AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

2024, Vol. 131, No. 4, 858–890
https://doi.org/10.1037/rev0000457

Sensory Perception Is a Holistic Inference Process

Jiang Mao and Alan A. Stocker
Department of Psychology, University of Pennsylvania

Sensory perception is widely considered an inference process that reflects the best guess of a stimulus feature based on uncertain sensory information. Here we challenge this reductionist view and propose that perception is rather a *holistic inference process* that operates not only at the feature but jointly across all levels of the representational hierarchy. We test this hypothesis in the context of a commonly used psychophysical matching task in which subjects are asked to report their perceived orientation of a test stimulus by adjusting a probe stimulus (method-of-adjustment). We introduce a holistic matching model that assumes that subjects' reports reflect an optimal match between the test and probe stimulus, both in terms of their inferred feature (orientation) and also their higher level representation (orientation category). Validation against several existing data sets demonstrates that the model accurately and comprehensively predicts subjects' response behavior and outperforms previous models both qualitatively and quantitatively. Moreover, the model generalizes to other feature domains and offers an alternative account for categorical color perception. Our results suggest that categorical effects in sensory perception are ubiquitous and can be parsimoniously explained as optimal behavior based on holistic sensory representations.

Keywords: Bayesian observer, efficient coding, El Greco fallacy, not reductionism, natural scene statistics

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

A quantitative validation of the Bayesian estimation framework crucially depends on an accurate specification of these prior beliefs. Visual orientation is one of the few stimulus features for which prior beliefs can be well specified in form of a probability distribution that reflects the local orientation statistics of 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), which is seemingly "anti-Bayesian" because the higher prior probabilities for cardinal orientations would rather predict a bias toward those orientations. However, the efficient coding hypothesis (Attneave, 1954; Barlow, 1961) provides a powerful constraint on sensory uncertainty to resolve this apparent paradox, leading to a consistent Bayesian interpretation of visual orientation perception (Wei & 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 not only offer an accurate model for orientation perception (Taylor & Bays, 2018; Wei & Stocker, 2015) but to provide a unifying account for human behavior in a wide variety of other cognitive tasks (e.g., Fritsche et al., 2020; Langlois et al., 2021; Ni & Stocker, 2023; Polania et al., 2019; Prat-Carrabin & Woodford, 2021; Taylor & Bays, 2018; Zhang & Stocker, 2022).

Despite its promise, however, there are several reasons to question this model's ability to provide a unifying account of sensory perception. First, a full quantitative validation of the model against data is still outstanding; previous studies mainly focused on

This article was published Online First February 15, 2024.
Alan A. Stocker https://orcid.org/0000-0002-2041-1515
The authors thank the members of the Computational Perception and Cognition Laboratory for many fruitful discussions and feedback. They thank Rafael Polania for helpful comments on an early version of the article. This work has been supported in part by the National Science Foundation of the United States of America (Award IIS-1912232 to Alan A. Stocker) and in part by the University of Pennsylvania.

All data in this article have been previously published and were obtained directly from the corresponding authors. Matlab code for model simulations is available at https://github.com/cpc-lab-stocker/Holistic-matching-model. This study has not been preregistered. Preliminary results have been presented at the Annual Meeting of the Vision Science Society 2021 and are described in a preprint (bioRxiv, https://doi.org/10.1101/2022.06.24.497534).

Jiang Mao played a lead role in data curation, formal analysis, software, visualization, and writing–original draft and an equal role in conceptualization, investigation, methodology, validation and writing–review and editing. Alan A. Stocker played a lead role in funding acquisition, project administration, resources, and supervision, a supporting role in writing–original draft, and an equal role in conceptualization, formal analysis, investigation, methodology, validation, visualization, and writing–review and editing.

Correspondence concerning this article should be addressed to Alan A. Stocker, Department of Psychology, University of Pennsylvania, Goddard Laboratories, 3710 Hamilton Walk, Philadelphia, PA 19106, United States. Email: astocker@psych.upenn.edu

summary statistics such as estimation bias (Wei & Stocker, 2015). Some studies have found reasonable fits to subjects' orientation reports (Bays, 2014; Pratte et al., 2017; Van den Berg et al., 2012) both in terms of bias and variance (Taylor & Bays, 2018), while other data sets using similar stimuli show substantially different bias/variance patterns that are difficult to reconcile with the existing efficient Bayesian estimator (De Gardelle et al., 2010; Noel et al., 2021). More importantly, the model struggles to even qualitatively account for some existing psychophysical data sets. Specifically, Tomassini et al. (2010) reported the results of a typical orientation-matching experiment where subjects were asked to report the orientation of a test stimulus by adjusting a probe stimulus. In half of the trials, however, the stimuli used as tests and probes were interchanged. If subjects' percepts of the test and probe orientations were to reflect independent estimates, one would expect a pair of perceptually matched stimuli to be matched under both conditions, yielding opposite estimation errors when the assignment of test and probe is switched. However, subjects' matching behavior did not show a flip of the bias pattern in those trials. This suggests that matching behavior is not based on independent perceptual estimates and therefore rules out any observer model that reduces perception to the process of estimating the stimulus feature. Last, but not least, there is the long-standing notion that higher level, categorical representations influence perception at the feature level (see Goldstone & Hendrickson, 2010, for a review). Several studies have suggested that perception of visual orientation is affected by a cardinal/oblique category distinction (Durgin & Li, 2011; Rosielle & Cooper, 2001; Wakita, 2004). The efficient Bayesian estimator does not provide the possibility to formally incorporate categorical effects unless the orientation prior implicitly reflects the categories, that is, has peaks at orientations that correspond to the category centers (e.g., Bae et al., 2015). Applying such compound orientation priors, however, violates the normative Bayesian assumption that the prior distribution reflects the statistical distribution of visual orientations.

Here, we introduce a hierarchical inference model of perception that resolves all these issues. We assume perception to be intrinsically holistic, such that the percept of a stimulus is *jointly represented* by the inference outcomes (i.e., the posteriors) at every level of a representational hierarchy. For example, in the specific context of orientation perception, this implies that perceived orientation is represented as the result of inference at both the feature (orientation) as well as at higher level representations (e.g., orientation categories). Importantly, the model assumes that cognitive processes downstream of perception (e.g., a decision stage) operate on these holistic perceptual representations and are not reduced to computations at only a single level of the representational hierarchy. This fundamentally separates our proposal from any previous model.

We tested our hypothesis in the context of a typical psychophysical matching task in which subjects are asked to report the perceived orientation of a test stimulus by adjusting a probe stimulus. We introduce a holistic matching model that provides a highly accurate account of four different existing data sets: the model not only correctly predicts that the bias pattern does not flip when test and probe stimuli are switched (see above; Tomassini et al., 2010) but also provides a superior quantitative account for the full error distributions of subjects' perceptual reports in other experimental studies, including those that show different bias/variance patterns (Bays, 2014; De Gardelle et al., 2010; Noel et al., 2021). Finally, we also demonstrate that our model presents an alternative explanation for "categorical perception" by providing an accurate account of human behavioral data in a color-matching experiment typically thought of being affected by color categories (Bae et al., 2015).

Holistic Matching Model

Perception of a stimulus feature is commonly assessed with a psychophysical matching task often referred to as "the method of adjustment." In this task, a subject is asked to adjust, for example, the orientation of a probe stimulus in order to match the perceived orientation of a test stimulus (Figure 1a). Typically, the probe stimulus is unambiguous and noiseless, leading to the general assumption that the reported probe orientation is a direct reflection of the subject's perceived test stimulus orientation, aside from some potential corruption with motor noise. Previous studies have suggested that the efficient Bayesian estimator provides a qualitatively accurate account of subjects' reported probe orientations in these matching experiments (Taylor & Bays, 2018; Wei & Stocker, 2015, 2017).

The proposed holistic matching model builds on the idea of efficient Bayesian inference. However, it assumes a hierarchical generative process where each stimulus orientation θ is associated with a category C distinguishing cardinal and oblique orientations (Figure 1b). What distinguishes it from previous hierarchical models is that perceptual inference is performed at all levels of the hierarchy. The outcome is thus a holistic, probabilistic representation of the perceived orientation stimulus, jointly represented by the posteriors at both the orientation and the category levels, $p(\theta|m)$ and $p(C|m)$, respectively. A 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 (Figure 1c). We express this as finding the probe orientation that minimizes a weighted average of the expected mismatch (loss) at the orientation (L_θ) and the category levels (L_c), thus

$$L_{\text{tot}}(\theta, \theta_p, C, C_p) = (1 - w)L_\theta(\theta, \theta_p) + wL_c(C, C_p), \qquad (1)$$

where $0 < w < 1$ is the relative contribution of the categorical mismatch. Finally, we assume that subjects' reported probe orientations θ_p^* represent noisy samples of the optimally matched probe orientation θ_p due to some additive motor noise.

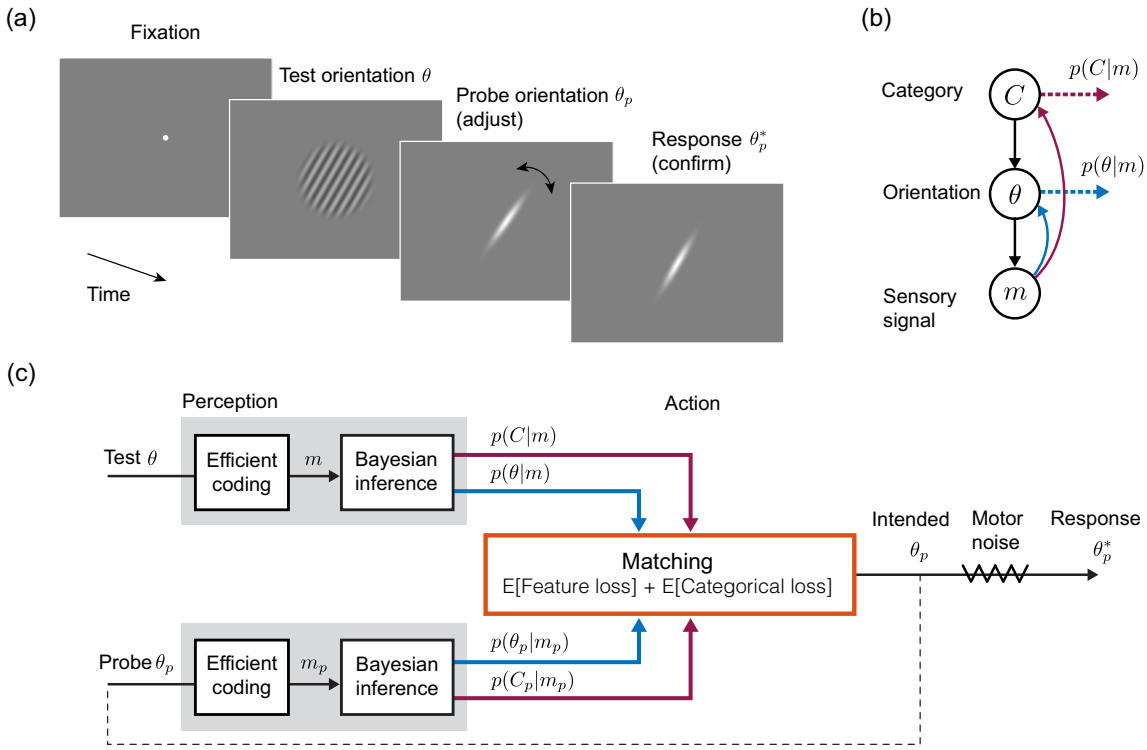
Model Details

In the following, we provide a more detailed description of the perception and action components of the holistic matching model outlined in Figure 1c.

Efficient Coding and Feature Inference

We follow Wei and Stocker (2015) in formulating feature inference that is constrained by efficient coding. 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

Figure 1
Holistic Perceptual Matching



Note. (a) Typical psychophysical matching task to characterize visual orientation perception (often referred to as "method-of-adjustment"). 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) The holistic matching model assumes that both the test θ and the probe θ_p orientations are efficiently encoded according to the prior distribution $p(\theta)$ (Wei & Stocker, 2015), resulting in sensory measurements m and m_p, respectively. Bayesian inference (according to the generative model in (b)) 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, the model computes the probe orientation that optimally matches the test orientation. Note, that a nonholistic 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 noiseless. See the online article for the color version of this figure.

measurement (approximated by Fisher information, Wei & 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

$$p(\theta) \propto \sqrt{J(\theta)}. \tag{2}$$

Sensory noise: We consider a sensory space in which Fisher information is uniform (i.e., the sensory noise is uniform). Efficient coding (Equation 2) defines the optimal mapping $\tilde{\theta} = F(\theta)$ from stimulus to this sensory space to be 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 a homogeneous likelihood function in sensory space $p(\tilde{m}|\tilde{\theta})$ obtained by assuming uniform sensory noise according to a von Mises distribution,

$$p(\tilde{m}|\tilde{\theta}) = \mathrm{vm}(\tilde{m}; \tilde{\theta}, \kappa_i), \tag{3}$$

with κ_i representing the sensory noise magnitude.

Stimulus noise (for the experiment by Tomassini et al., 2010): We assume that the test stimulus in each trial reflects a noisy sample θ′ of the test orientation θ drawn from a von Mises distribution,

$$p(\theta'|\theta) = \mathrm{vm}(\theta'; \theta, \kappa_e), \tag{4}$$

where κ_e represents the constant 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 Equation 3, hence,

$$p(\tilde{m}|\tilde{\theta'}) = \mathrm{vm}(\tilde{m}; \tilde{\theta}', \kappa_i). \quad (5)$$

The distribution of the sensory measurement m in stimulus space is

$$p(m|\theta') = p(\tilde{m}|\tilde{\theta'})F'(m), \quad (6)$$

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

$$p(m|\theta) = \int p(m|\theta')p(\theta'|\theta)\, d\theta'. \quad (7)$$

Finally, based on the generative model (Figure 1b) the posterior over stimulus orientation given the sensory measurement is

$$p(\theta|m) \propto p(m|\theta)\sum_i p(\theta|C_i)p(C_i) \propto p(m|\theta)p(\theta), \quad (8)$$

where the orientation prior $p(\theta)$ represents the natural orientation statistics (Figure 2a).

