Behavioral/Cognitive

Prior Expectations in Visual Speed Perception Predict Encoding Characteristics of Neurons in Area MT

Ling-Qi Zhang and Alan A. Stocker

Bayesian inference provides an elegant theoretical framework for understanding the characteristic biases and discrimination thresholds in visual speed perception. However, the framework is difficult to validate because of its flexibility and the fact that suitable constraints on the structure of the sensory uncertainty have been missing. Here, we demonstrate that a Bayesian observer model constrained by efficient coding not only well explains human visual speed perception but also provides an accurate quantitative account of the tuning characteristics of neurons known for representing visual speed. Specifically, we found that the population coding accuracy for visual speed in area MT ("neural prior") is precisely predicted by the power-law, slow-speed prior extracted from fitting the Bayesian observer model to psychophysical data ("behavioral prior") to the point that the two priors are indistinguishable in a cross-validation model comparison. Our results demonstrate a quantitative validation of the Bayesian observer model constrained by efficient coding at both the behavioral and neural levels.

Key words: **Bayesian model; efficient coding; neural representation; speed prior; Weber's law**

Significance Statement

Statistical regularities of the environment play an important role in shaping both neural representations and perceptual behavior. Most previous work addressed these two aspects independently. Here we present a quantitative validation of a theoretical framework that makes joint predictions for neural coding and behavior, based on the assumption that neural representations of sensory information are efficient but also optimally used in generating a percept. Specifically, we demonstrate that the neural tuning characteristics for visual speed in brain area MT are precisely predicted by the statistical prior expectations extracted from psychophysical data. As such, our results provide a normative link between perceptual behavior and the neural representation of sensory information in the brain.

Introduction

Human perception of visual speed is typically biased and depends on stimulus attributes other than the actual motion of the stimulus. Contrast, for example, strongly affects perceived stimulus speed such that a low-contrast drifting grating typically appears to move slower than a high-contrast grating (Thompson, 1982; Stone and Thompson, 1992; Blakemore and Snowden, 1999; Stocker and Simoncelli, 2006). These biases and perceptual distortions are qualitatively consistent with a Bayesian observer that combines noisy sensory measurements with a prior preference for lower speeds (Simoncelli, 1993; Weiss et al., 2002; Stocker, 2006). Previous work has also shown that by embedding the Bayesian observer within a two-alternative forced choice (2AFC) decision process one can "reverse-engineer" the noise characteristics (i.e., likelihood) and prior expectations of individual human subjects from their behavior in a speed-discrimination task (Stocker and Simoncelli, 2004, 2006). This provided both a quantitative validation of the Bayesian observer model and a normative interpretation of human behavior in visual speed perception tasks, which has been confirmed in various later studies (Welchman et al., 2008; Hedges et al., 2011; Sotiropoulos et al., 2014; Jogan and Stocker, 2015). However, recovering the parameters of a Bayesian observer model from behavioral data is typically difficult because of the intrinsic nonspecificity of its probabilistic formulation, which has been grounds for a critical view of the Bayesian modeling approach altogether (Jones and Love, 2011; Bowers and Davis, 2012). The reverse-engineered speed priors in previous studies indeed all showed large variations across subjects, indicating a potential case of overfitting because of insufficient model constraints (Stocker and Simoncelli, 2006; Hedges et al., 2011; Sotiropoulos et al., 2014; Jogan and Stocker, 2015).

In this article, we show how we addressed this potential problem by developing and validating a tightly constrained Bayesian

Received Sep. 23, 2021; revised Jan. 18, 2022; accepted Jan. 19, 2022.

Author contributions: L.-Q.Z. and A.A.S. designed research; L.-Q.Z and A.A.S. performed research; L.-Q.Z. analyzed data; L.-Q.Z. and A.A.S. wrote the paper.

We thank Greg DeAngelis for sharing electrophysiological MT data; Benjamin Chin for work on an earlier implementation of the model; Emily Cooper, Mike Landy, and Rafael Polania for helpful comments on the manuscript; and the members of the Computational Perception and Cognition Laboratory for many fruitful discussions of the work.

The authors declare no competing financial interests.

Correspondence should be addressed to Alan A. Stocker at astocker@psych.upenn.edu.

https://doi.org/10.1523/JNEUROSCI.1920-21.2022

observer model. We followed a recent proposal to use efficient coding as a constraint that links the likelihood function to the prior expectations of a Bayesian observer (Wei and Stocker, 2012, 2015). The efficient coding hypothesis posits that biological neural systems allocate their limited coding capacity such that overall information transmission is optimized given the stimulus distribution in the natural environment (Barlow, 1961; Laughlin, 1981). It thus establishes a direct relationship between the stimulus distribution and the accuracy of neural representations in sensory systems (Linsker, 1988; McDonnell and Stocks, 2008; Wang et al., 2012; Ganguli and Simoncelli, 2014; Yerxa et al., 2020; Roy et al., 2021). Wei and Stocker (2015) showed how to formulate efficient coding as an information constraint that can be embedded within the probabilistic language of the Bayesian framework. The resulting Bayesian observer model has proven to account for a wide range of phenomena in perception including repulsive biases in perceived visual orientation (Wei and Stocker, 2015; Taylor and Bays, 2018) and the lawful relationship between perceptual bias and discrimination threshold (Wei and Stocker, 2017), but also in more cognitive domains such as subjective preferences judgments (Polania et al., 2019) or the representation of numbers (Prat-Carrabin and Woodford, 2021).

The overall goal of our current work was twofold. First, we aimed for quantitative validation of this new Bayesian observer model in the domain of visual speed perception. We fit the model to speed discrimination data collected by Stocker and Simoncelli (2006). We found that compared with the model in this original study, the new model allowed us to reverse-engineer much more reliable and consistent estimates of subjects' prior beliefs while still accurately accounting for subjects' psychophysical behavior. Second, based on the efficient coding hypothesis, we wanted to test whether the reverse-engineered prior expectations are mirrored in the population-encoding characteristics of neurons in the motion-sensitive area in the primate brain. The middle temporal (MT) area is widely recognized as the cortical area in the primate brain that selectively represents direction and speed of moving visual stimuli (Zeki, 1974; Newsome and Pare, 1988; Britten et al., 1993; Movshon and Newsome, 1996; Priebe et al., 2003). By analyzing single-cell recordings of a large population of MT neurons (Nover et al., 2005), we found that the sensitivity with which visual speed is encoded in this population ("neural prior") is precisely predicted by the prior beliefs extracted from the psychophysical data ("behavioral prior"). Our results provide important quantitative validation of the Bayesian observer model constrained by efficient coding at both the behavioral and neural levels.

Materials and Methods

Behavioral prior: Bayesian observer model constrained by efficient coding

Data. We reanalyzed the 2AFC speed discrimination data from Stocker and Simoncelli (2006). In each trial of the experiment, a subject was shown a pair of horizontally drifting gratings (reference and test), and was asked to choose which one of them was moving faster. The reference grating had one of two contrast levels (0.075, 0.5) and one of six different drifting speeds (0.5, 1, 2, 4, 8, 12°/s). The test grating had one of seven different contrast levels (0.05, 0.075, 0.1, 0.2, 0.4, 0.5, 0.8), and its speed was determined by an adaptive staircase procedure (one-up/one-down). There were 72 different individual conditions (i.e., psychometric curves), and each condition contained 80 trials, resulting in a total of 5760 trials. We excluded one subject (labeled as "Subject 3" in the study by Stocker and Simoncelli (2006)) that was tested at only two contrasts and two test speed levels. While we were able to recover a prior from this subject that was highly consistent with the rest of the subjects, it was not possible to perform a meaningful model comparison and cross-validation because of the low number of trials.

