.

Ronald Van den Berg, Hongsup Shin, Wen-Chuang Chou, Ryan George, and Wei Ji Ma. Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109(22):8780–8785, 2012.

K Vinken, X Boix, and G Kreiman. Incorporating intrinsic suppression in deep neural networks captures dynamics of adaptation in neurophysiology and perception. *Science Advances*, 6(42): eabd4205, 2020.

Martin J Wainwright. Visual adaptation as optimal information transmission. *Vision research*, 39 (23):3960–3974, 1999.

Masumi Wakita. Categorical perception of orientation in monkeys. *Behavioural processes*, 67(2): 263–272, 2004.

Z. Wang, A.A. Stocker, and D.D. Lee. Efficient neural codes that minimize L_p reconstruction error. *Neural Computation*, 28(12):2656–2686, December 2016. doi: 10.1162/NECO_a_00900.

Eiji Watanabe, Akiyoshi Kitaoka, Kiwako Sakamoto, Masaki Yasugi, and Kenta Tanaka. Illusory motion reproduced by deep neural networks trained for prediction. *Frontiers in psychology*, page 345, 2018.

Michael A Webster and Eriko Miyahara. Contrast adaptation and the spatial structure of natural images. *Josa a*, 14(9):2355–2366, 1997.

Michael A Webster, Daniel Kaping, Yoko Mizokami, and Paul Duhamel. Adaptation to natural facial categories. *Nature*, 428(6982):557–561, 2004.

X.-X. Wei and A. A. Stocker. Mutual information, Fisher information, and efficient coding. *Neural Computation*, 28(2):305–326, February 2016.

X.-X. Wei and A.A. Stocker. Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference. In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems NIPS 25*, pages 1313–1321. MIT Press, December 2012.

Xue-Xin Wei and Alan A Stocker. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature neuroscience*, 18(10):1509, 2015.

Xue-Xin Wei and Alan A Stocker. Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences*, 114(38):10244–10249, 2017.

Xue-Xin Wei, Pedro Ortega, and Alan Stocker. Perceptual adaptation: getting ready for the future. *Journal of Vision*, 15(12):388–388, 2015.

Gerald Westheimer and Angela Gee. Orthogonal adaptation and orientation discrimination. *Vision Research*, 42(20):2339–2343, 2002.

Jonathan Winawer, Nathan Witthoft, Michael C Frank, Lisa Wu, Alex R Wade, and Lera Boroditsky. Russian blues reveal effects of language on color discrimination. *Proceedings of the national academy of sciences*, 104(19):7780–7785, 2007.

Christoph Witzel and Karl R Gegenfurtner. Categorical sensitivity to color differences. *Journal of vision*, 13(7):1–1, 2013.

Jeremy M Wolfe, Stacia R Friedman-Hill, Marion I Stewart, and Kathleen M O'Connell. The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1):34, 1992.

Sichao Yang, Johannes Bill, Jan Drugowitsch, and Samuel J Gershman. Human visual motion perception shows hallmarks of bayesian structural inference. *Scientific reports*, 11(1):3714, 2021.

Andrew W Young, Deborah Hellawell, and Dennis C Hay. Configurational information in face perception. *Perception*, 16(6):747–759, 1987.