Categorical Inference

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

$$p(C = \text{'H'}|\theta; \mu_H) = \alpha \frac{\mathrm{vm}(\theta; \mu_H, \kappa_c)}{\mathrm{vm}(\mu_H; \mu_H, \kappa_c)}, \quad (9)$$

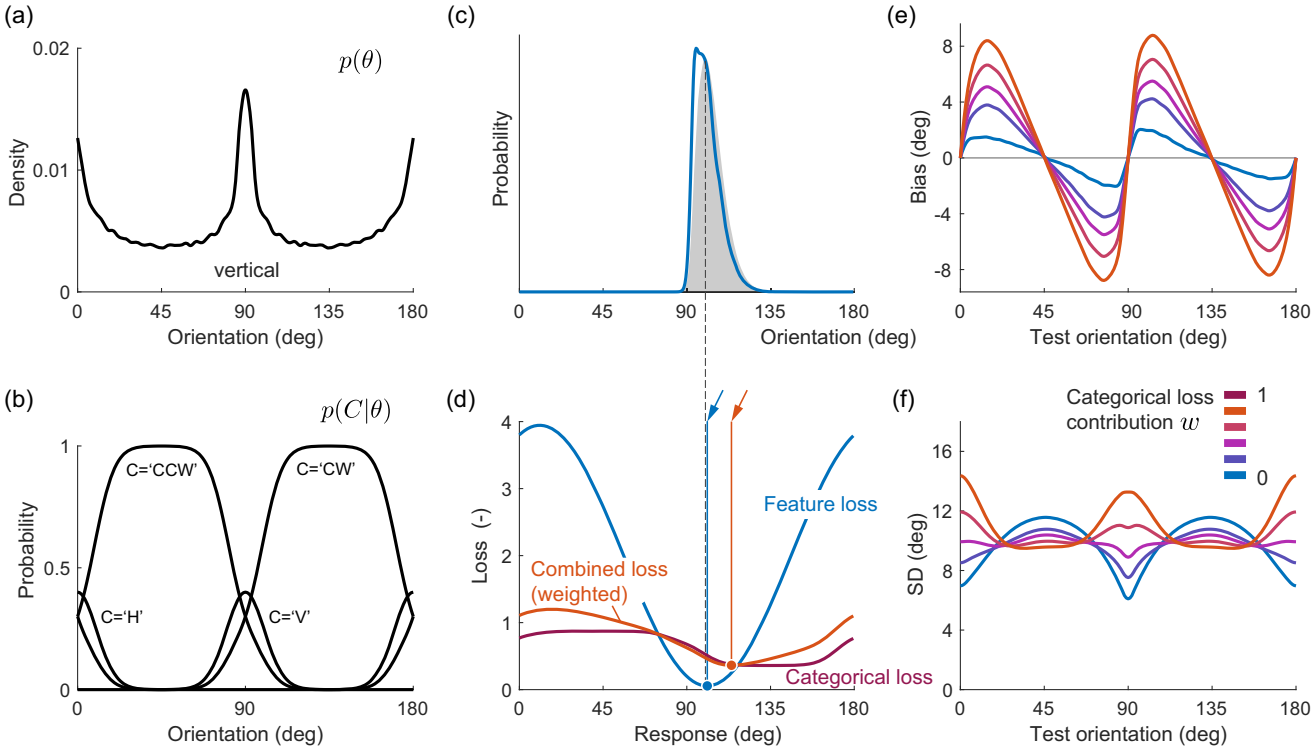
where α is the probability of the horizontal category at μ_H, κ_c represents the uncertainty in the categorical boundaries, and μ_H represents a noisy signal of the horizontal orientation that may stochastically vary across trials according to

$$p(\mu_H) = \mathrm{vm}(\mu_H; 0\,\mathrm{deg}, \kappa_c). \quad (10)$$

The vertical category is similarly defined with its center μ_V always 90 deg away from μ_H. The oblique categories are the orientations in

Figure 2

Model Simulations for Matching Task With Noiseless Probe (Typical Condition)



Note. (a) Prior distribution $p(\theta)$ used for model simulations and fits throughout this article. It reflects the statistics of local visual orientation averaged across natural indoor and outdoor scenes (Coppola et al., 1998). (b) Category structure $p(C|\theta)$ assumed by the holistic matching model. Together with the prior $p(\theta)$, it defines $p(C)$ and $p(\theta|C)$ of the generative model (Figure 1b). (c) Likelihood (shaded area) and posterior (blue curve) of test orientation for a given sensory measurement m (dashed line). (d) Expected feature (blue), categorical (purple), and total loss (orange) given m, and the optimal match predicted by the efficient Bayesian estimator (blue arrow) and the holistic matching model (orange arrow). (e) Predicted bias pattern and (f) standard deviation in probe responses as a function of w. Note that for $w = 0$, the model is equivalent to the efficient Bayesian estimator when the probe stimulus is noiseless (blue curves). See the online article for the color version of this figure.

between the cardinal categories with a smooth transition given by the cumulative von Mises distributions,

$$p(\text{'CCW'}|\theta;\mu_H) = (\text{cum vm}(\theta;\mu_H,\kappa_c) - \text{cum vm}(\theta;\mu_V,\kappa_c)) \times (1 - p(\text{'H'}|\theta) - p(\text{'V'}|\theta)), \quad (11)$$

and

$$p(\text{'CW'}|\theta;\mu_H) = (\text{cum vm}(\theta;\mu_V,\kappa_c) - \text{cum vm}(\theta;\mu_H,\kappa_c)) \times (1 - p(\text{'H'}|\theta) - p(\text{'V'}|\theta)), \quad (12)$$

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 (Figure 1b), the posterior probability over category C can be computed as follows:

$$\begin{aligned} p(C|m;\mu_H) &= \frac{1}{p(m)}\int_\theta p(m|\theta)p(\theta|C;\mu_H)p(C) \\ &= \frac{1}{p(m)}\int_\theta p(m|\theta)p(\theta)p(C|\theta;\mu_H) \\ &= \int_\theta p(C|\theta;\mu_H)p(\theta|m). \end{aligned} \quad (13)$$

Matching

We assume that participants adjust the probe stimulus while obtaining continuous visual feedback about the probe orientation. 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 has been optimally adjusted (Figure 1b).

Matching is assumed to minimize the total loss L_{tot} consisting of a weighted sum of the feature and the categorical loss (Equation 1). We define feature loss to be the cosine of the difference between the probe and the test orientation, thus

$$L_\theta(\theta,\theta_p) = 2(1 - \cos(2(\theta - \theta_p))). \quad (14)$$

This loss function 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 categorical loss is defined as whether the category of the probe is different from the category of the test orientation C or not, thus,

$$L_c(C,C_p) = \begin{cases} 0 & \text{if } C_p = C \\ 1 & \text{otherwise.} \end{cases} \quad (15)$$

Given the sensory measurements m and m_p, the expected total loss is

$$\begin{aligned} E\left[L_{\text{tot}}|m,m_p;\mu_H\right] &= (1-w)\iint L_\theta(\theta,\theta_p)p(\theta|m)p(\theta_p|m_p)\,d\theta\,d\theta_p \\ &\quad + w\sum_{C_0\in\mathbb{C}} p(C=C_0|m;\mu_H) \times (1 \\ &\quad - p(C_p = C_0|m_p;\mu_H)). \end{aligned} \quad (16)$$

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

$$\begin{aligned} E\left[L_{\text{tot}}|m,\theta_p;\mu_H\right] &= (1-w)\int L_\theta(\theta,\theta_p)p(\theta|m)d\theta \\ &\quad + w\sum_{C_0\in\mathbb{C}} p(C=C_0|m;\mu_H) \\ &\quad \times (1 - p(C_p = C_0|\theta_p;\mu_H)). \end{aligned} \quad (17)$$

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

$$\hat{\theta}_p(m;\mu_H) = \arg\min_{\theta_p} E[L_{\text{tot}}|m,\theta_p;\mu_H]. \quad (18)$$

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,

$$\hat{m}_p(m;\mu_H) = \arg\min_{m_p} E\left[L_{\text{tot}}|m,m_p;\mu_H\right]. \quad (19)$$

Predicted Response Distribution

Given the optimal probe response for a single sensory measurement (Equations 18 and 19, respectively), we can now calculate the predicted response distribution for a given test orientation θ.

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 Equation 7 and is not affected by any nonuniform prior assumption, hence,

$$p(\hat{\theta}_p|\hat{m}_p(m;\mu_H)) \propto p(\hat{m}_p(m;\mu_H)|\hat{\theta}_p). \quad (20)$$

Because $\hat{m}_p(m;\mu_H)$ is a deterministic function (Equation 19), we can rewrite the probability distribution as follows:

$$p(\hat{\theta}_p|m;\mu_H) = p(\hat{\theta}_p|\hat{m}_p(m;\mu_H)). \quad (21)$$

For a noiseless probe stimulus, Equation 21 turns into a Dirac delta distribution at the optimal $\hat{\theta}_p$ (Equation 18).

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

$$p(\hat{\theta}_p^*|\hat{\theta}_p) = \text{vm}(\hat{\theta}_p^*;\hat{\theta}_p,\kappa_m), \quad (22)$$

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 follows:

$$p(\hat{\theta}_p^*|\theta) = \iiint p(\hat{\theta}_p^*|\hat{\theta}_p)p(\hat{\theta}_p|m;\mu_H)p(m|\theta)p(\mu_H)\,d\hat{\theta}_p\,dm\,d\mu_H, \tag{23}$$

with the terms in the integral given by Equations 22, 21, 7, and 10, respectively.

Nonholistic Matching Model

For comparison, we also consider the nonholistic version of the matching model. It is identical to the holistic model but does not consider categorical inference (Figure 1b). Furthermore, 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 (Equation 23) except that it is not dependent on category noise μ_H.

If the probe stimulus is noiseless, the nonholistic matching model is equivalent to the *efficient Bayesian estimator* (Wei & 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 (Equation 14), aside from some additive motor noise. The efficient Bayesian estimator shares the same efficient feature encoding as the holistic matching model. In contrast, the *standard Bayesian estimator* assumes homogeneous encoding such that the sensory measurements m given the stimulus sample θ' follow the von Mises distribution,

$$p(m|\theta') = \mathrm{vm}(m;\theta',\kappa_i), \tag{24}$$

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

Model Simulations

Simulations illustrate how and why the predictions of the holistic matching model differ from those of the efficient Bayesian estimator (Figure 2). In order to reduce model complexity, we constrain the prior distribution for visual orientation $p(\theta)$ to reflect the statistics of local orientations in natural scenes. These statistics are 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 manmade objects typically show less pronounced peaks at the cardinals compared to indoor scenes (Coppola et al., 1998; Straub & Rothkopf, 2021; see Appendix Figure C2). We use the average distribution across both indoor and outdoor scenes measured by Coppola et al. (1998) as the fixed orientation prior $p(\theta)$ for all simulations and fits presented in the article (Figure 2a).

Furthermore, we consider four orientation categories: vertical ("V"), horizontal ("H"), clockwise ("CW"), or counterclockwise ("CCW") oblique relative to vertical (Figure 2b). Uncertainty associated with the categorical representation is expressed in overlapping categorical distributions as well as in noisy centers of the 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 Appendix Figures C3 and C4).

Efficient encoding leads to likelihood functions that have long tails away from the nearest cardinal orientation (Wei & Stocker, 2015). Figure 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 toward 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 & Stocker, 2015). Figure 2d illustrates this by plotting the feature loss L_θ and the matching percept represented by the point of minimal loss (blue arrow). The point of minimal loss, however, is different when considering the combined loss L_{tot} of the holistic matching model. The point of minimal category loss L_c does not coincide with the location of the minimal feature loss L_θ but is shifted toward the center of the most probable category of the test stimulus. This results in a larger repulsive bias away from the category boundaries.

With all else equal, both the magnitude of the bias (Figure 2e) as well as the pattern in the variability of the predicted probe reports (Figure 2f) depend on the relative contribution w of the categorical loss to the total loss. Compared to the efficient Bayesian observer, the holistic matching model generally predicts larger repulsive biases for the same level of stimulus uncertainty. Increasing values of w predict larger bias magnitudes but also a change of the variability pattern such that the loci of largest variability switch from being at oblique (small w) to being at cardinal orientations (large w). Note that the efficient Bayesian estimator ($w = 0$) always predicts the smallest variability to occur at cardinal orientations, independent of the level of sensory or stimulus uncertainty.

Empirical Validation

Matching Experiments With a Noiseless Probe Stimulus

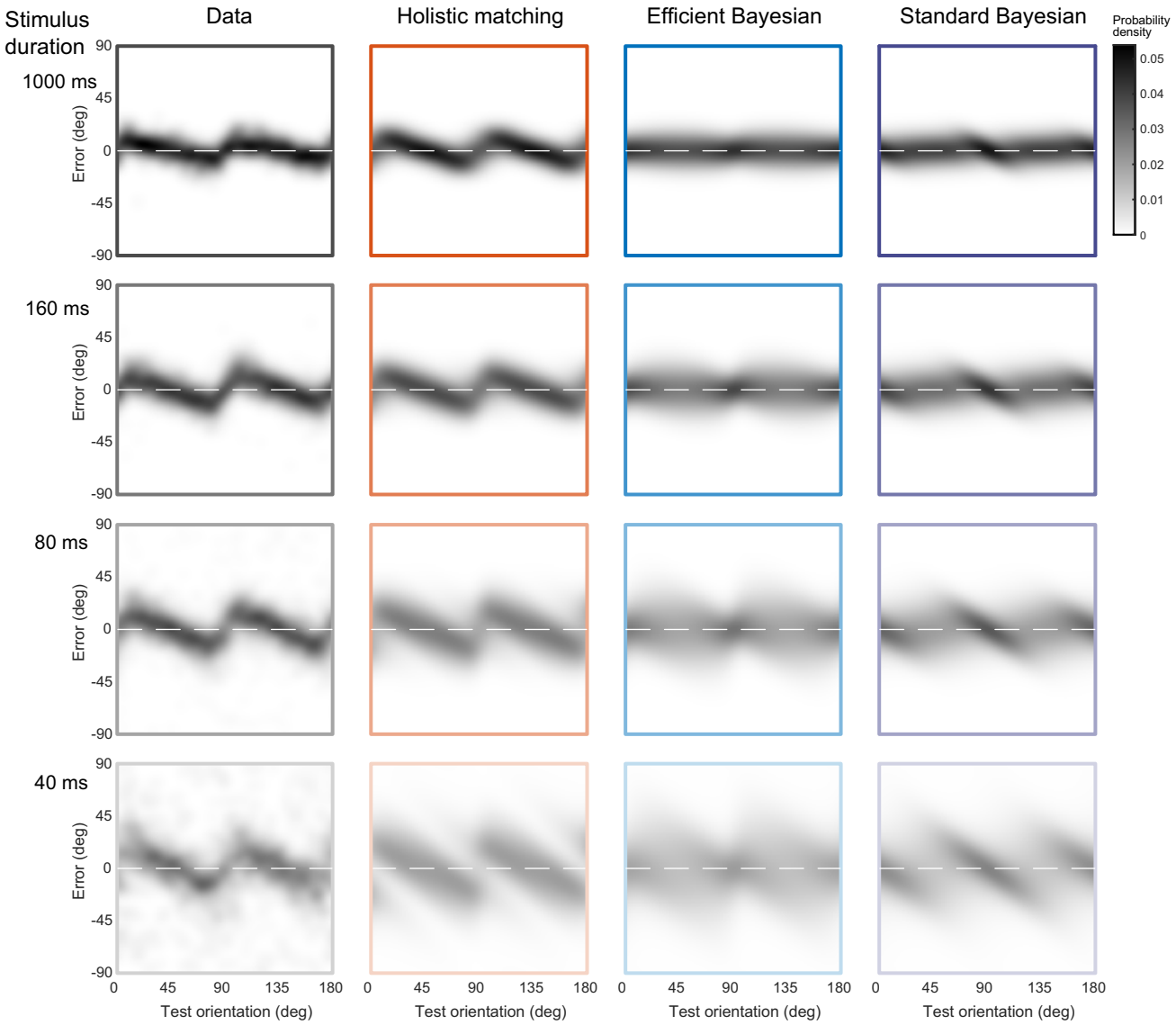
We first validated the holistic matching model in the typical, noiseless probe condition against three different existing data sets. All three data sets are from matching experiments that share the same general experimental design (see Figure 1a), yet allow us to test different aspects of our model.

Data by De Gardelle et al. (2010): Effect of Sensory Noise

In the experiment of this study, human subjects were asked to report the orientation of a briefly presented peripheral Gabor patch (test stimulus) by adjusting the orientation of a subsequently shown probe stimulus. Sensory noise of the test stimulus was modulated by varying its presentation duration (see Appendix A, for a more detailed description of the experiment). Figure 3 shows the full error distributions of the combined subject data for each of the four presentation durations. Note that the number of trials per subject and test orientation were too low to reliably analyze data from individual subjects. The distributions exhibit the characteristic repulsive bias away from cardinal orientations and show no apparent asymmetry between the two cardinal orientations (Wei & Stocker, 2015). Bias and variability increase with decreasing stimulus presentation duration, which is a fundamental characteristic of Bayesian perception.

Figure 3

Data and Model Fits for Matching Task With Noiseless Probe (De Gardelle et al., 2010)

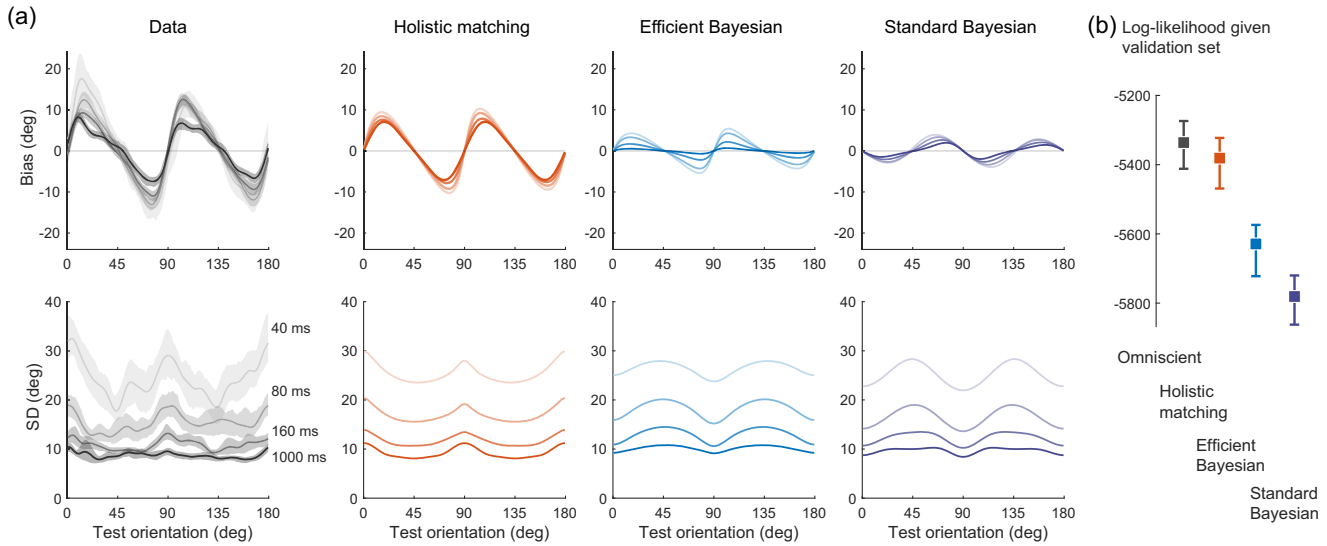


Note. Shown are the error distributions of the matching responses for different test stimulus durations (combined subject). Columns show the data and the corresponding best fit model predictions, respectively. Data distributions show clear repulsive biases away from the cardinal orientations. Bias and 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 the online article for the color version of this figure.