Model formulation. We use the Bayesian observer model by Wei and Stocker (2015) and embed it within a decision process to predict the binary judgments in the 2AFC experiment (Stocker and Simoncelli, 2006).

Specifically, we assume that "encoding" of the stimulus is governed by an efficient coding constraint such that encoding accuracy, measured as the square root of Fisher Information (FI), is proportional to the stimulus prior (Brunel and Nadal, 1998; McDonnell and Stocks, 2008; Wei and Stocker, 2016), hence:

$$\sqrt{I_F}(v) \propto p(v). \tag{1}$$

Encoding is described as the conditional probability distribution $p(m|v)$. It determines how stimulus speed v is transformed probabilistically into a noisy sensory measurement, m. We can satisfy the efficient coding constraint (Eq. 1) by assuming the following encoding distribution:

$$p(m \mid v) = \mathcal{N}\Big(m; \mu = F(v), \sigma^2 = h^2(c)\Big), \tag{2}$$

where $F(v) = \int_{-\infty}^{v} p(v)dv$ is the cumulative density function (CDF) of v. We parameterized the speed prior distribution as the following, modified power-law function:

$$p(v) \propto (|v| + c_1)^{c_0} + c_2, \tag{3}$$

where $c_{0,1,2}$ are free and unconstrained parameters. We also tested alternative parameterizations (see Fig. 4). The scalar $h(c)$ determines the amount of total encoding resources (i.e., the overall magnitude of internal noise) at different contrast levels. It can be shown that:

$$\sqrt{I_F(v)} = F'(v)/\sigma = p(v)/h(c). \tag{4}$$

The total amount of encoding resource is measured by $\int\sqrt{I_F(v)}dv$, which equals to $1/h(c)$. To numerically handle the unbounded nature of a magnitude variable such as speed (compared with a circular variable such as orientation), we added a small constant ($2.5 * 10^{-3}$) to $p(v)$ such that its CDF did not saturate (i.e., $F(v)$ is not upper bounded by 1).

To decode (i.e., estimate) the stimulus v, given a particular sensory representation, m, we first determine the likelihood function:

$$l(v) = p(m \mid v), \tag{5}$$

by considering the encoding distribution (Eq. 2) as a function of v. Applying Bayes' rule and multiplying the likelihood function with the prior $p(v)$ (Eq. 3), we then can compute the posterior as follows:

$$p(v \mid m) \propto l(v)p(v). \tag{6}$$

Assuming an L_0 loss function (Stocker and Simoncelli, 2006), the estimate $\hat{v}$ of the stimulus v is given as:

$$\hat{v} = \underset{\hat{v}}{\operatorname{argmax}}\ l(\hat{v})p(\hat{v}). \tag{7}$$

The estimate represents the optimally decoded stimulus $\hat{v}$ given m. It is a deterministic function of m [implicit in the likelihood function $l(v)$], which we can explicitly express as $\hat{v}(m)$. However, m is not directly observable in a psychophysical experiment. Thus, we marginalize over m to obtain the estimate distribution for a given stimulus, v, as follows:

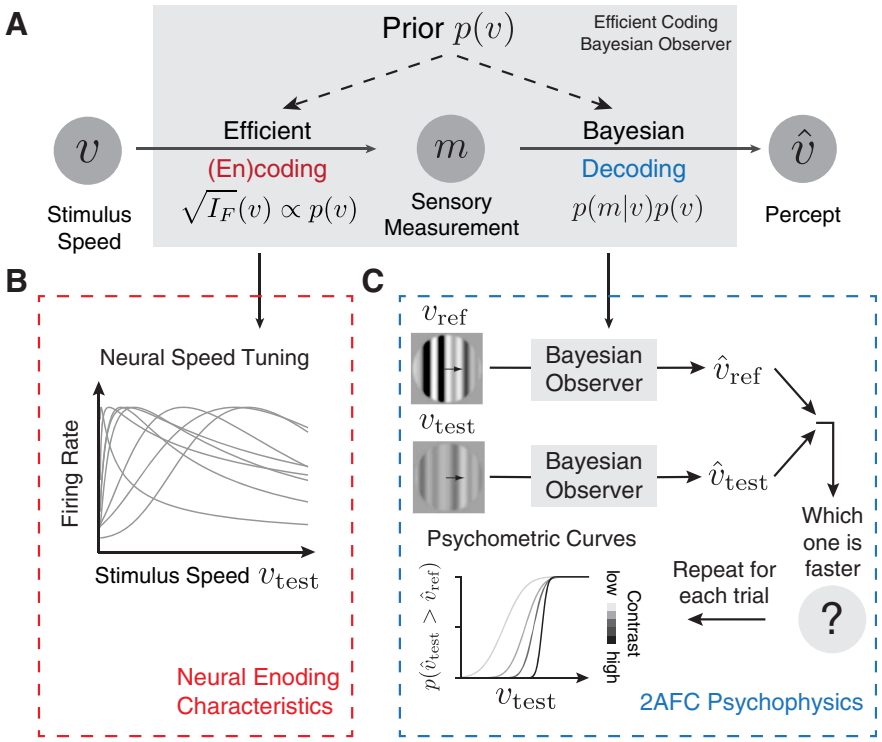

Figure 1. Bayesian observer model constrained by efficient coding. ***A***, We model speed perception as an efficient encoding, Bayesian decoding process (Wei and Stocker, 2012, 2015). Stimulus speed v is encoded in a noisy and resource-limited sensory measurement, m, with an encoding accuracy that is determined by the stimulus prior $p(v)$ via the efficient coding constraint (Eq. 1). Ultimately, a percept is formed through a Bayesian decoding process that combines the likelihood $p(m|v)$ and prior $p(v)$ to compute the posterior $p(v|m)$, and then selects the optimal estimate $\hat{v}$ according to a loss function. Encoding and decoding are linked and jointly determined by the prior distribution over speed. ***B***, Efficient coding determines the accuracy of the neural representation of visual speed (i.e., the tuning characteristics of neurons in area MT). ***C***, Embedding the Bayesian observer within a decision process provides a model to predict psychophysical behavior in a 2AFC speed discrimination task.

$$p(\hat{v} \mid v) = \int p(\hat{v} \mid m)p(m \mid v)dm = \int \delta(\hat{v} - \hat{v}(m))p(m \mid v)dm, \tag{8}$$

where $\delta(\cdot)$ is the Dirac delta function.