We fit the holistic matching model as well as its nonholistic variants (i.e., the efficient Bayesian estimator and the standard Bayesian estimator with homogeneous sensory encoding) to the response distribution data. We assumed the probe stimulus to be noiseless as it consisted of a Gabor patch with one visible strip that was continuously present until subjects confirmed their choice. All models use the same formulation of the feature loss L_θ and the fixed natural orientation prior (Figure 2a). The holistic matching model fully captures the entire shape of the error distributions across all noise conditions, which is not the case for the two Bayesian estimators (Figure 3). Their predicted error distributions are mostly centered around zero and show substantially larger variability for oblique than for cardinal orientations. Furthermore, as expected, the standard Bayesian estimator predicts attractive bias near cardinal orientations. In general, both estimation models exhibit distribution patterns that are substantially different from the data. The differences are evident when comparing the mean and standard deviation of the error distributions with the predictions of the models

Figure 4

Model Comparison and Cross-Validation (De Gardelle et al., 2010)



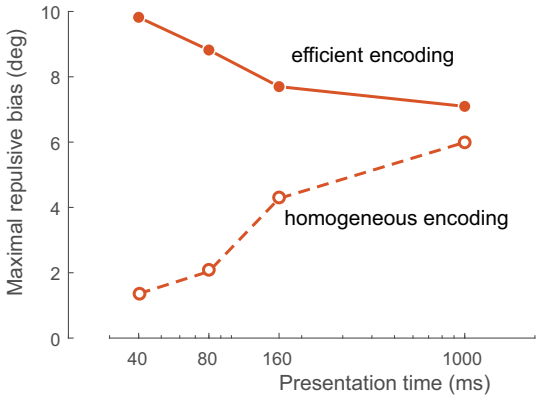
Note. (a) Mean (bias) and standard deviation (*SD*) of the error distributions shown in Figure 3. Subjects exhibit increasing repulsive bias with decreasing presentation duration. The holistic matching model fits both the pattern and the magnitude of the bias. The bias predicted by the efficient Bayesian model is smaller than the observed bias. In contrast, the standard Bayesian model predicts attractive bias. 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. Shaded areas represent 95% confidence intervals from 100 bootstrap runs. (b) 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 standard Bayesian observer model. The "omniscient" model is an empirical model that uses the data distribution in the training set as a predictor of the validation data using optimal kernel density estimation (see Appendix B). See the online article for the color version of this figure.

(Figure 4a). Human subjects exhibit repulsive bias away from cardinal orientations, 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 when minimizing the categorical loss. This additional categorical bias offsets the smaller sensory variability at cardinal orientations due to efficient coding.

We used cross-validation for a quantitative comparison of the models. Cross-validation intrinsically corrects for differences in model complexity (i.e., number of parameters); overly complex models that overfit the training data typically score low in predicting the test 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 for the test data using kernel density estimation (see Appendix B). The omniscient observer serves as an upper bound representing the best possible statistical prediction of the test set given the training set. As shown in Figure 4b, the holistic matching model predicts the data substantially better than the efficient and standard Bayesian model. Its performance is almost at the level of the omniscient model with error bars that largely overlap. Cross-validation confirms that the holistic matching model provides an excellent account of the data with a model complexity that does not lead to overfitting.

How important is the efficient encoding assumption of the model? To answer this question, we compared the predictions of the fit holistic matching model with and without efficient sensory coding (Figure 5). Without efficient sensory encoding (i.e., assuming uniform sensory accuracy), the model predicts increasing bias magnitudes with increasing stimulus presentation time, which is opposite to the pattern seen in the data (Figure 4a). Bias in our model is modulated by sensory noise via three different processes: 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 (Figure 2c). At the same time, the difference in category posterior probability of the test stimulus decreases, leading to a flatter categorical loss curve (see Figure 2d). Both effects lead to less repulsive bias as the sensory noise increases, which is the outcome shown in Figure 5 (dashed line). Efficient coding introduces repulsive biases via asymmetric likelihood functions that have long tails away from the peak of the prior at cardinal orientations (Wei & Stocker, 2015). However, larger sensory noise

Figure 5
Efficient Sensory Encoding



Note. 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), which is consistent with the data. With homogeneous coding and all else equal, however, it predicts the opposite pattern. See the online article for the color version of this figure.

leads to larger asymmetry in the likelihood function and therefore larger repulsive biases. Thus, efficient coding is the one component of the model that causes larger repulsive biases with higher sensory noise. As such, efficient sensory encoding is an indispensable assumption of the holistic matching model for an accurate account of the data.

Data by Noel et al. (2021): Individual Subject Differences

Figure 6a shows validation against data from another recent study, investigating the differences in orientation perception between individuals with an autism spectrum disorder and individuals from a neurotypical control group (Noel et al., 2021). The study used a similar experimental design as in De Gardelle et al. (2010) and found similar behavior signatures such as the shape of the error distribution, the relatively large bias magnitude, and the higher standard deviation at cardinal compared to oblique orientations. The holistic matching model well accounts for the data from both groups, suggesting that Bayesian inference as such is not compromised in autistic individuals (Noel et al., 2020).

Because of its larger numbers of trials per subject and condition, this data set allows us to extend model validation to individual subject data. Appendix Figures C5–C8 show the measured estimation biases and variability together with the predictions of the individually best fit model for all 42 subjects across both groups. There are substantial individual differences that are well accounted for by the model. Figure 6b shows the distributions of the best fit model parameters across all subjects. A comparison of the fit model parameters indicates that the difference between the two groups is mainly limited to a difference in sensory noise (autism spectrum disorder: higher sensory noise).

Data by Bays (2014): Inverted Variability Pattern

Subjects do not always show the exact same behavior that we saw in the first two data sets. Some studies found that subjects' response variability can be lower for cardinal compared to oblique test orientations (Taylor & Bays, 2018). Note that it is important to disambiguate variability from discriminability since the latter is always lower at oblique orientations, which is typically referred to as the "oblique effect" (Appelle, 1972).

We thus validated our model against a third data set for which subjects' response variability was generally lower at cardinal compared to oblique orientations (Bays, 2014). Subjects performed a working memory experiment where they had to recall the orientation of a set of line elements after a memory delay using a probe stimulus (see Appendix A, for a more detailed description of the experiment). As shown in Figure 7, subjects' error distributions of their recalled orientations exhibit similar repulsive biases at cardinal orientations as in the two previous data sets (e.g., Figure 6). Variability in subjects' reported orientations, however, is indeed lowest at cardinal orientations. Nonetheless, the proposed holistic matching model provides a detailed and accurate account also for this data set. Comparison of the fit parameters shows that the main difference between the fits of the two previous data sets is the reduced relative weight w of the categorical loss (Appendix A and Appendix Table B1), which is in line with the initial simulations (Figure 2). This also explains why a previous study suggested that the efficient Bayesian estimator ($w = 0$) provides a reasonable fit to this data set (see Taylor & Bays, 2018). A closer comparison between the two model fits, however, reveals that the holistic matching model provides a better account of human behavior, in particular with regard to the magnitude of the bias and the variability around cardinal orientations (Figure 7b).

Validation across the three different data sets demonstrates that the holistic matching model provides a parsimonious explanation for characteristically different human orientation-matching behavior at both the group and the individual subject levels.

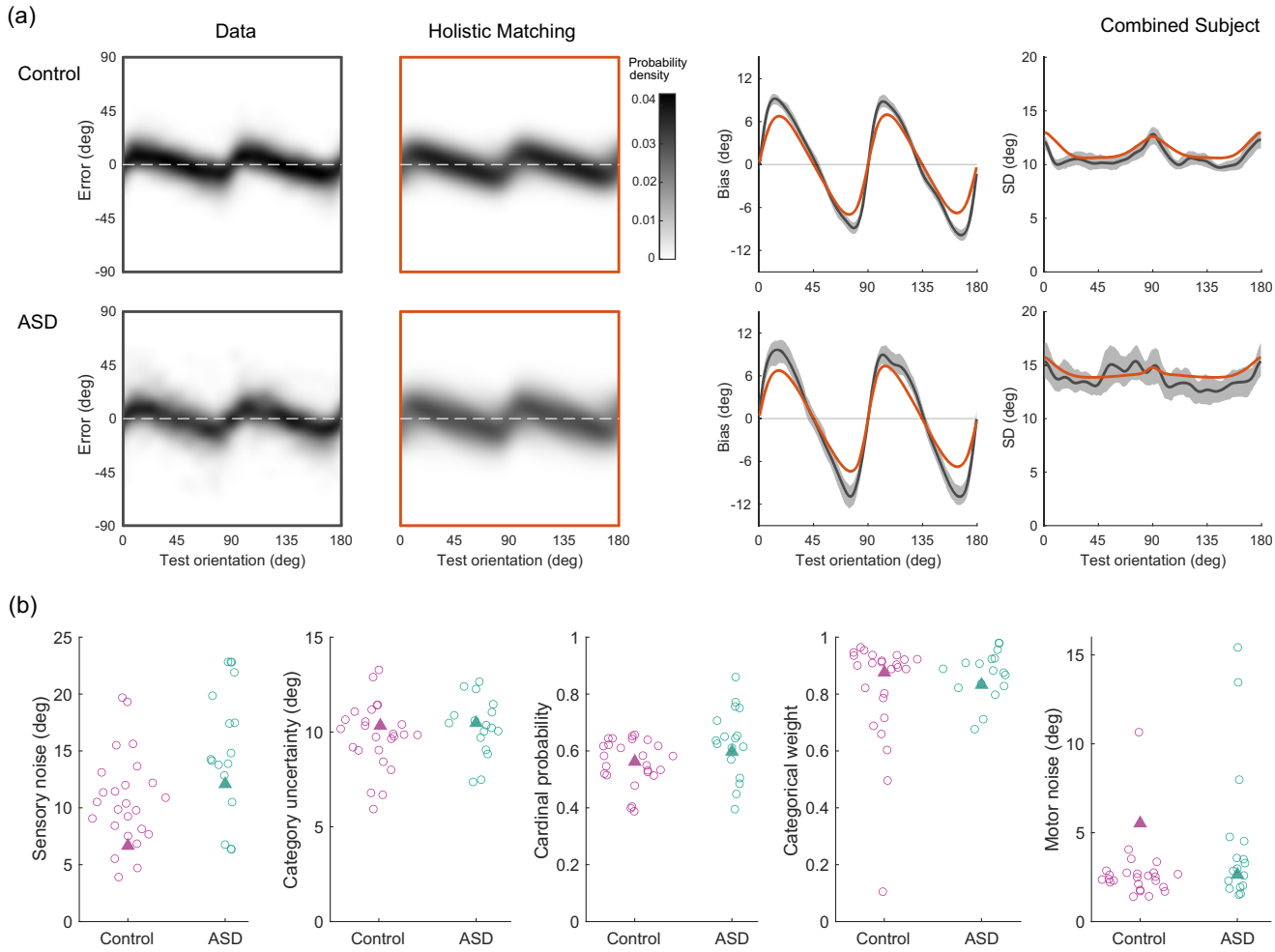
Matching Experiment With Noisy Probe Stimulus

In most perceptual matching experiments, such as the experiments discussed above, the probe stimulus is unambiguous and noiseless, and the percept of the probe stimulus is typically considered to be veridical. However, the holistic matching model explicitly captures the perception of the probe stimulus and thus can make predictions for far more general experimental conditions (Figure 1b).

In the following, we consider the case where sensory uncertainties in the test and probe stimuli are reversed. Any model that compares the test and the probe stimuli only at the feature level would predict a reversal of the bias pattern when the roles of the two stimuli are switched. While the matching stimulus configuration (i.e., the orientations of the two stimuli for which they perceptually match) is identical, the roles of the test and probe are switched and thus lead to a bias that is flipped (Figure 8a). However, the holistic matching model makes a qualitatively different prediction. Because matching also operates at the category level, adjusting the probe orientation toward the

Figure 6

Data and Model Fits for Matching Task With Noiseless Probe in Neurotypical and Autistic (ASD) Subjects (Noel et al., 2021)



Note. (a) Error distributions, bias and standard deviation, and the corresponding predictions of the best fit holistic matching model (combined subjects). Data are from the control experiment (no feedback), for which both subject groups have been identified to have prior expectations that match the assumed prior distribution $p(\theta)$. Shaded areas represent 95% confidence intervals from 100 bootstrap runs. (b) Distributions of model parameters obtained from model fits to individual subjects' data in both groups. Filled triangles indicate parameters for the combined subjects shown in (a). ASD = autism spectrum disorder. See the online article for the color version of this figure.

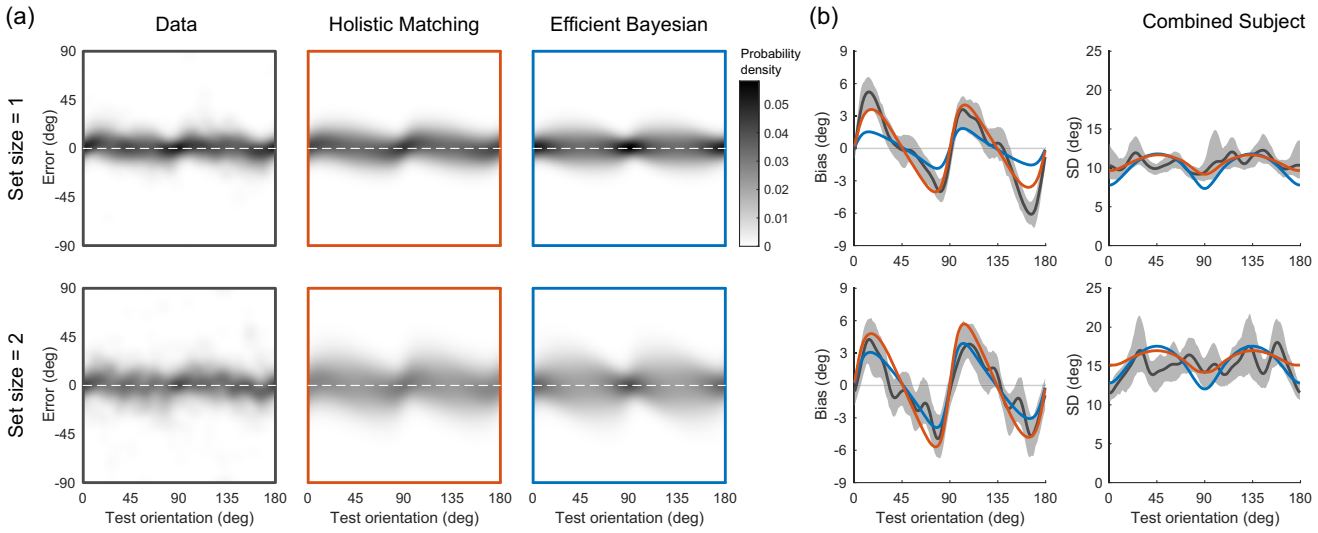
center of the most probable category of the test stimulus always reduces the expected categorical loss L_c. This leads to a stable repulsive bias pattern whether the test and probe stimulus are switched or not. Thus, the model does not predict a flip of the bias (Figure 8b). More generally, it predicts that the matching stimulus configuration depends on which stimulus is adjusted.

Tomassini et al. (2010) performed an orientation-matching experiment where test and probe stimuli were interchanged (see Appendix A, for a detailed description of the experiment). 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 white dots (noiseless probe) such that it matched the mean orientation of the Gabor patches (see Figure 8a). Throughout the trial, both test and probe stimuli were constantly present. 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 Figure 9. As predicted by the holistic matching model, the biases did not flip and are indeed repulsive under both conditions. We performed a joint model fit to the data across all conditions. The model well predicts 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. 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 (Figure 9a). When the probe stimulus is noisy, the bias is smaller in

Figure 7

Data and Model Fits for Matching Task With Noiseless Probe in a Working Memory Experiment (Bays, 2014)



Note. Data represent the combined subject shown for set sizes 1 and 2. (a) Error distributions and (b) bias and standard deviation; and the corresponding best fit model predictions of both the holistic matching model and the efficient Bayesian estimator. The error distributions and the magnitude of the bias indicate that the distortions in subjects' reports are smaller than in the two previous data sets. Note also that the standard deviation shows the opposite pattern compared to the two previous data sets, with smaller variability at the cardinal compared to oblique orientations. The holistic matching model captures all of these characteristics well and consistently outperforms the efficient Bayesian estimator. Shaded areas represent 95% confidence intervals from 100 bootstrap runs. See the online article for the color version of this figure.

the small noise condition compared to the large noise condition, which matches the pattern in the data (Figure 9b). The standard deviation predicted by the model is for the 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 nonholistic estimation model.

Categorical Color Perception

Finally, we extend validation to color perception, commonly considered to be categorical (Cibelli et al., 2016; Hardman et al., 2017; Witzel & Gegenfurtner, 2013). We use the data set from a previous study, in which subjects performed two color-matching experiments reporting the color (hue) of either a previously (delayed condition) or a simultaneously presented test color (Bae et al., 2015; see Appendix A, for experimental details).

Validation requires us to first specify the color category structure and the color prior. Bae et al. (2015) initially ran a color-naming experiment where subjects were asked to select the name that best described the test color out of a set of basic color names. Appendix Figure C10a shows subjects' probabilities for choosing each color name given a test color. We extract the category structure from these naming probabilities. Following the original study, we assume six color categories ($C \in \mathbb{C} = \{C_1, C_2, \ldots, C_6\}$). 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

$$\mu_j = b_j + \Delta\mu \ (j = 1, 2, \ldots, 6), \tag{25}$$

and the deviation follows a von Mises distribution

$$p(\Delta\mu) = \mathrm{vm}(\Delta\mu; 0, \kappa_b), \tag{26}$$

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

$$p(\hat{C} = C_j|\theta) = \mathrm{cum\ vm}(\theta; b_j, \kappa_b) - \mathrm{cum\ vm}(\theta; b_{j+1}, \kappa_b). \tag{27}$$

We fit this probability to the color-naming data to obtain κ_b and b_j ($j = 1, 2, \ldots, 6$). As with orientation, we assume that color categories overlap according to cumulative von Mises distributions,

$$p(C_j|\theta; \Delta\mu) = \mathrm{cum\ vm}(\theta; \mu_j, \kappa_c) - \mathrm{cum\ vm}(\theta; \mu_{j+1}, \kappa_c), \tag{28}$$

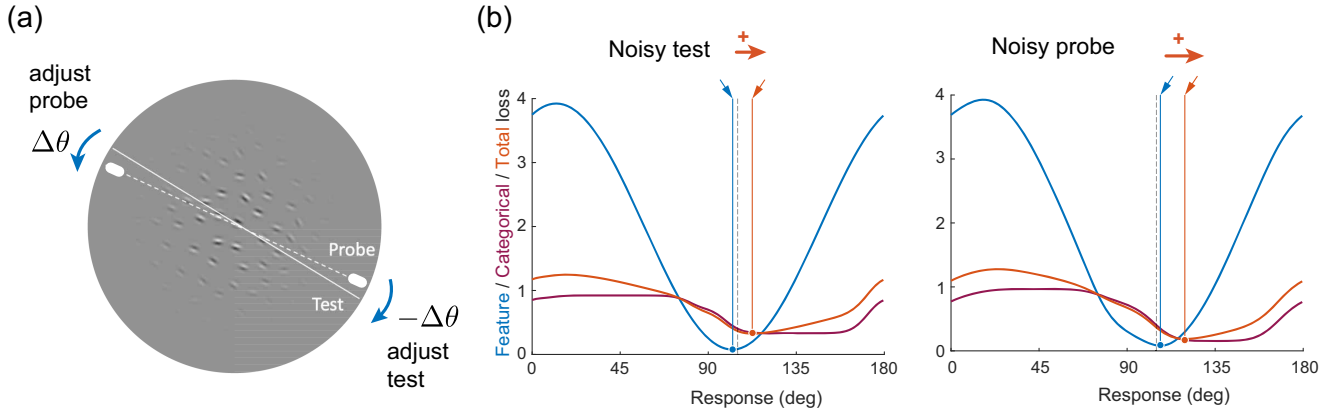
where κ_c specifies the overlap between neighboring categories. Finally, the posterior probabilities of each category are computed according to Equation 13.

One of the advantages of validating our model against data of orientation perception is the availability of reliable measurements of the natural distributions of local visual orientations. For color spectra, such measures are technically more difficult to obtain and were not available for the used CIELAB color space. As a result, we resorted to an approximation of the hue prior $p(\theta)$ using the Cramer–Rao bound

$$\sqrt{J(\theta)} \propto \frac{|1 + b'(\theta)|}{\sigma(\theta)}, \tag{29}$$

Figure 8

Orientation Matching Experiment With Noisy Probe Stimulus



Note. (a) When the test and the probe stimulus in the matching experiment are interchanged and thus stimulus uncertainties are reversed, any model that matches independent perceptual estimates of the test and the probe orientation predicts a reversal of the bias pattern. For example, let us assume that the perceived average orientation of an array of Gabor stimuli (dashed line) is different from its true orientation (solid line). Then a subject would adjust the noiseless probe orientation (bold white line segments) such that its orientation matches the perceived array orientation, leading to a bias $\Delta\theta$ (probe minus test orientation). If probe and test are switched, the matching stimulus configuration is identical, but now the bias is reversed. Figure replotted from Tomassini et al. (2010). (b) Holistic matching. In contrast to the efficient Bayesian estimator (blue arrow), the optimal response according to the holistic matching model (orange arrow) remains on the same side because of the influence of the categorical loss. The predicted biases are larger in the switched condition (noisy probe) because the feature bias and categorical bias add up. The illustration is shown for a single pair of sensory measurements (dashed line). Note, we illustrate the situation for the large stimulus noise condition in Tomassini et al. (2010) for which the efficient Bayesian estimator actually predicts an attractive bias (Wei & Stocker, 2015). See the online article for the color version of this figure.

and the efficient coding constraint Equation 2 (Noel et al., 2021; Wei & Stocker, 2017). Together, they define a relation between the bias $b(\theta)$ and standard deviation $\sigma(\theta)$ of an efficient estimator and the prior. Using the measured biases and standard deviations from the color-matching experiment by Bae et al. (2015), we thus can extract an approximation of the prior. We considered the data measured for the delayed condition because bias and standard deviation are larger, and thus effects of any late noise (e.g., motor noise) are relatively smaller. Reconstruction is based on a polynomial fit (degree 20) to the measured bias and standard deviation, respectively (see Appendix Figure C9). The reconstructed hue prior $p(\theta)$ can be seen in Appendix Figure C10b. Note that similar to orientation matching, we expect subjects' color-matching behavior to substantially differ from the predicted behavior of an efficient estimator. Thus, we predict the extracted prior to be a relatively coarse approximation of the true hue prior and likely to already reflect some of the categorical information.

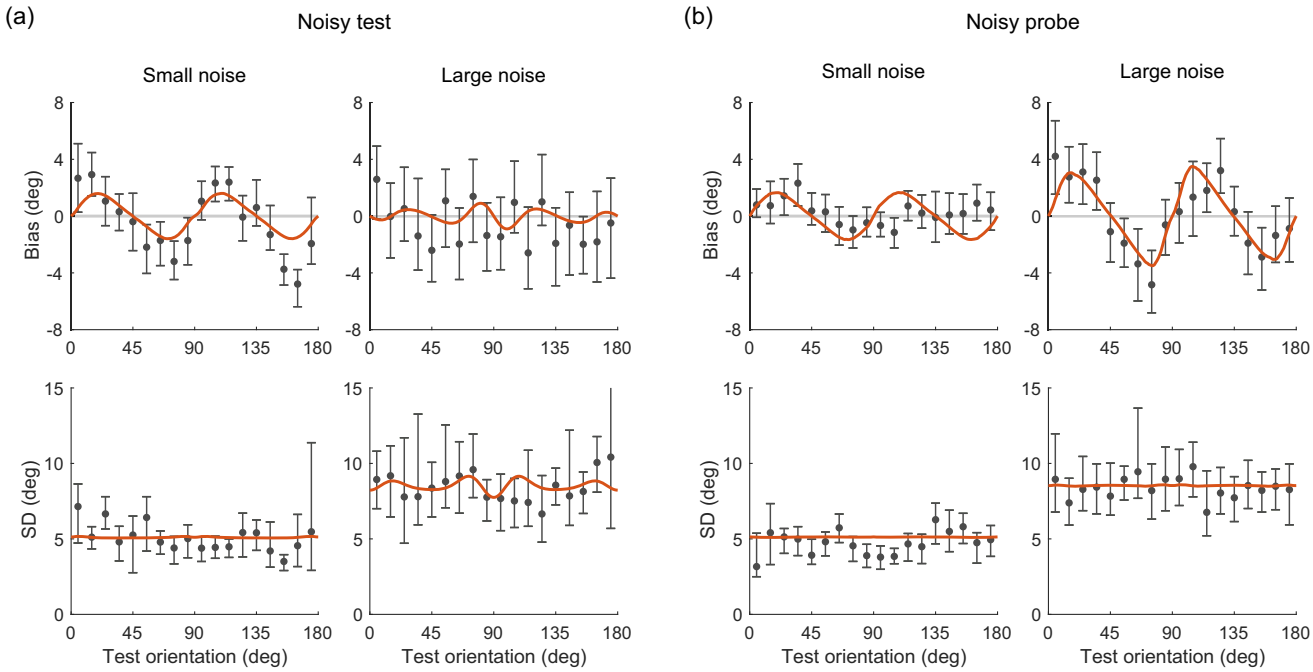
Having extracted approximations of the categorical structure and hue prior, the predicted response distributions of the holistic matching model are given by Equation 23 like in the orientation case, with θ and θ_p representing hue angle instead of orientation. Note that μ_H is replaced by $\Delta\mu$ where $p(\Delta\mu)$ is given by Equation 26. 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 Figure 10. Similar to the orientation-matching data, the hierarchical matching model well captures the entire shape of the error distributions for both conditions, especially the shifts of the distributions at category boundaries. In contrast, although the efficient Bayesian estimator qualitatively accounts for the repulsive bias pattern in the data and its dependency on sensory noise, the predicted bias is generally too small, similar to the orientation-matching data (Figure 4). A direct comparison of the biases and variances predicted by the two models makes this more explicit (Figures 10b and 10c). Note that because the approximated hue prior may already contain categorical information, the efficient Bayesian estimator likely overperforms in this comparison.

It is worth pointing out that the current implementation of the holistic matching model does not perfectly fit the measured behavior. In particular, it does not well capture the fact that the subjects' standard deviations are typically lower for hue angles that lie in the middle as opposed to close to the boundaries of a color category (e.g., the "blue" category with focal angle of 240°). We strongly suspect that this is caused by the current approximation of the hue prior, which shows unexpected troughs in the middle of each color category that seem difficult to justify (Appendix Figure C10). Future research is needed to obtain direct measures of the hue distributions in natural scenes that will lead to a statistically better constrained model.

The limited trial data and the above-discussed limitations in obtaining a good approximation of the prior distribution prevent a meaningful, cross-validated model comparison with the currently best fitting model of the color-matching data, the "CATMET" model proposed by Bae et al. (2015). Nonetheless, a model fit comparison using the correlation analysis as proposed by Bae et al. (2015) indicates that performance is at least comparable (Appendix Figure C12). More importantly, however, is that the "CATMET" model is an estimation

Figure 9

Data and Model Fits for Interchanging Test and Probe Stimuli for Two Different Stimulus Noise Levels (Tomassini et al., 2010)



Note. (a) Bias and standard deviation of subjects' matching response data when the test stimulus is noisy and the probe is noiseless (stimulus setup as in Figure 8a). (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 are reanalyzed from Tomassini et al. (2010). Error bars represent 95% confidence intervals from 100 bootstrap samples of the data. See the online article for the color version of this figure.

model. It assumes that subjects' reported matching color reflects an optimal estimate of the test color conditioned on the most likely color category (Stocker & Simoncelli, 2007). While such a conditioned inference model has previously been shown to account for choice-induced categorical effects in orientation perception (Luu & Stocker, 2018, 2021), it remains an estimator, and as such fails to provide an explanation for subjects' matching behavior when test and probe uncertainty are switched (Figure 8a).

We conclude that the proposed holistic matching model is currently the only model that can account for human matching behavior across all the different data sets we presented here. It represents a general, parsimonious description of human matching behavior that is subject to categorical influences, often referred to as "categorical perception."

Discussion

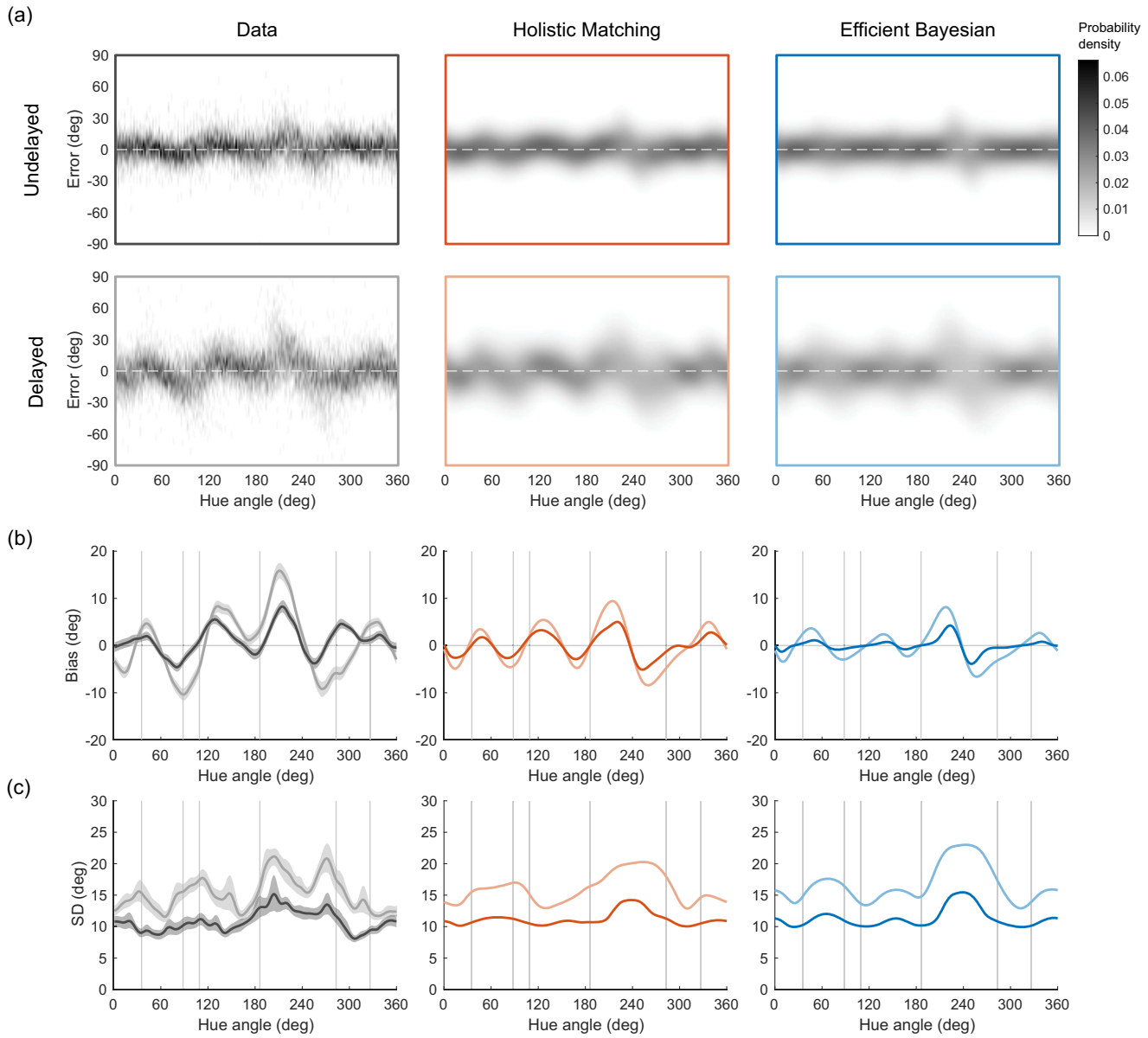
We demonstrated that human sensory perception can be interpreted as a holistic inference process where the percept of a visual stimuli is a joint representation across different levels of a sensory representational hierarchy. Based on this idea, we introduced a holistic matching model to account for human behavior in a widely used perceptual matching task (method-of-adjustment). The model assumes that a subject's report represents an optimal match between the probe and the test stimulus in terms of both feature value and category identity. We quantitatively validated the model against existing data from four different psychophysical studies that probed human orientation perception under different conditions. We showed that, in addition to an efficient sensory encoding of the stimulus orientation, holistic stimulus representations are a necessary assumption in order to provide an accurate account of subjects' full response distributions in all data sets. Specifically, the fact that subjects' response bias is not inverted when switching the role of the test and the probe stimulus in the matching experiment is strong evidence in favor of the proposed holistic matching process; any model only operating at the feature level will 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 is to show that the brain intrinsically operates on holistic stimulus representations when performing perceptual tasks, even when being presented with the simplistic, nonnaturalistic stimuli typically used in psychophysical experiments. These results have profound implications for the correct interpretation of subjects' reports in popular method-of-adjustment experiments. Also, they suggest that "categorical perception" is not per se a perceptual effect but is rather induced by downstream decision processes operating on holistic perceptual representations.