In a 2AFC speed discrimination experiment, subjects report a binary decision and not a continuous estimate. We assume subjects make their choices (i.e., which one is faster) by comparing their estimate $\hat{v}_r$ of the reference stimulus with their estimate $\hat{v}_t$ of the test stimulus. For a pair of v_t and v_r across many repeated trials, these choices follow a binomial distribution with the probability of the test stimulus being perceived faster given as follows:

$$p_{v_t,v_r}(\hat{v}_t > \hat{v}_r) = \int_{-\infty}^{+\infty} p(\hat{v}_t \mid v_t) \int_{-\infty}^{\hat{v}_t} p(\hat{v}_r \mid v_r)d\hat{v}_r d\hat{v}_t. \tag{9}$$

Model fitting. If we represent the data in our experiment as N triplets (v_{ir}, v_{it}, k_i), where $k_i \in \{0,1\}$ represents the binary choice, then the overall log-likelihood of the model given the data is:

$$\mathcal{L} = \sum_{i=1}^{N} \{k_i \log[p_{v_{it},v_{ir}}(\hat{v}_{it} > \hat{v}_{ir})] + (1 - k_i)\log[1 - p_{v_{it},v_{ir}}(\hat{v}_{it} > \hat{v}_{ir})]\}. \tag{10}$$

We find the model parameters c_0, c_1, c_2, and $h(c)$ by maximizing $\mathcal{L}$ using the MATLAB *fminsearchbnd* algorithm. Note that the model is highly constrained: for each subject, we jointly fit a single three-parameter prior distribution plus one scalar noise parameter, $h(c)$, for each of the seven contrast levels to the data from all 72 conditions.

Alternative prior parameterization. To assess the consistency and stability of our reverse-engineered prior distributions, we also tested the following two alternative parameterizations (see Fig. 4):

- a Gamma distribution: $p(v; \alpha, \beta) \propto |v|^{\alpha-1} e^{-\beta|v|}$
- a piece-wise log-linear function with 18 sample points, $v^*_{1:18}$, equally distributed in logarithmic space in the range v = 0...50°/s. Each corresponding $p(v^*_{1:18})$ value is a free prior parameter; prior density values are linearly interpolated between those values.

For comparison, we also fit a Gaussian prior with $p(v; \sigma^2) = \mathcal{N}(v; \mu = 0, \sigma^2)$.

Weber's law and power-law prior. With our model, it is possible to analytically predict discrimination threshold Δ_v and Weber fraction $\frac{\Delta_v}{v}$ for any given prior distribution. It has been shown that the discrimination threshold is inversely proportional to the square root of FI (Seriès et al., 2009; Wei and Stocker, 2017):

$$\Delta_v \propto \frac{1}{\sqrt{I_F}}. \tag{11}$$

According to the efficient coding constraint (Eq. 1), we can substitute $\sqrt{I_F}$ with $p(v)$ and find:

$$\Delta_v \propto \frac{1}{p(v)}. \tag{12}$$

This equation allows us to predict discrimination threshold for any prior density (up to a scale factor). For the modified power-law prior with exponent $c_0 = -1$ and $c_2 = 0$ (Eq. 3), we can find the following:

$$\Delta_v \propto (v + c_1). \tag{13}$$

By further setting $c_1 = 0$ we obtain $v \propto \Delta_v$, which is the definition of Weber's law (i.e., a constant Weber fraction). For nonzero c_1, the Weber fraction changes to:

$$\frac{\Delta_v}{v} \propto \left(1 + \frac{c_1}{v}\right). \tag{14}$$

At high speeds, $\frac{c_1}{v} \approx 0$, and thus the Weber fraction is constant. At low speeds, $\frac{c_1}{v} \to \infty$, which causes $\frac{\Delta_v}{v}$ to increase.

Our efficient coding constraint implies that the stimulus is transformed according to the CDF of v (Eq. 2). For a power-law prior with exponent $c_0 = -1$ and $c_2 = 0$, the CDF is as follows:

$$\int z_1 (v + c_1)^{-1} dv = z_1 \log(v + c_1) + z_2, \tag{15}$$

which is precisely the logarithmic transformation that has been previously used for describing the speed tuning of MT neurons (Nover et al., 2005).

Neural prior: MT encoding analysis

Data. We reanalyzed the electrophysiological recording data from the study by Nover et al. (2005). Neurons in area MT of several macaque monkeys were individually identified. Each identified neuron was then tested with a random-dot motion stimulus moving with one of eight speeds (0, 0.5, 1, 2, 4, 8, 16, 32°/s). Stimulus location, direction, size, and

disparity were individually optimized for each neuron. Every stimulus speed was presented three to seven times. We considered the mean firing rate over the entire stimulus duration (1.5 s) as the single-trial response of a neuron. We analyzed a total of 480 neurons.

Population Fisher Information. Following Nover et al. (2005), we fit the mean firing rate of each neuron as a function of stimulus speed with a Gaussian tuning curve in log-speed:

$$R(v) = R_0 + A\exp\left(-\frac{\log[q(v)]^2}{2\sigma^2}\right), \qquad (16)$$

where $q(v) = \frac{v+v_0}{v_p+v_0}$. Parameters R_0, A, σ^2, v_0, and v_p are determined by minimizing the sum of squared difference of the observed and predicted firing rates. A maximum-likelihood fit assuming Poisson distributed firing rate variability produced very similar results.

We computed the population FI for different assumptions about the response variabilities of neurons and their pair-wise noise correlations within the population. First, we assumed that response noise is independent between neurons in the population, and response variability is well described by a Poisson process. In this case, the population FI is calculated as follows:

$$I_F(v) = \sum_{i=1}^{N} \frac{[R_i'(v)]^2}{R_i(v)}. \qquad (17)$$

The neural prior (the prior that corresponds to the measured MT encoding precision, assuming efficient encoding) is then equivalent to the normalized square-root of FI, thus:

$$p(v) = \frac{\sqrt{I_F(v)}}{\int\sqrt{I_F(v)}dv}. \qquad (18)$$

As in the study by Nover et al. (2005), we also repeated the above analysis using an alternative tuning-curve model (Gamma distribution function) and obtained very similar results.

Next, we estimated the population FI by adjusting the Poisson model with an explicit estimate of the Fano factor F_i for each neuron as follows:

$$I_F(v) = \sum_{i=1}^{N} \frac{[R_i'(v)]^2}{F_i R_i(v)}, \qquad (19)$$

where F_i was obtained by linearly regressing the firing rate variance of the neuron against its firing rate mean. Finally, we computed the linear Fisher Information (Kanitscheider et al., 2015; Kohn et al., 2016), as follows:

$$I_F(v) = \vec{R}'(v)^T \Sigma^{-1} \vec{R}'(v), \qquad (20)$$

where Σ is the noise correlation matrix and is the identity matrix for the independent noise case. To understand the effect of speed tuning preference-dependent noise correlations observed in area MT (Huang and Lisberger, 2009), we adopted the following limited-range correlation model:

$$\Sigma_{ij} = \sigma^2 \rho_{ij}$$
$$\rho_{ij} = \exp(-|\Delta_{ij}|/L), \qquad (21)$$

where Δ_{ij} is the difference in log-speed preference between neuron i and neuron j, σ^2 is the noise variance, and L is the overall correlation strength (Abbott and Dayan, 1999). For the simulations seen in Figure 7*E*, we set $\sigma^2 = 1$ as the prior is only determined by the shape of the population FI and L = [0.5, 1.0, 2.5] for low, medium, and high correlation strengths.