Its holistic nature fundamentally separates our model from other hierarchical Bayesian models that have been proposed to describe categorical effects in perception (Bill et al., 2020; Feldman et al., 2009; Gifford et al., 2014; Kronrod et al., 2016; Landy et al., 2017).

Figure 10

Data and Model Fits for the Color-Matching Experiment



Note. (a) Error Distributions, (b) Bias, and (c) Standard Deviation. Columns show the data and the corresponding best fit model predictions. Bias and variability are larger in the delayed condition than in 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. Vertical lines show categorical boundaries. Data are reanalyzed from Bae et al. (2015). Shaded areas represent 95% confidence intervals from 1,000 bootstrap runs. See the online article for the color version of this figure.

While these models share a similar hierarchical generative process (Figure 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 and thereby reduces inference to a nonhierarchical process with a heterogeneous prior determined by the weighted sum of the stimulus prior given each category. Thus, the predictions of these models are qualitatively identical to those of the nonholistic models considered in our study. 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 (Bae et al., 2015; Ding et al., 2017; Luu & Stocker, 2018; Qiu et al., 2020; Stocker & Simoncelli, 2007) or the updated category belief (Lange et al., 2021), respectively. Although these "self-consistent" hierarchical 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 independent estimate at the feature level. As such, these models too cannot explain human behavior when interchanging the probe and the test stimuli in a matching task (Tomassini et al., 2010; Figure 9). Sims et al. (2016) proposed a rate-distortion theory-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, rate-distortion theory is intrinsically a single-channel estimation model that is difficult to adapt to a matching process between a test and probe stimulus under more general conditions (i.e., with noisy probe stimuli). As a result, this model is also too limited to account for the data by Tomassini et al. (2010).

It is worth highlighting some other 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 contrasts with studies that limit model validation to comparisons of summary statistics such as the average response (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 (Appendix Figure C2). Furthermore, the model assumes that perceptual inference operates on efficient sensory representations and thus incorporates and extends previous work showing that human perception ubiquitously exhibits lawful hallmarks of efficient coding in combination with optimal Bayesian inference (Wei & Stocker, 2017). Thus, aside from the specification of the noise levels, the free model choices are essentially limited to the specification of the categorical structure of orientation. Little is known about the natural structure of orientation categories. Thus, our choice of "cardinal" and "oblique" categories is somewhat arbitrary, albeit intuitive, and shared with previous studies (e.g., Rosielle & Cooper, 2001; Wakita, 2004). It is reassuring, however, that assuming a categorical structure that only distinguishes between clockwise and counterclockwise orientations across the vertical meridian does not significantly change the model behavior (see Appendix Figures C3 and C4, for the two-category model fit to both data sets). Future experiments are necessary to better constrain the categorical structure of visual orientation in human observers or to impose experimentally well-defined categories.

Note that although our model is formally normative, a clear rationale for why humans would optimize the particular objective function in Equation 1 is difficult to provide. One possibility is that operating on holistic representations is the visual system's default mode because ecologically relevant tasks under natural conditions most often rely on holistic comparisons (e.g., "Is this the same person even though the haircut is different?"). Furthermore, because subjects did not receive feedback on their responses in any of the studies included in our analysis, they simply may not have had access to the necessary error signals needed to properly adjust their objectives during the experiments. Future studies are necessary to elucidate in more detail the role of feedback in subjects' matching behavior. Another possibility is that there exist asymmetries between the inference processes at the different levels of the representational hierarchy that are currently not considered in our model. For example, the computational costs for inference could be different at the different levels of the hierarchy, in which case a combined objective may reflect an optimal cost-accuracy trade-off. Comparing the objective function across the three data sets with a noiseless probe reveals that the objective is more categorical for the first two studies (De Gardelle et al., 2010; Noel et al., 2021; w = [0.83, 0.87, 0.91]) compared to the third (Bays, 2014; w = 0.49). Interestingly, the main experimental difference between the three studies is that only in the first two studies, subjects were presented with a mask stimulus right after the test stimulus presentation. We can speculate that the mask may have led to interferences at the slower (i.e., more costly) inference process at the feature level resulting in a less informative posterior $p(\theta|m)$, while the faster inference process at the category level may have already completed its computation (Hochstein & Ahissar, 2002). If so, then a combined objective that weighs categorical information depending on its relative informativeness would be the optimal strategy (Qiu et al., 2020) and would explain the observed differences in w. Future experiments combined with theoretical considerations will help to probe these hypotheses further.

Finally, as low-level perception has been shown to follow common principles of sensory inference (Wei & Stocker, 2017), there is good reason to believe that our model generalizes to stimulus domains other than visual orientation or color. In particular, various forms of direction perception, such as motion direction (Rauber & Treue, 1998), pointing direction (Smyrnis et al., 2014), and visual and vestibular heading direction (Cuturi & MacNeilage, 2013), exhibit repulsive biases away from as well as better discrimination at cardinal directions (Gros et al., 1998). Similarly, studies of visuospatial memory distortions have found biases toward landmarks, an effect that 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 data sets requires models that not only consider efficient sensory representations but also a holistic 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 have provided meaningful (normative) explanations for many of the characteristic bias and variability patterns observed in perceptual estimation tasks. 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 psychophysical matching task.

References

Appelle, S. (1972). Perception and discrimination as function of stimulus orientation. *Psychological Bulletin*, *78*, 266–278. https://doi.org/10.1037/h0033117

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193. https://doi.org/10.1037/h0054663

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). 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), 744–763. https://doi.org/10.1037/xge0000076

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, *1*(1), 217–233.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632–3645. https://doi.org/10.1523/JNEUROSCI.3204-13.2014

Bill, J., Pailian, H., Gershman, S. J., & Drugowitsch, J. (2020). Hierarchical structure is employed by humans during visual motion perception. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(39), 24581–24589. https://doi.org/10.1073/pnas.2008961117

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf hypothesis and probabilistic inference: Evidence from the domain of color. *PLOS ONE*, *11*(7), Article e0158725. https://doi.org/10.1371/journal.pone.0158725

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 4002–4006. https://doi.org/10.1073/pnas.95.7.4002

Cuturi, L. F., & MacNeilage, P. R. (2013). Systematic biases in human heading estimation. *PLOS ONE*, *8*(2), Article e56862. https://doi.org/10.1371/journal.pone.0056862

De Gardelle, V., Kouider, S., & Sackur, J. (2010). An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. *Journal of Vision*, *10*(10), Article 6. https://doi.org/10.1167/10.10.6

Ding, S., Cueva, C., Tsodyks, M., & Qian, N. (2017). Visual perception as retrospective Bayesian decoding from high- to low-level features. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), E9115–E9124. https://doi.org/10.1073/pnas.1706906114

Durgin, F. H., & Li, Z. (2011). The perception of 2D orientation is categorically biased. *Journal of Vision*, *11*(8), Article 13. https://doi.org/10.1167/11.8.13

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. https://doi.org/10.1037/a0017196

Fritsche, M., Spaak, E., & de Lange, F. P. (2020). A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife*, *9*, Article e55389. https://doi.org/10.7554/eLife.55389

Gifford, A., Cohen, Y., & Stocker, A. A. (2014). Characterizing the impact of category uncertainty on human auditory categorization behavior. *PLOS Computational Biology*, *10*(7), Article 1003715. https://doi.org/10.1371/journal.pcbi.1003715

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. https://doi.org/10.1038/nn.2831

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science*, *1*(1), 69–78. https://doi.org/10.1002/wcs.26

Gros, B. L., Blake, R., & Hiris, E. (1998). Anisotropies in visual motion perception: A fresh look. *Journal of the Optical Society of America, A: Optics, Image Science & Vision*, *15*(8), 2003–2011. https://doi.org/10.1364/JOSAA.15.002003

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 30–54. https://doi.org/10.1037/xhp0000290

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376. https://doi.org/10.1037/0033-295x.98.3.352

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590

Kim, S., & Burge, J. (2018). The lawful imprecision of human surface tilt estimation in natural scenes. *elife*, *7*, Article e31448. https://doi.org/10.7554/eLife.31448

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review*, *23*(6), 1681–1712. https://doi.org/10.3758/s13423-016-1049-y

Landy, D., Crawford, L. E., & Corbin, J. (2017). *A hierarchical Bayesian model of individual differences in memory for emotional expressions* [Conference session]. Proceedings of the Annual Meeting of the Cognitive Science Society.

Lange, R. D., Chattoraj, A., Beck, J. M., Yates, J. L., & Haefner, R. M. (2021). A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *PLOS Computational Biology*, *17*(11), Article 1009517. https://doi.org/10.1371/journal.pcbi.1009517

Langlois, T. A., Jacoby, N., Suchow, J. W., & Griffiths, T. L. (2021). Serial reproduction reveals the geometry of visuospatial representations. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(13). https://doi.org/10.1073/pnas.2012938118

Luu, L., & Stocker, A. A. (2018). Post-decision biases reveal a self-consistency principle in perceptual inference. *elife*, *7*, Article e33334. https://doi.org/10.7554/eLife.33334

Luu, L., & Stocker, A. A. (2021). Categorical judgments do not modify sensory representations in working memory. *PLOS Computational Biology*, *17*(6), Article e1008968. https://doi.org/10.1371/journal.pcbi.1008968

Ni, L., & Stocker, A. A. (2023). Efficient sensory encoding predicts robust averaging. *Cognition*, *232*, Article 105334. https://doi.org/10.1016/j.cognition.2022.105334

Noel, J.-P., Lakshminarasimhan, K. J., Park, H., & Angelaki, D. E. (2020). Increased variability but intact integration during visual navigation in autism spectrum disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(20), 11158–11166. https://doi.org/10.1073/pnas.2000216117

Noel, J.-P., Zhang, L.-Q., Stocker, A. A., & Angelaki, D. E. (2021). Individuals with autism spectrum disorder have altered visual encoding capacity. *PLOS Biology*, *19*(5), Article e3001215. https://doi.org/10.1371/journal.pbio.3001215

Polania, R., Woodford, M., & Ruff, C. C. (2019). Efficient coding of subjective value. *Nature Neuroscience*, *22*(1), 134–142. https://doi.org/10.1038/s41593-018-0292-0

Prat-Carrabin, A., & Woodford, M. (2021). *Efficient coding of numbers explains decision bias and noise*. bioRxiv. https://doi.org/10.1101/2020.02.18.942938

Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017). 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), 6–17. https://doi.org/10.1037/xhp0000302

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

Rauber, H.-J., & Treue, S. (1998). Reference repulsion when judging the direction of visual motion. *Perception*, *27*(4), 393–402. https://doi.org/10.1016/S0042-6989(99)00025-5

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

Sims, C. R., Ma, Z., Allred, S. R., Lerch, R. A., & Flombaum, J. I. (2016). *Exploring the cost function in color perception and memory: An information-theoretic model of categorical effects in color matching* [Conference session]. Proceedings of the Annual Meeting of the Cognitive Science Society.

Smyrnis, N., Mantas, A., & Evdokimidis, I. (2014). Two independent sources of anisotropy in the visual representation of direction in 2-D space. *Experimental Brain Research*, *232*(7), 2317–2324. https://doi.org/10.1007/s00221-014-3928-7

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585. https://doi.org/10.1038/nn1669

Stocker, A. A., & Simoncelli, E. P. (2007). A Bayesian model of conditioned perception. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems NIPS 20* (pp. 1409–1416). MIT Press.

Straub, D., & Rothkopf, C. A. (2021). Looking for image statistics: Active vision with avatars in a naturalistic virtual environment. *Frontiers in Psychology*, *12*, Article 641471. https://doi.org/10.3389/fpsyg.2021.641471

Taylor, R., & Bays, P. M. (2018). Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *Journal of Neuroscience*, *38*(32), 7132–7142. https://doi.org/10.1523/JNEUROSCI.1018-18.2018

Tomassini, A., Morgan, M. J., & Solomon, J. A. (2010). Orientation uncertainty reduces perceived obliquity. *Vision Research*, *50*(5), 541–547. https://doi.org/10.1016/j.visres.2009.12.005

Van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

Wakita, M. (2004). Categorical perception of orientation in monkeys. *Behavioural Processes*, *67*(2), 263–272. https://doi.org/10.1016/j.beproc.2004.04.005

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

Wei, X.-X., & Stocker, A. (2012). Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference. In P. Bartlett, F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems NIPS 25* (pp. 1313–1321). MIT Press.

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, *18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Wei, X.-X., & Stocker, A. A. (2016). Mutual information, Fisher information, and efficient coding. *Neural Computation*, *28*(2), 305–326. https://doi.org/10.1162/NECO_a_00804

Wei, X.-X., & Stocker, A. A. (2017). Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(38), 10244–10249. https://doi.org/10.1073/pnas.1619153114

Witzel, C., & Gegenfurtner, K. (2013). Categorical sensitivity to color differences. *Journal of Vision*, *13*(7), Article 1. https://doi.org/10.1167/13.7.1

Zhang, L.-Q., & Stocker, A. A. (2022). Prior expectations in visual speed perception predict encoding characteristics of neurons in area MT. *Journal of Neuroscience*, *42*(14), 2951–2962. https://doi.org/10.1523/JNEUROSCI.1920-21.2022

(*Appendices follow*)

Appendix A

Psychophysical Data

Data Set by De Gardelle et al. (2010)

Each trial began with a background noise texture, and then a test stimulus (Gabor patch) was presented for a variable duration at a random location 6.5 deg 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*). Subjects did not receive feedback (only in training). Presentation times of the test Gabor were (1,000, 160, 80, 40, 20) ms and 0 ms (no stimulus presented), randomly intermixed. Forty-six 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 of 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 data set contained 1,103, 2,187, 2,140, and 1,383 trials for each presentation duration, respectively.

Illustrations of the data distributions (Figure 3) and bias and standard deviation (Figure 4a) 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 deg; see Appendix Figure C1). 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 Figures 3 and 4.

Data Set 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 three 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 (both combined and individual subjects), we only included trials in the no-feedback block. We also excluded trials in which the responses were 3 *SDs* away from the mean response.

Illustrations of the data distributions, bias, and standard deviation (Figure 6a) of the combined subject are based on smoothing the raw trial data with a symmetric Gaussian kernel centered at each data point with a standard deviation of 5 deg. Distributions are normalized to indicate the conditional probability of response for each test orientation.

Data Set by Bays (2014)

Each trial began with a white fixation cross at the center of a gray background. Once stable fixation was maintained within a 2 deg radius of the cross, the stimulus array was presented for 2 s. The stimulus array consisted of 1, 2, 4, or 8 oriented colored bars (2 deg $\times$ 0.3 deg), each presented at one of eight equally spaced locations at 6 deg eccentricity from the fixation cross. Colors were randomly selected on each trial without repetition. After a 1s blank, a randomly chosen bar with a new random orientation appeared as the probe. Participants adjusted an input dial to match the orientation of the probe to the remembered orientation of the test bar that occurred in the same location of the stimulus array. No feedback was provided. Eight subjects participated in the experiment. Each subject completed 900 trials, 225 trials for each set size, randomly interleaved. For our analysis, we combined the data across all subjects. We only included trials where the set size was 1 or 2, for they are most similar to a common perceptual estimation task.

Illustrations of the data distributions, bias, and standard deviation (Figure 7) are based on smoothing the raw trial data with a symmetric Gaussian kernel centered at each data point with a standard deviation of 5 deg. Distributions are normalized to indicate the conditional probability of response for each test orientation.

Data Set 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. The orientation of the test stimulus was randomly selected from 18 orientations each 10 deg 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. In all conditions, no feedback was provided. Five subjects participated in the main experiment, each completing eight 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 but only included the trials with response-terminated presentations. We also excluded trials in which the responses were 3 *SDs* away from the mean response.

(*Appendices continue*)

Data Set by Bae et al. (2015)

In the color-naming experiment, subjects viewed a colored square and selected the color name (out of eight basic color terms) that most closely described the test color. In the matching experiments, subjects viewed a colored square and chose the color that best matched the test color by clicking on a color wheel. No feedback was provided. In the undelayed condition, the colored square remained on the screen until the subject responded. In the delayed condition, there was a delay period after the colored square disappeared before the color wheel was presented. Ten subjects participated in the color-naming experiment, each completed six trials for each test color. Eight subjects participated in the undelayed condition, each completing 16 trials per test color for half of the test colors, resulting in 64 trials per test color from all subjects. Three subjects participated in the delayed condition, 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 experiment. We excluded trials in which the responses were 5 *SDs* away from the mean response in the estimation experiments.

In Figure 10, illustrations of the estimation error distributions are based on smoothing the raw trial data with a 1D Gaussian kernel along the error axis centered at each data point with a standard deviation of 3 deg. Similarly, illustrations of the bias and standard deviation are based on smoothing the raw trial data with a running Gaussian window with a standard deviation of 5 deg.

Appendix B

Model Fit

We jointly fit the model to the data of all the conditions in each data set by maximizing the likelihood of the model given the data, that is,

$$\arg\max_{\rho} p(D|\rho) = \prod_{j=1}^{n} p(D_j|\rho) = \prod_{j=1}^{n} p(\hat{\theta}_j|\rho, \theta_j), \qquad \text{(B1)}$$

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 to the orientation data sets, representing the average natural orientation statistics extracted from indoor and outdoor scene images as reported by Coppola et al. (1998). More specifically, we use a spline approximation of the orientation histograms for indoor and outdoor scenes, also assuming that the distributions are symmetric around the vertical orientation (Appendix Figure C2a), and then take the average of the two spline fits as the orientation prior $p(\theta)$ (Figure 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 eight 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; and
- 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 five 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; and
- κ_m for motor noise.

For fitting the data by Bays (2014), we assume no stimulus noise and two sensory noise levels corresponding to the two different set sizes, resulting in a total of six free parameters:

- a group of two parameters κ_i for two different set sizes;
- κ_c for category uncertainty;
- α for the probability of cardinal category;
- w for the weight of the categorical loss; and
- 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 seven 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; and
- 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 (Equation 27), with seven free parameters for the mean boundary positions b_j ($j = 1, 2, \ldots, 6$) and the uncertainty in boundary positions κ_b. For fitting the matching data (Figure 10), we assume no stimulus noise and two sensory noise levels corresponding to the undelayed and delayed conditions, resulting in a total of five free parameters:

(*Appendices continue*)

- 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; and
- 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).

Fit Parameter Values

Table B1
Best Fitting Model Parameters for Data From All Matching Experiments With Noiseless Probe (Combined Subjects)

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	0.91
κ_m: motor noise	34.63
Noel et al. (2021): [control, ASD]	
κ_i: sensory noise	[19.17, 6.19]
κ_c: category uncertainty	[8.25, 8.04]
α: cardinal probability	[0.56, 0.60]
w: categorical weight	[0.87, 0.83]
κ_m: motor noise	[27.56, 120.18]
Bays (2014)	
κ_i: sensory noise	[12.33, 4.65]
κ_c: category uncertainty	7.59
α: cardinal probability	0.54
w: categorical weight	0.49
κ_m: motor noise	217.71

Note. ASD = autism spectrum disorder.

Table B2
Best Fitting Model Parameters for Data From Matching Experiments With Noisy Probe (Combined Subject)

Parameter	Value
Tomassini et al. (2010)	
κ_i: sensory noise	696.62
κ_e: stimulus noise	[694.30, 17.76]
κ_c: category uncertainty	6.63
α: cardinal probability	0.54
w: categorical weight	0.36
κ_m: motor noise	38.61

Table B3
Best Fitting Parameters of the Two-Category Holistic Matching Model 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	0.87
κ_m: motor noise	24.32
Tomassini et al. (2010)	
κ_i: sensory noise	689.42
κ_e: stimulus noise	[681.23, 17.61]
κ_b: boundary noise	9.41
κ_c: category overlap	2.22
w: categorical weight	0.42
κ_m: motor noise	37.76

Table B4
Best Fitting Parameters of the Holistic Matching Model for the Color-Naming and Color Estimation Data in Bae et al. (2015)

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	[135.89, 23.77]
κ_c: category overlap	7.89
w: categorical weight	[0.23, 0.37]
κ_m: motor noise	36.65

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, and 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.

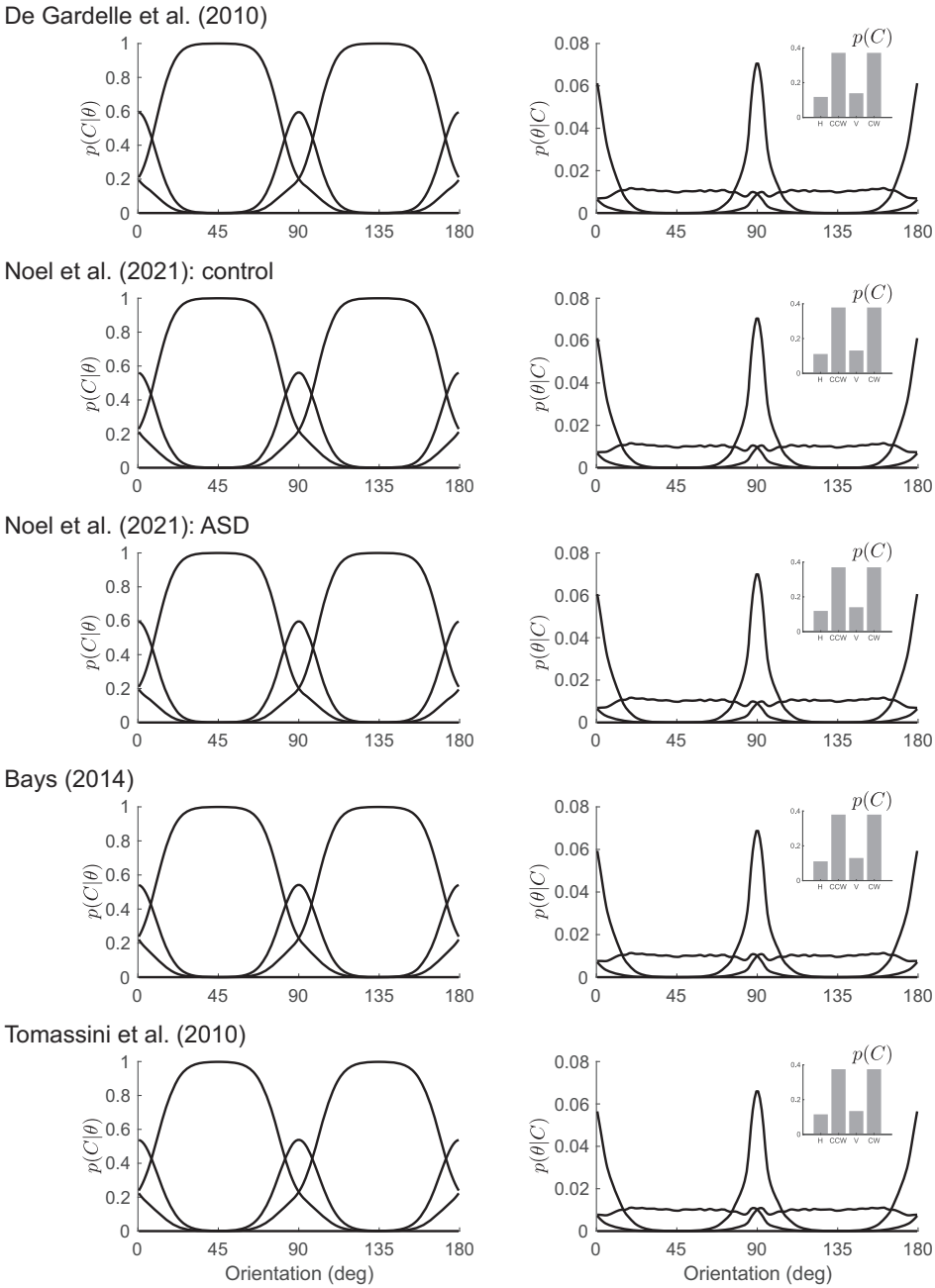
The "Omniscient" Observer Model

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 overfits 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 deg leads to the best performance (Appendix Figure C1).

(*Appendices continue*)

Figure B1

Best Fitting Categories for All Four Orientation Data Sets



Note. ASD = autism spectrum disorder.

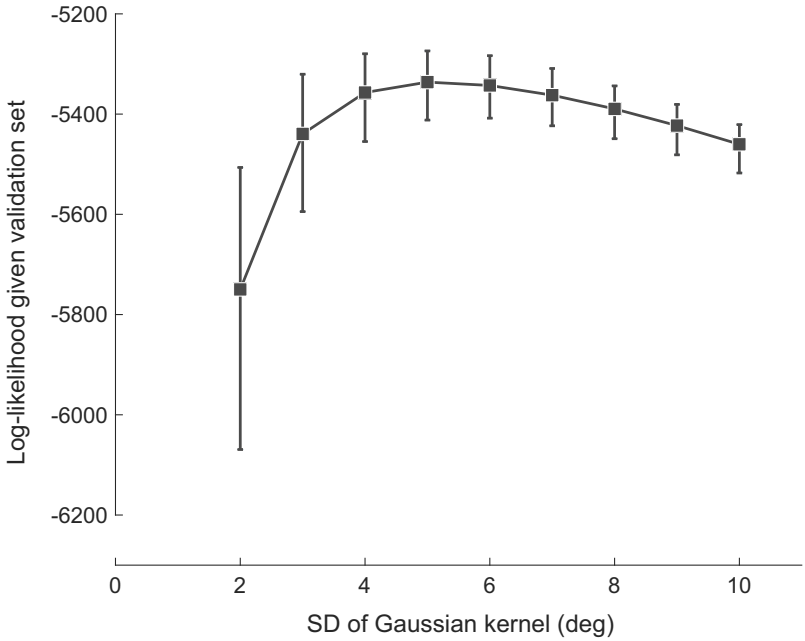
(*Appendices continue*)

Appendix C

Extended Data Illustrations

Figure C1

Cross-Validation of the Kernel Density Estimation Accuracy for the Omniscient Model as a Function of Different Gaussian Kernel Size

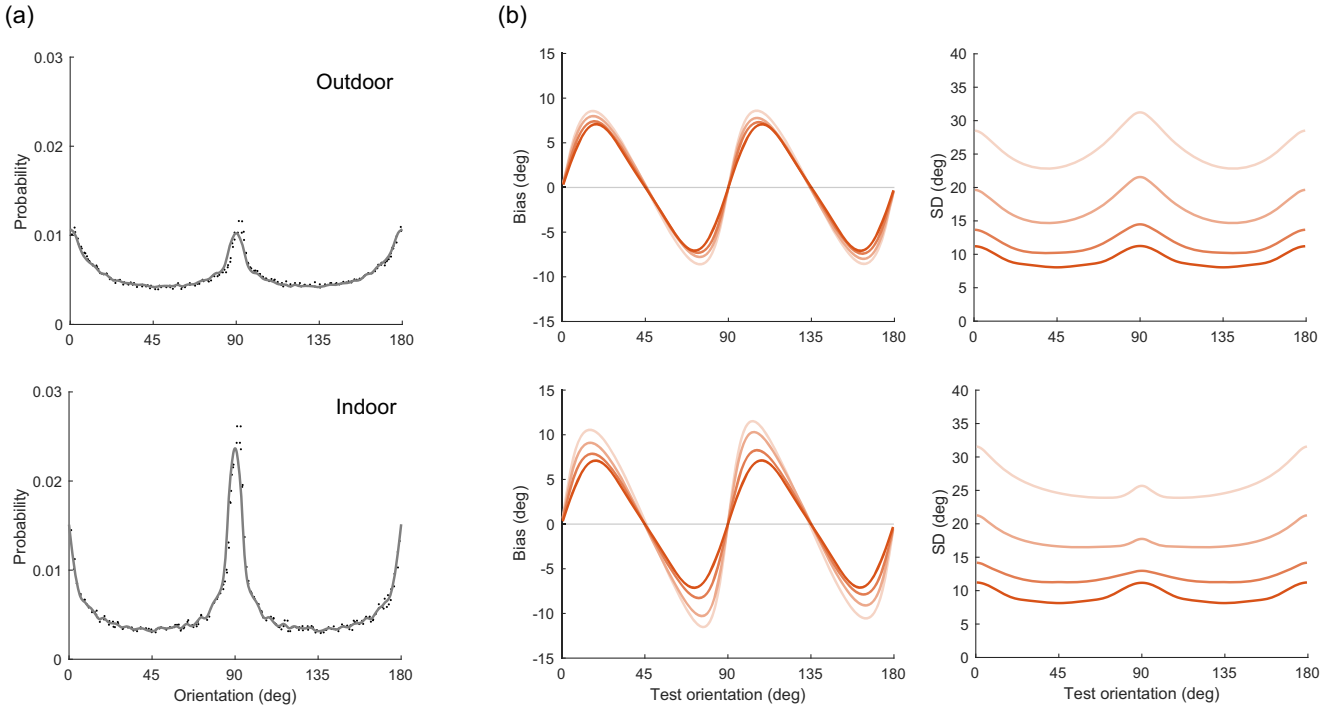


Note. Squares represent the median, and error bars represent 95% confidence intervals of 100 repetitions of a repeated random subsampling cross-validation procedure. Accuracy shows a lawful dependency on kernel size with a standard deviation of 5 deg providing the largest median likelihood value.

(*Appendices continue*)

Figure C2

Predictions of the Holistic Matching Model Using "Outdoor" and "Indoor" Orientation Priors

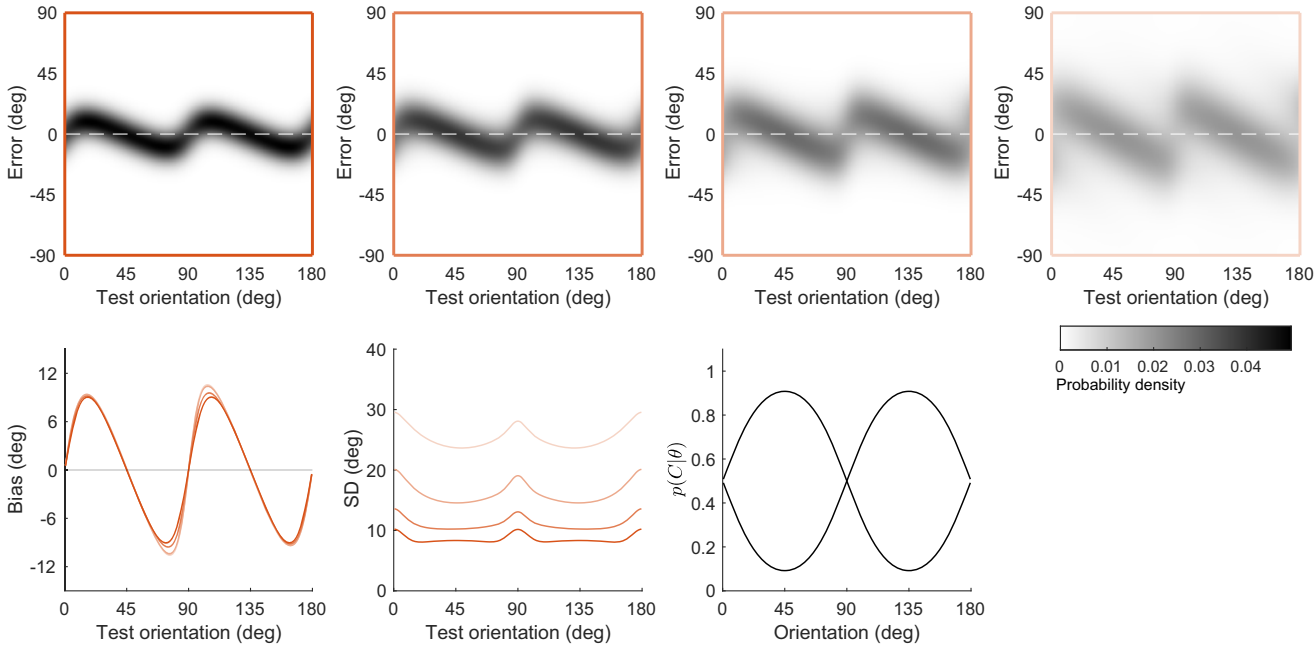


Note. Data: De Gardelle et al. (2010). (a) Image statistics of indoor and outdoor natural scenes (dots) and their smooth spline interpolations representing the corresponding prior distributions (lines). We assume the distributions to be symmetric around vertical. Data reanalyzed from Coppola et al. (1998). (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 Appendix Table B1. 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 (Figure 2). See the online article for the color version of this figure.