Cross-validation. We performed a fivefold cross-validation procedure. The trial data for each condition were first randomly and equally divided into five groups. For each group, the model was fit to the data of the remaining four groups (training), and then evaluated on the group's data (validation). Model validation performance was measured as the log-likelihood of the fit model given the validation data. The entire

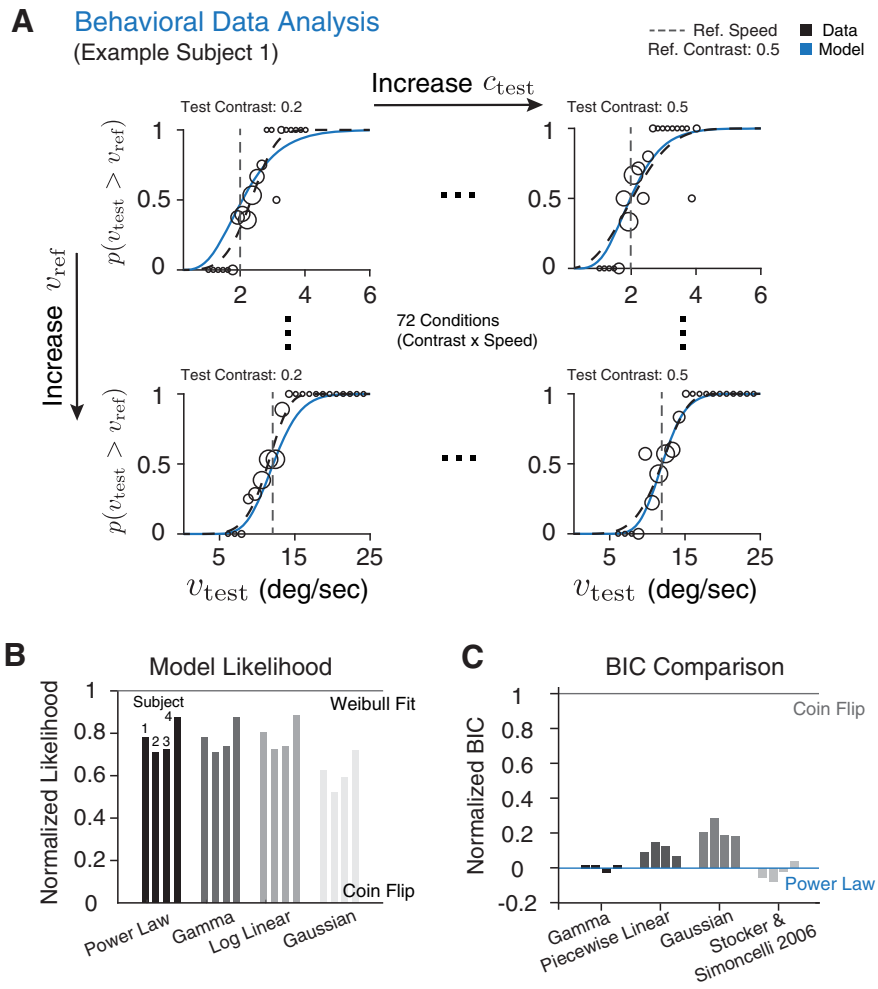

Figure 2. Extracting the behavioral prior. ***A***, We jointly fit the Bayesian observer model to psychophysical data across all contrast and reference speed conditions (72 conditions total). Shown are a few conditions for exemplary Subject 1. Circle sizes are proportional to the number of trials at that test speed. The dashed curves are Weibull fits to each condition, and the solid blue curves represent the model prediction. See the Data availability section for instructions to create a full display of psychometric curves and model fits of all 72 conditions for individual subjects. ***B***, Log-likelihood values of the best-fitting model for each subject using four different prior parameterizations including a power-law function, Gamma distribution, piece-wise log-linear function, and Gaussian distribution, respectively. Values are normalized to the range set by a coin flip model (lower bound) and Weibull fits to individual psychometric curves (upper bound). ***C***, The relative BIC values for the different parameterizations as well as the original, less constrained Bayesian observer model by Stocker and Simoncelli (2006). Values are normalized to the range set by the efficient Bayesian observer model with power-law parameterization and the coin flip model (lower is better). For details, see Materials and Methods.

Table 1. Fit parameter values of the prior density $p(v) \propto (|v|+c_1)^{c_0}+c_2$ and the contrast-dependent noise $\sigma = h(c)$ for every subject

Subject	c_0	c_1	c_2	$h(0.05)$	$h(0.075)$	$h(0.10)$	$h(0.20)$	$h(0.40)$	$h(0.50)$	$h(0.80)$
1	−0.790	0.003	$6 * 10^{-5}$	0.035	0.027	0.028	0.019	0.016	0.013	0.012
2	−0.867	0.002	10^{-8}	0.045	0.041	0.038	0.030	0.023	0.020	0.010
3	−1.045	0.220	$1 * 10^{-5}$	0.061	0.045	0.041	0.029	0.017	0.019	0.026
4	−1.097	0.248	$7 * 10^{-6}$	0.043	0.040	0.036	0.028	0.018	0.017	0.016

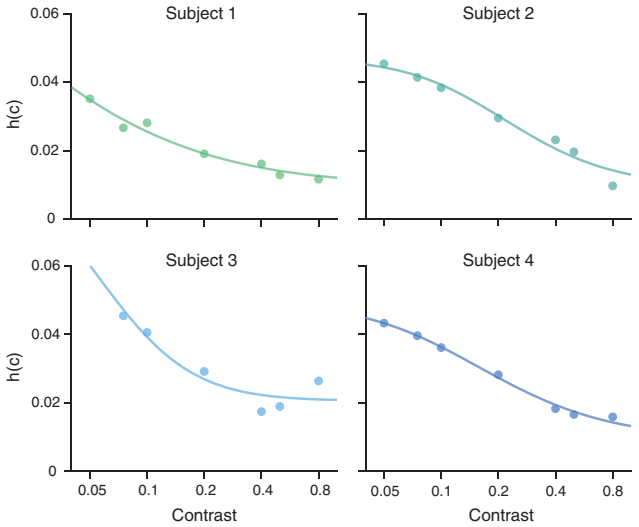

Figure 3. Contrast-dependent noise. Fit parameter values are shown as a function of stimulus contrast plotted for every subject. Bold lines represent fits with a parametric description of the contrast response function of cortical neurons $h(c) = [r_{max}c^q/(c^q+c_{50}^q)+r_{base}]^{-1/2}$ (Albrecht and Hamilton, 1982; Sclar et al., 1990; Heuer and Britten, 2002).

procedure was repeated 20 times, resulting in 100 estimates of the model validation likelihood. For the behavioral prior condition, we considered the full observer model using the fit prior for that run. For the neural prior condition, we assumed that the prior is fixed and equal to the prior extracted from the population FI analysis with only the contrast-dependent noise parameters being fit on each run. The same procedure was used to compute the validation likelihoods of the original, less constrained Bayesian observer model (Stocker and Simoncelli, 2006); and of individual Weibull fits to every condition. The log-likelihood values seen in Figure 8*B* were normalized to the range set by a lower bound given by the log-likelihoods of a coin-flip model for the decision (i.e., a model with a fixed decision probability of 0.5) and an upper bound determined by the values of the Weibull fits.

Data availability

Data and analysis code, including the instruction to create a full display of the psychometric curves and model fits for individual subjects, are available through GitHub (https://github.com/cpc-lab-stocker/Speed_Prior_2021).

Results

We model speed perception as an efficient encoding, Bayesian decoding process (Fig. 1*A*). On any given trial, the speed v of a visual stimulus is represented by a noisy and bandwidth-limited sensory measurement, m. Following Wei and Stocker (2012, 2015), we assume that encoding of the stimulus is governed by an efficient coding constraint (Eq. 1) such that encoding accuracy, measured as the square root of FI, is proportional to the stimulus prior $p(v)$ (Brunel and Nadal, 1998; McDonnell and Stocks, 2008; Wei and Stocker, 2016). This constraint promotes a more accurate encoding of speeds for which the prior density is high. It determines the observer's uncertainty about the actual stimulus speed given a particular sensory measurement [i.e., the likelihood function $p(m|v)$]. For "decoding", this likelihood function is combined with the stimulus prior $p(v)$, resulting in the posterior $p(v|m)$. Last, a percept, $\hat{v}$ (i.e., an estimate), is computed based on the posterior and a loss function (for details, see Materials and Methods).