(*Appendices continue*)

Figure C3

Two-Category Holistic Matching Model Fit for Matching Task With Noiseless Probe

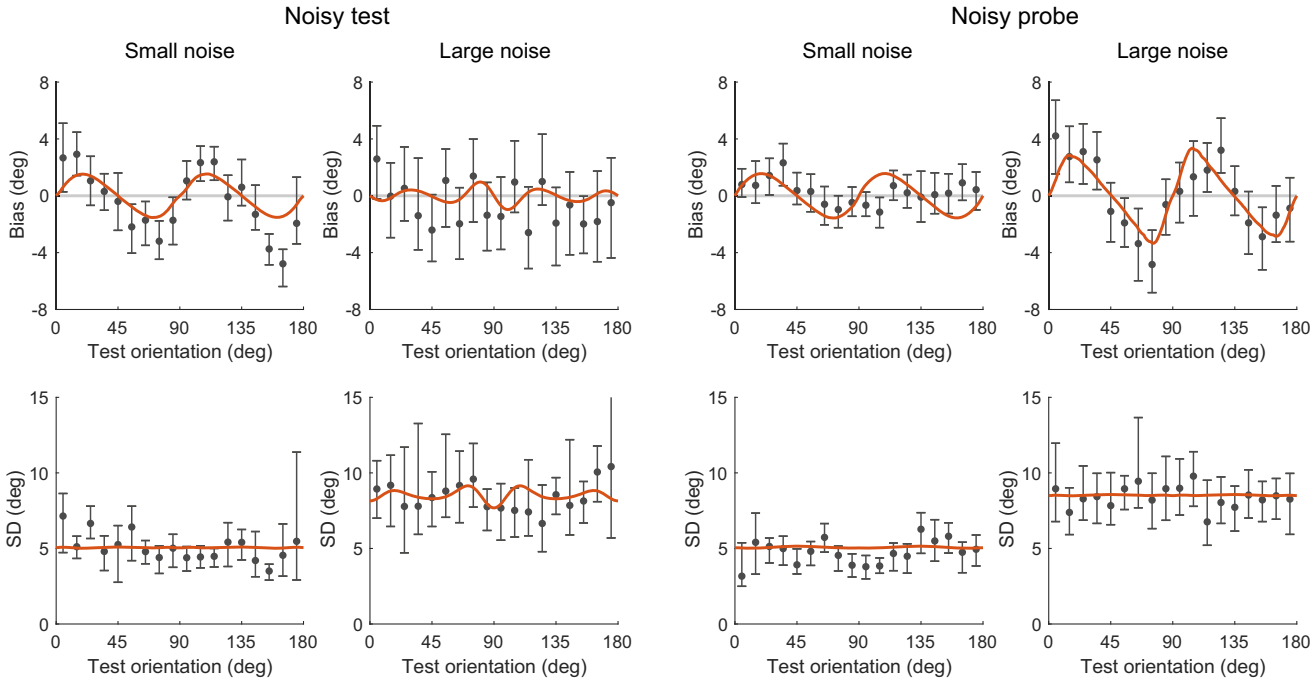


Note. Data: De Gardelle et al. (2010). The cardinal probability α is set to zero, and the parameters for boundary noise and category overlap are allowed to vary independently (Equation 9). The fitting procedure is otherwise identical to the model with four categories. Fit parameter values are listed in Appendix Table B3. See the online article for the color version of this figure.

(*Appendices continue*)

Figure C4

Two-Category Holistic Matching Model Fit for Matching Task With Noisy Probe

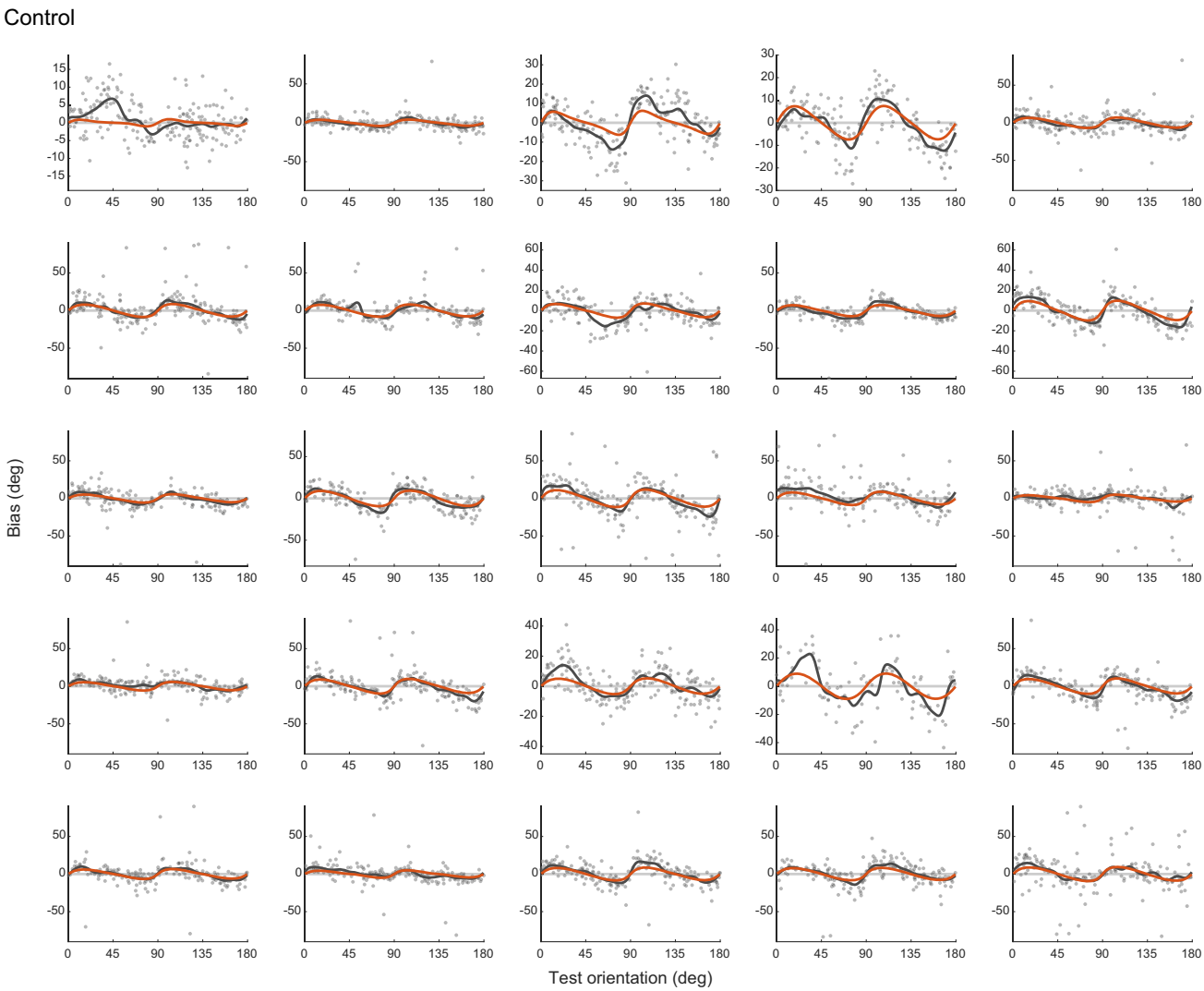


Note. Data: Tomassini et al. (2010). The cardinal probability α is set to zero, and the parameters for boundary noise and category overlap are allowed to vary independently (Equation 9). The fitting procedure is otherwise identical to the model with four categories. Fit parameter values are listed in Appendix Table B3. See the online article for the color version of this figure.

(*Appendices continue*)

Figure C5

Bias Curves and Model Fits for Individual Subjects in Noel et al. (2021): Control Group (N = 25)

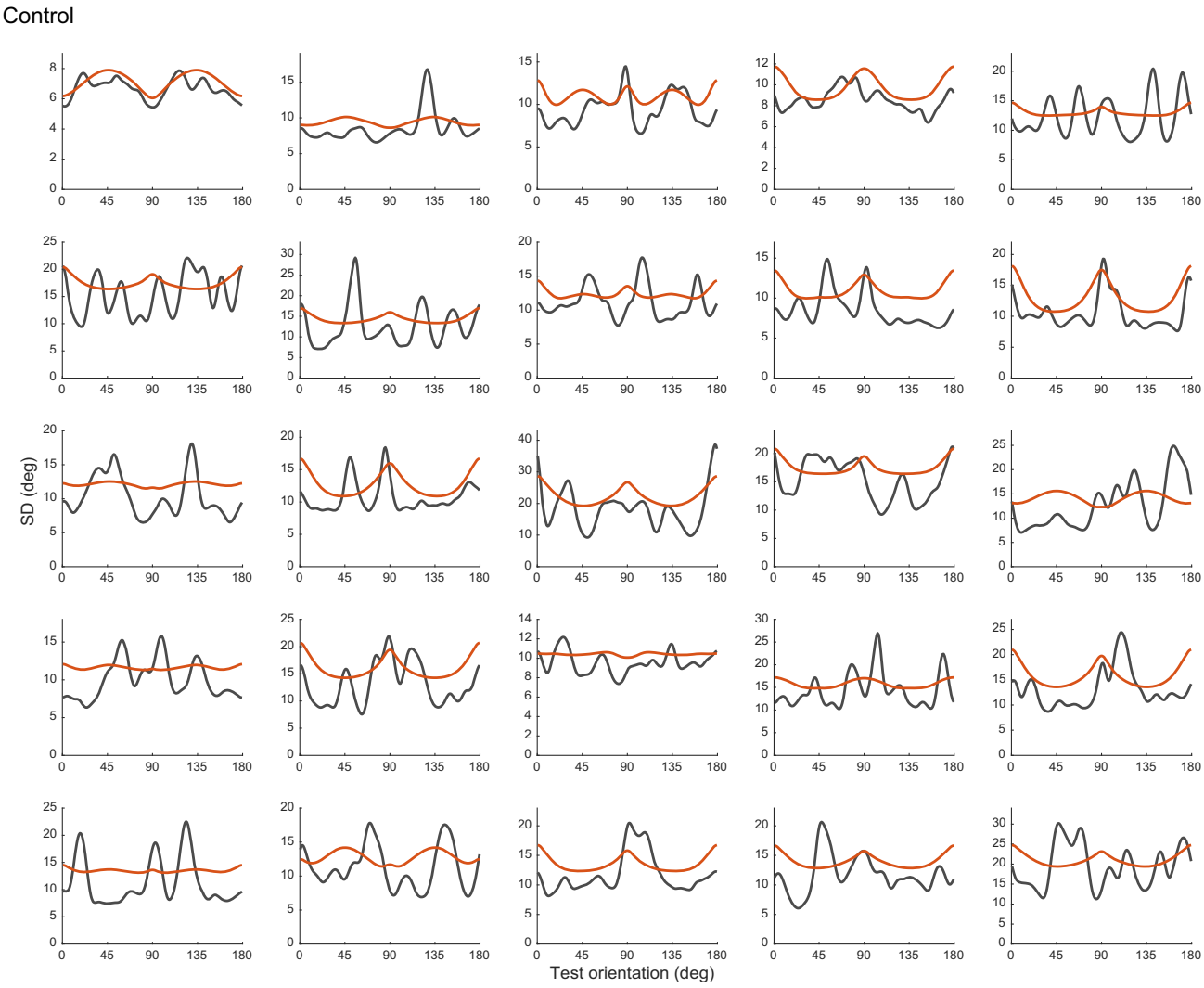


Note. See the online article for the color version of this figure.

(*Appendices continue*)

Figure C6

Standard Deviation Curves and Model Fits for Individual Subjects in Noel et al. (2021): Control Group (N = 25)

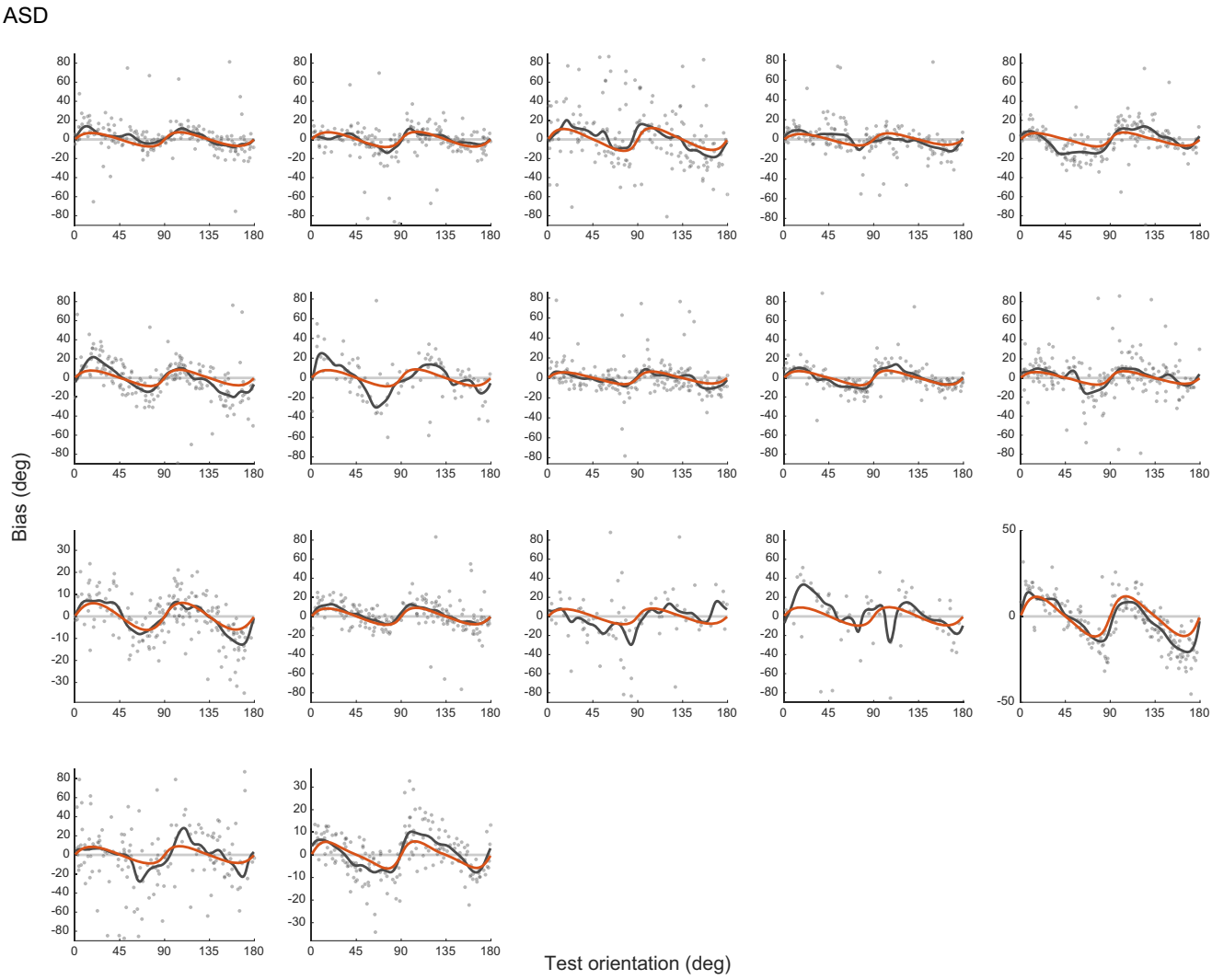


Note. Note that panels at the same grid locations in Figure C5 represent the corresponding bias curves for individual subjects. See the online article for the color version of this figure.