A unique feature of the new model is that the stimulus distribution jointly determines encoding and decoding of the Bayesian observer model (Wei and Stocker, 2012). Thus, both the encoding characteristics of neurons representing visual speed (Fig. 1*B*), and the psychophysical behavior of subjects in speed perception (Fig. 1*C*) should be consistent with the prior belief of the observer about the statistical regularities of visual speed.

Extracting the behavioral prior

We fit our model to the psychophysical speed discrimination data collected by Stocker and Simoncelli (2006). On each trial of their experiment, subjects were shown a pair of horizontally drifting gratings (reference and test stimulus) and were asked to choose which one was moving faster (Fig. 1*C*). For each combination of stimulus contrast and reference speed, a full psychometric curve was measured by repeating the trials at different test speeds chosen by an adaptive staircase procedure. A total combination of 72 conditions representing reference and test stimuli at different speeds and contrast levels were tested, resulting in 72 different psychometric functions (for details, see Materials and Methods; Stocker and Simoncelli, 2006).

In contrast to the original model (Stocker and Simoncelli, 2006), the new observer model directly links the likelihood function and the prior distribution (Wei and Stocker, 2012). Thus, perceived speed is fully determined by subjects' prior expectations and a contrast-dependent internal noise parameter that reflects the total amount of represented sensory information (Wei and Stocker, 2015, 2016). Our goal was to find the prior distribution $p(v)$ and the noise parameters $h(c)$ that best accounted for subjects' individual perceptual behavior. To fit the observer model, we embedded it within a binary decision process (Fig. 1*C*). On each trial, speed estimates for both the reference and the test stimuli are obtained, and then subjects are assumed to respond according to which estimate is faster. Entire psychometric functions are predicted by marginalizing over the unobserved sensory measurement.

We jointly fit our model for every subject to all 72 conditions using a maximum-likelihood procedure. The free parameters of the model consisted of a parametric description of the prior and one noise parameter for each stimulus contrast. Following previous studies (Stocker and Simoncelli, 2006; Hedges et al., 2011; Jogan and Stocker, 2015), we parameterized the prior distribution as a modified power-law function (Eq. 3). Figure 2*A* shows the data and model fit for a few example conditions for exemplary Subject 1. Fit parameter values for all subjects are listed in Table 1. Overall, the model predicts psychometric curves that are similar to those obtained from fitting a Weibull function. The log-likelihood of the

new model is close to that of separate Weibull fits to every individual condition (Fig. 2*B*). Figure 2*C* further illustrates that the new model performs as well as the original, less constrained Bayesian observer model (Stocker and Simoncelli, 2006). A more detailed model comparison using cross-validation is provided in a later section.

Importantly, the reverse-engineered prior expectations are much more consistent across subjects than those obtained from using the original model (Stocker and Simoncelli, 2006; Hedges et al., 2011; Sotiropoulos et al., 2014; Jogan and Stocker, 2015). The exponent c_0, for example, is now close to a value of −1 for every subject rather than varying over an order of magnitude (Table 1). Furthermore, values of the contrast-dependent noise parameter monotonically decrease as a function of contrast as expected and are consistent with the functional description of the contrast response curve of cortical neurons (Fig. 3).

To test the impact of choosing a power-law parameterization for the prior distribution (Eq. 3), we performed model fits using two other parameterizations with increasing degrees of freedom (i.e., a Gamma distribution and a piece-wise log-linear function), and also a Gaussian prior for comparison (see Materials and Methods). The model fits well for all but the Gaussian prior, resulting in similar log-likelihood values (Fig. 2*B*), although the Bayesian information criterion (BIC) value is higher for the log-linear parameterization because of its large number of parameters (Fig. 2*C*). Crucially, however, the shapes of the fit prior distributions are very similar across the different parameterizations, all exhibiting a power-law like, slow-speed preferred distribution (Fig. 4). The obvious exception is the Gaussian parameterization because it is unsuited to approximate a power-law function.

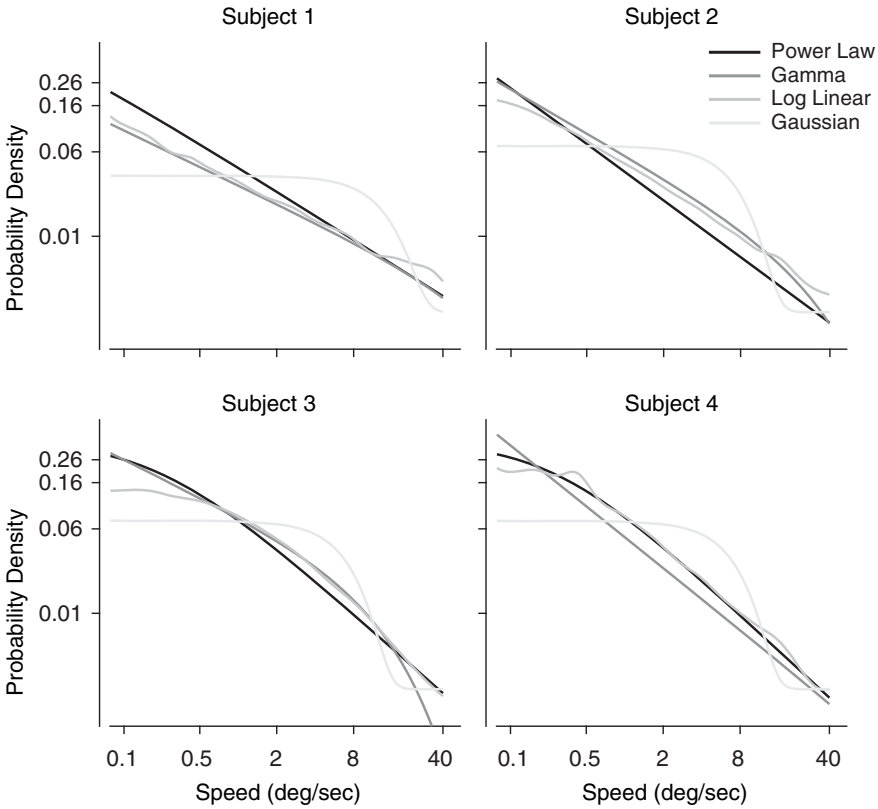

Figure 4. Effect of prior parameterization. Shown are the best fit prior density functions for each subject using four different parameterizations including a power-law function, a Gamma distribution, a piece-wise log-linear function, and a Gaussian distribution, respectively. Note that the Gaussian provides a relative poor fit of the data (see also Fig. 2*B*).

We further validated our model by comparing its predictions for contrast-induced biases and discrimination thresholds to subjects' data. To quantify bias, we computed the ratio of test speed to reference speed at the point of subjective equality (PSE; defined as the 50% point of the psychometric curve). If a lower-contrast test stimulus is indeed perceived to be slower, then its physical speed will need to be higher to match the perceived speed of the higher-contrast reference. Thus, a contrast-induced slow-speed bias is manifested by a PSE ratio greater than one when the test contrast is lower than the reference, and vice versa. As shown in Figure 5*A*, subjects clearly underestimated the speeds of low-contrast stimuli, an effect that occurred at any contrast level and speed. Furthermore, subjects' thresholds increase monotonically with speed (Fig. 5*B*). While they follow Weber's law at higher speeds, they deviate from a constant Weber fraction at slow speeds, which is well documented (McKee et al., 1986; De Bruyn and Orban, 1988; Stocker and Simoncelli, 2006). Our new model is able to capture both the contrast-induced slow-speed bias and the discrimination threshold behavior with an accuracy comparable to the original model (Stocker and Simoncelli, 2006; see Fig. 2*C*).

It is worth noting that the predicted higher threshold values for the low-contrast stimulus condition are not particularly evident in the data (Fig. 5*B*). Previous studies, however, have convincingly demonstrated that lower stimulus contrasts lead to higher speed discrimination thresholds under various stimulus configurations (Panish, 1988; Turano and Pantle, 1989; Horswill and Plooy, 2008; Champion and Warren, 2017). Thus, we believe that the experimental design in the study by Stocker and Simoncelli (2006), in particular the deliberate compromise in choosing a low number of trials per condition (only 80 trials per psychometric curve) to test subjects over a large range of different contrast/speed combinations, may be responsible for the noisy, overlapping threshold estimates. The ability to obtain reliable threshold estimates was further limited by the use of a staircase procedure optimized for inferring the PSE rather than the slope of the psychometric curves (for details, see Stocker and Simoncelli, 2006). Future investigations will be required to fully resolve this discrepancy.

Despite its constrained nature, the model can well account for individual differences across subjects. Differences in the values of the contrast-dependent noise parameter determine individual variations in bias and threshold magnitude. In addition, when the prior exponent is close to −1, the PSE ratios are mostly constant across different speeds (e.g., Fig. 5*A*, Subjects 3 and 4; Wei and Stocker, 2017), whereas an exponent larger than −1 predicts relative biases that decrease for higher stimulus speeds (e.g., Fig. 5*A*, Subject 1).

Previous models have mainly focused on the contrast-induced speed bias and how it can be attributed to a slow-speed prior that shifts the percept toward slower speeds for increasing levels of sensory uncertainty (Hürlimann et al., 2002; Weiss et

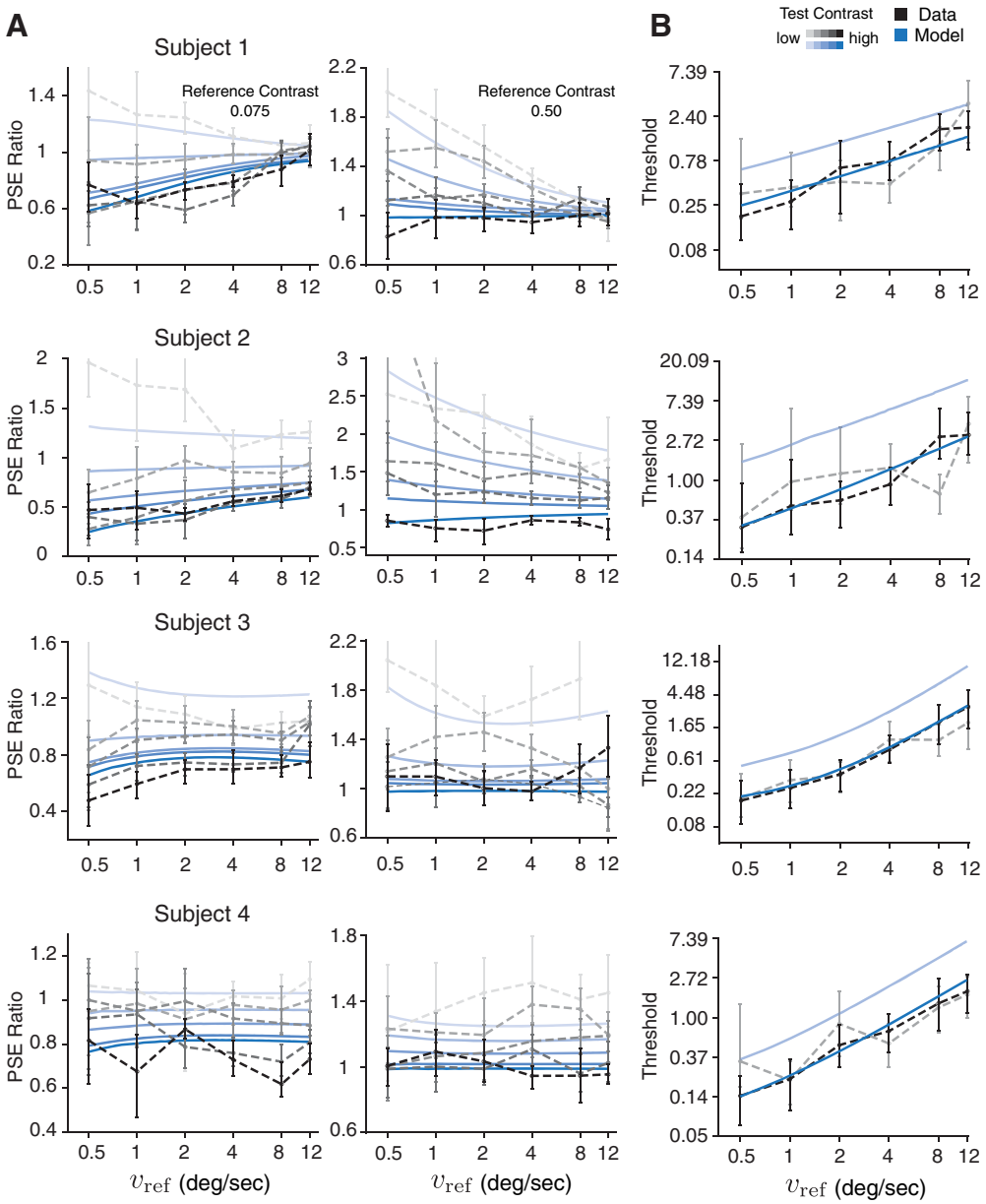

Figure 5. Predicted contrast-dependent bias and discrimination threshold. ***A***, Ratios of the test relative to reference speed at the PSE extracted from individual Weibull fits to the data (black) and our model (blue). Shading levels correspond to different contrast levels (0.05, 0.1, 0.2, 0.4, 0.8) of the test stimulus (darker means higher contrast). The reference stimulus has a contrast of 0.075 in the left column, and 0.5 in the right column. ***B***, Speed discrimination thresholds, defined as the difference in stimulus speed at the 50% and 75% points of the psychometric curve, at two different contrast levels (0.075, 0.5), extracted from individual Weibull fits to the data (black) and our model (blue). Error bars indicate the 95% confidence interval across 500 bootstrap runs.

al., 2002; Stocker, 2006; Stocker and Simoncelli, 2006; Hedges et al., 2011; Lakshminarasimhan et al., 2018; Rokers et al., 2018). While this is still the case for our new model, the monotonic increase in threshold is now also a direct consequence of the slow-speed prior: since higher speeds are less likely, efficient coding dictates that less neural resources are allocated for their representation, resulting in a larger threshold. In fact, the predicted Weber fractions based on subjects' reverse-engineered priors closely resemble previous psychophysical measurements (Fig. 6; also see Materials and Methods).