(*Appendices continue*)

Figure C7

Bias Curves and Model Fits for Individual Subjects in Noel et al. (2021): ASD Group (N = 17)

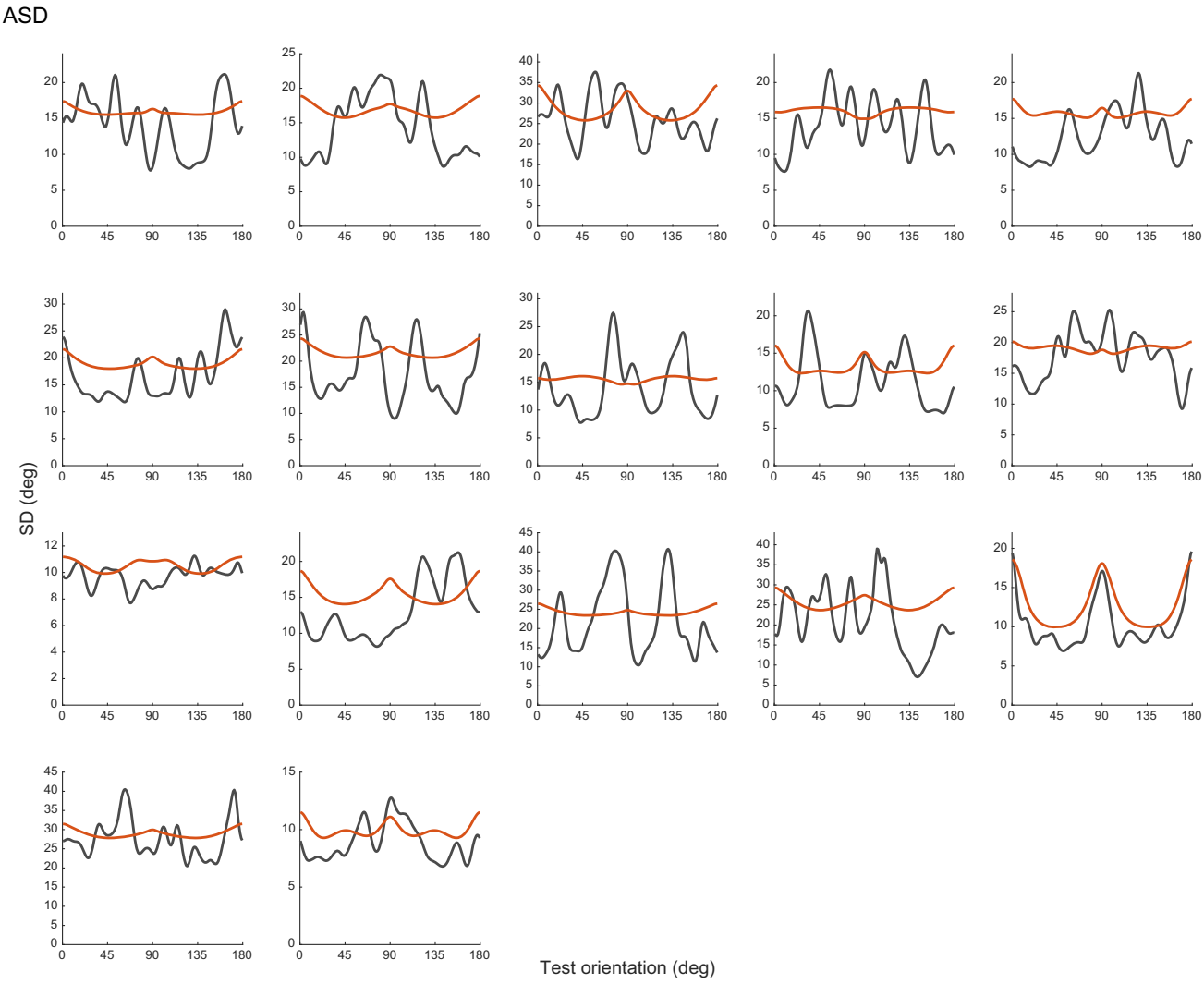


Note. ASD = autism spectrum disorder. See the online article for the color version of this figure.

(*Appendices continue*)

Figure C8

Standard Deviation Curves and Model Fits for Individual Subjects in Noel et al. (2021): ASD Group (N = 17)

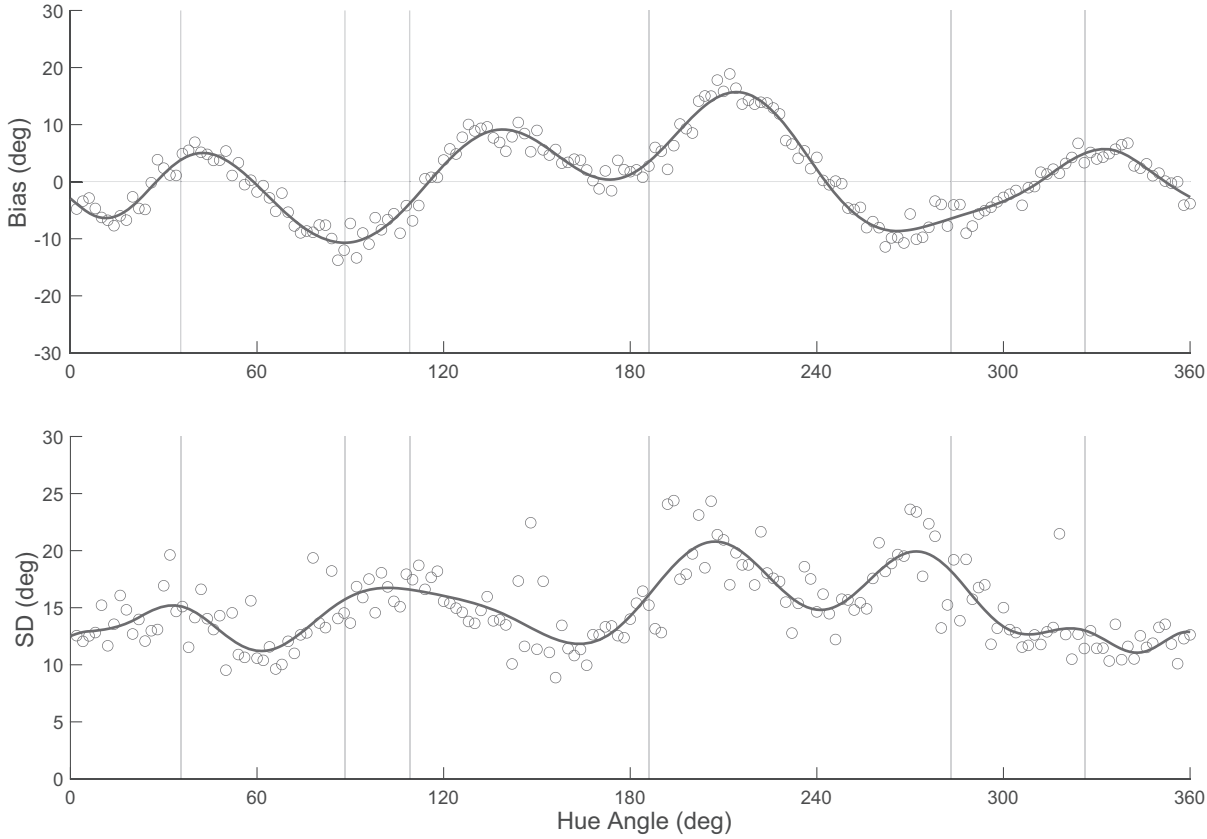


Note. Note that panels at the same grid locations in Figure C7 represent the corresponding bias curves for individual subjects. ASD = autism spectrum disorder. See the online article for the color version of this figure.

(*Appendices continue*)

Figure C9

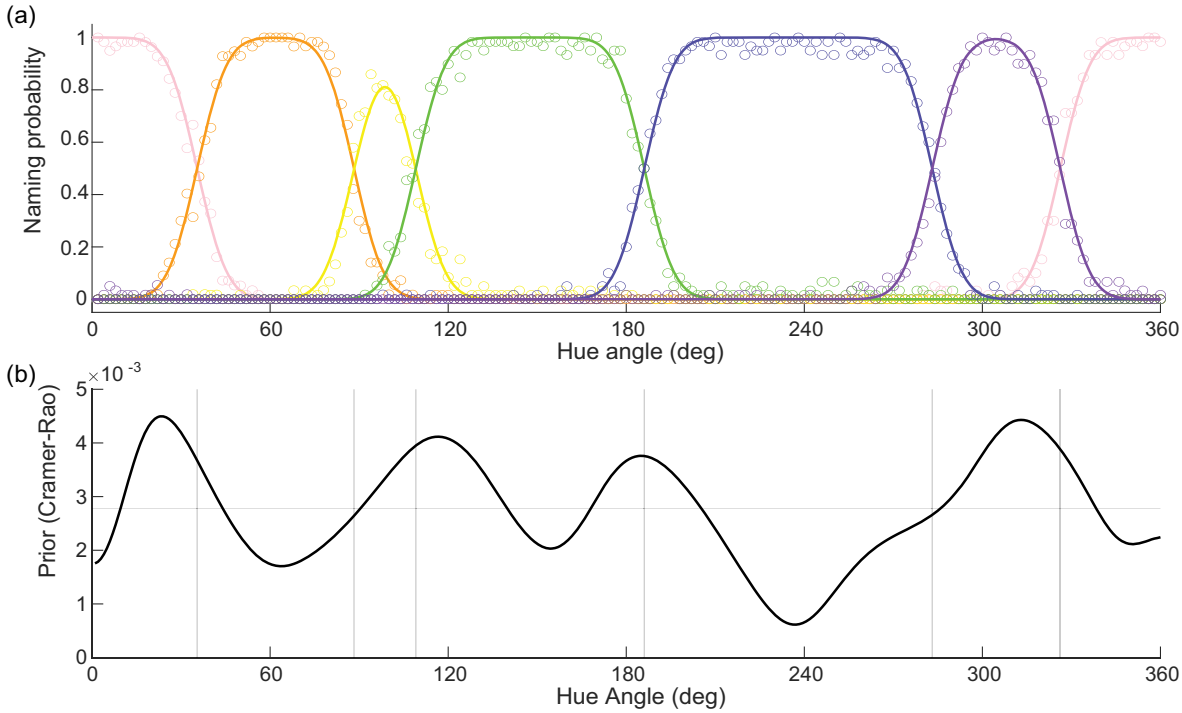
Polynomial Fit of Degree 20 to the Bias and Standard Deviation of Subjects' Color-Matching Responses in the Delayed Condition in Bae et al. (2015)



(*Appendices continue*)

Figure C10

Categorical Structure and Prior of Color

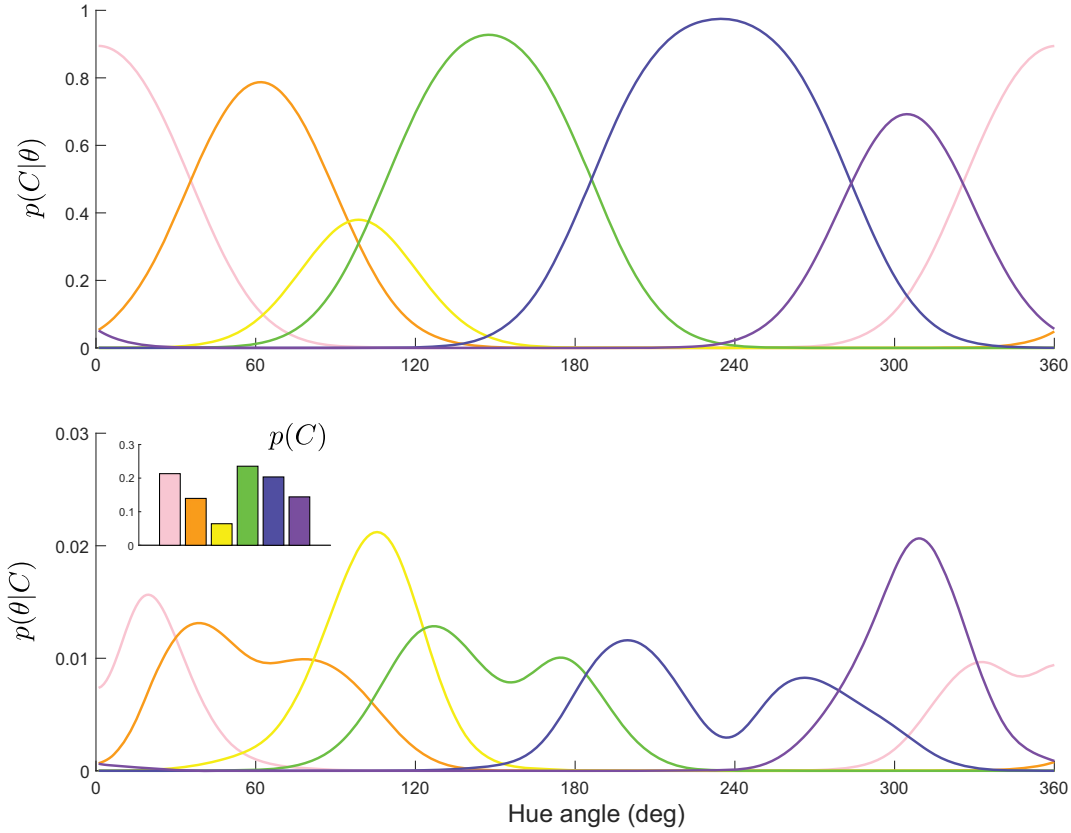


Note. (a) Data from the color-naming experiment in Bae et al. (2015) and smooth approximations using cumulative von Mises distributions (solid lines). These naming probabilities served as proxies for the underlying categorical structure $p(C|\theta)$. (b) Prior extracted from the bias and standard deviations of participants' response in the color-matching experiment, based on the Cramer–Rao bound and the assumption that sensory encoding is efficient (Noel et al., 2021). See the online article for the color version of this figure.

(*Appendices continue*)

Figure C11

Best Fitting Color Categories

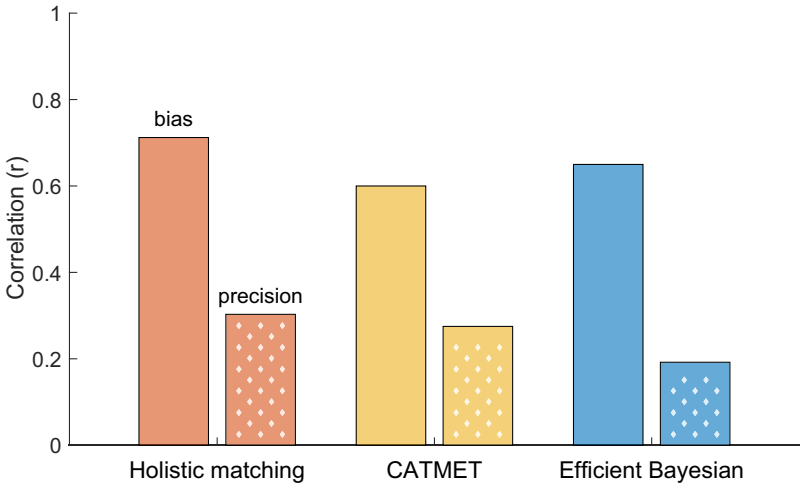


Note. See the online article for the color version of this figure.

(*Appendices continue*)

Figure C12

Model Comparison for Color-Matching Data



Note. Correlation values for bias and precision between fit model predictions and data obtained for 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 article. For our models, we compute correlations in the same way as Bae et al. (2015). We fit a weighted sum of a von Mises distribution and a uniform distribution to the response to each test color, thus $p(\hat{\theta}|\theta) = \beta \mathrm{vm}(\hat{\theta}; \theta + b(\theta), \kappa(\theta)) + (1 - \beta)\frac{1}{2\pi}$, where $1 - \beta$ is the guess rate, $b(\theta)$ is the bias, and $\kappa(\theta)$ is the precision. Since the guess rate is taken into account, all trials are included. We compute the bias and precision based on the response distribution predicted by the models as well, assuming zero guess rate. Then we calculate the correlation for bias and precision between the data and model predictions. 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). See the online article for the color version of this figure.

Received October 22, 2022
Revision received September 24, 2023
Accepted October 7, 2023 ■