Extracting the neural prior

The efficient coding constraint of the model predicts that the neural encoding of visual speed should reflect the stimulus prior distribution (Wei and Stocker, 2012, 2016; Ganguli and Simoncelli, 2014). Thus, if our new model is correct, then the reverse-engineered behavioral prior should be a good predictor of the neural encoding characteristics of visual speed. Specifically, we expect the neural encoding accuracy, measured as the square root of the neural population FI, to match the extracted speed prior (Eq. 1).

To test this prediction, we analyzed the encoding characteristics of a large population of neurons in area MT. Our analysis was based on single-cell recorded data from the macaque brain (Nover et al., 2005). The data contained repeated spike counts from 480 MT neurons responding to random dot motion stimuli moving at eight different speeds. Following the original study (Nover et al., 2005), we fit a log-normal speed-tuning curve model for each neuron in the dataset (Fig. 7*A*). This tuning curve model accurately described the mean firing rates of the majority of the neurons (Fig. 7*B*). Under the assumption that neural response variability was well captured by a Poisson distribution,

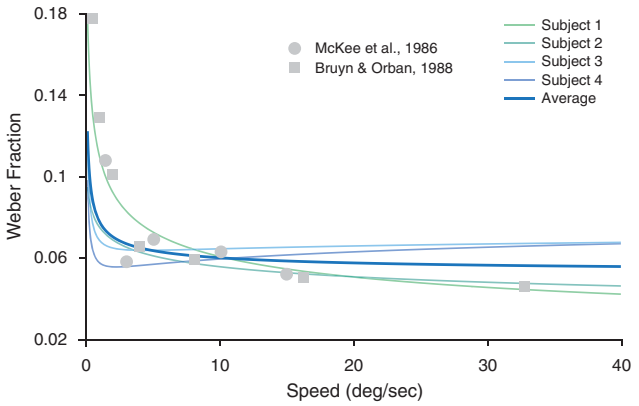

Figure 6. Predicted Weber fraction. Shown are the predicted Weber fractions $\Delta v/v$ based on the reverse-engineered behavioral priors of the four individuals and the average subject, compared with previously reported psychophysical measurements (McKee et al., 1986; De Bruyn and Orban, 1988). We can analytically show that the modified power-law prior with an exponent $c_0 = -1$ will predict both the constant Weber fraction at higher speed and its deviation at slow speeds (see Materials and Methods). Note that since the Weber fraction is predicted up to a factor, the predictions are scaled to the level of the data. De Bruyn and Orban (1988) also found deviations from Weber's law at extremely high speeds (256 deg/s), which is not depicted here.

it is then straightforward to compute the FI of individual neurons (Fig. 7*A*; see Materials and Methods).

Given the inherent limitations of single-unit recorded data, assumptions about the noise and its correlation structure in the population are necessary to compute the population FI (Abbott and Dayan, 1999; Averbeck et al., 2006; Moreno-Bote et al., 2014; Kohn et al., 2016). We first considered the noise to be independent among the neurons in the recorded population. In this simplified scenario, the population FI is the sum of the FI of individual neurons (Fig. 7*C*). The shape of the resulting population FI is very close to a power-law function; that is, when plotted on a log-log scale, it closely resembles a straight line. Two slightly different methods of calculating the FI of individual neurons, either by estimating the Fano factor explicitly or by computing the linear FI (Kanitscheider et al., 2015; Kohn et al., 2016), produced nearly identical estimates of the shape of the population FI (Fig. 7*D*; see Materials and Methods).

Further, assuming a correlation pattern that is speed independent simply reduces the magnitude of the population FI but does not change its overall shape compared with the independent noise assumption. However, speed tuning-dependent noise correlations between pairs of neurons have been reported for area MT (Huang and Lisberger, 2009). Thus, to assess the potential impact of such correlations (Zohary et al., 1994), we computed the linear population FI with a limited-range correlation model based on the relative speed preferences of individual neurons (Abbott and Dayan, 1999; see Materials and Methods). We found that although the magnitude of FI decreases with increasing correlation strength, the shape of the population FI is largely invariant within a large range of simulated correlation strengths (Fig. 7*E*). The reason these correlations have little effect on the shape of the population FI is that the tuning characteristics of MT neurons are relatively "homogeneous" (i.e., the parameters of the tuning curve, such as the tuning width, are mostly independent of speed preference) and close to uniformly tile the logarithmic speed space (Nover et al., 2005). Thus, we argue that, given the available evidence, estimating the shape of the population FI assuming independent noise is a reliable approximation.

The efficient coding constraint makes the additional prediction that the overall magnitude of the population FI corresponds to the total represented sensory information and, thus, should be directly related to the contrast-dependent noise parameter of our observer model (Wei and Stocker, 2015, 2016; Noel et al., 2021). Although the contrast-dependent noise parameter values are consistent with the typical contrast response function of cortical neurons (Fig. 3), a rigorous test of the prediction requires characterization of MT speed encoding at different levels of stimulus contrasts, which is something the current data do not provide. Preliminary neural data (Stocker et al., 2009) suggest, however, that stimulus contrast indeed simply scales the population FI without changing its shape. This is intriguing given the well documented diversity and heterogeneity by which stimulus contrast affects the shape and position of speed-tuning curves in area MT (Pack et al., 2005; Krekelberg et al., 2006; Stocker et al., 2009).

Comparing the behavior and neural prior

Finally, we compared the extracted behavioral and neural priors. If our observer model is correct, then the prior expectation with which a subject perceives the speed of a moving stimulus should be quantitatively identical to, and thus predictive of, the stimulus distribution that neural encoding is optimized for. Figure 8*A* shows the extracted behavioral prior of every subject and the neural prior. The prior distributions are indeed very similar and are consistent with a power-law function with an exponent of approximately −1.

To quantitatively assess the effective similarity between the behavior and neural prior, we constructed a "neural observer" model for which the prior was fixed to be the neural prior extracted from the MT data; only the contrast-dependent noise parameters $h(c)$ were free parameters. We used a cross-validation procedure to compare this neural observer with the unconstrained observer model and the original model (Stocker and Simoncelli, 2006). As illustrated in Figure 8*B*, the validation performances are highly similar across all three models and closely match the performance of individual Weibull fits. This comparison demonstrates several aspects. First, it confirms that the behavioral and neural priors are behaviorally indistinguishable and thus effectively equivalent; if the neural data did not exist, we would have been able to accurately predict the encoding accuracy of MT neurons at the population level. Second, it highlights the excellent quality of the Bayesian observer model as its account of human behavior is close to that of the best possible parametric description of the data (i.e., individual Weibull fits). And finally, it shows that the complexity of our new observer model is appropriate and does not lead to overfitting.

Weber–Fechner law

The power-law shape of the behavioral and neural prior distribution also sheds a new normative light on the interpretation of Weber's law. Famously, Fechner proposed that Weber's law emerges from a logarithmic neural encoding of a stimulus variable (Fechner, 1860). Neural encoding of visual speed in area MT is indeed considered logarithmic (Nover et al., 2005; Burge and Geisler, 2015): When analyzed in the logarithmic speed domain, the tuning curves of MT neurons are approximately bell shaped and scale invariant, and tile the stimulus space with nearly uniform density (Nover et al., 2005; Pack et al., 2005).

With our new model, we have already demonstrated that a modified power-law prior (Eq. 2) with an exponent of approximately −1 can well account for Weber's law behavior and the

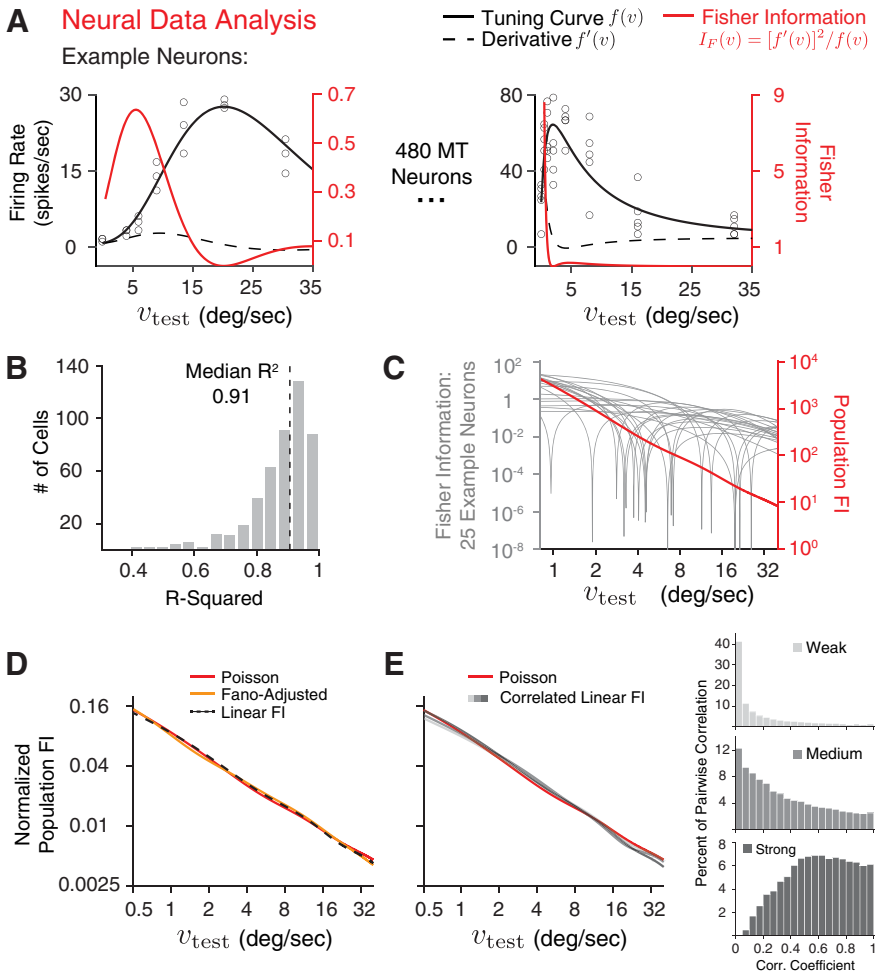

Figure 7. Extracting the neural prior. According to the efficient coding constraint (Eq. 1), the population FI should directly reflect the prior distribution of visual speed. ***A***, Single-trial mean firing rates as a function of stimulus speed v_{test} (dots) shown for two example MT neurons from the dataset (Nover et al., 2005), together with their fit log-normal tuning curves (black, left *y*-axis) and corresponding FI (red, right *y*-axis) assuming a Poisson noise model. ***B***, Histogram of the goodness of the log-normal tuning curve fit across all neurons measured by R^2. ***C***, Individual FI of 25 example neurons (gray, left *y*-axis) and the population FI (red, right *y*-axis), calculated as the sum of the FI over all 480 neurons in the dataset. ***D***, The normalized square root of population FI assuming independent Poisson noise (red), adjusted for the variance by estimating the Fano factor explicitly (orange), and the linear Fisher Information (dashed black). ***E***, The normalized square root of population FI assuming independent Poisson noise (red) and the linear population FI based on a speed tuning-dependent, limited-range noise correlation model (Abbott and Dayan, 1999) for three levels of correlation strength as illustrated by the histogram of pairwise correlation coefficients on the right. For details, see Materials and Methods.

deviation from it at slow speeds (Fig. 6). We can now show that this power-law prior also predicts logarithmic neural encoding. Specifically, one way to implement the efficient coding constraint (Eq. 1) is to assume a homogeneous neural encoding (i.e., identical tuning curves that uniformly tile the sensory space) of the variable of interest transformed by its CDF (Ganguli and Simoncelli, 2010; Wei and Stocker, 2012; Wang et al., 2016), which is sometimes also referred to as histogram equalization (Acharya and Ray, 2005). With the exponent $c_0 = -1$, the CDF of the speed prior is exactly the logarithmic function that well described MT tuning characteristics (Nover et al., 2005). Thus, the Bayesian observer model constrained by efficient coding provides a normative explanation for both Weber's law and the logarithmic encoding of visual speed in area MT.

Discussion

We presented a Bayesian observer model constrained by efficient coding for human visual speed perception. We fit this model to existing human 2AFC speed discrimination data recorded over a wide range of stimulus contrasts and speeds, which allowed us to reverse-engineer the behavioral prior that best accounts for the psychophysical behavior of individual subjects. In addition, we analyzed the population encoding the accuracy of visual speed based on an existing set of single-cell recordings in area MT, thereby extracting the neural prior according to the efficient coding constraint of our observer model. We found that the behavioral prior estimated from the psychophysical data accurately predicts the neural prior reflected in the encoding characteristics of the MT neural population.

Our results provide a successful, quantitative validation of the Bayesian observer model constrained by efficient coding in the domain of visual speed perception. We demonstrate that this model can accurately account for the behavioral characteristics of bias and threshold in visual speed perception if subjects' prior belief about the statistical distribution of visual speed resembles a power-law function with an exponent of approximately −1. Cross-validation revealed no significant difference between the best possible parametric description of the behavioral data (i.e., individual Weibull fits) and our model fits. Compared with the original, more flexible Bayesian model formulation (Stocker and Simoncelli, 2006), the added efficient coding constraint results in estimates of behavioral priors that are not only much more consistent across subjects but are also remarkably predictive of the population-encoding characteristics of neurons in the motion-sensitive area MT in the primate brain. Our work substantially strengthens the evidence for the slow-speed prior interpretation of motion illusions (Weiss et al., 2002; Stocker and Simoncelli, 2004, 2006; Welchman et al., 2008; Sotiropoulos et al., 2014; Jogan and Stocker, 2015; Senna et al., 2015; but see Rideaux and Welchman, 2020) by the demonstrated quantitative support from electrophysiological neural data.

We also offer an explanation for why certain perceptual variables have a logarithmic neural representation and thus follow Weber's law (Fechner, 1860). According to our model, logarithmic encoding and Weber's law both follow from the efficient representation of a perceptual variable with a power-law prior distribution. We thus predict that perceptual variables that conform to Weber's law have power-law distributions with an exponent of approximately −1 and are logarithmically encoded in the brain (although alternative encoding solutions that satisfy the efficient coding constraint are possible; Wei and Stocker, 2015). Indeed, perceptual variables that are known to approximately follow Weber's law, such as weight (Fechner et al., 1966), light intensity (Treisman, 1964), and numerosity (Nieder and Miller, 2003; Cheyette and Piantadosi, 2020; Prat-Carrabin and Woodford,

2021), exhibit heavy tails in their statistical distributions under natural environmental conditions (Dehaene and Mehler, 1992; Dror et al., 2004; Peters et al., 2015; Piantadosi and Cantlon, 2017), a defining feature of a power-law function. Conversely, any deviation from Weber's law and the logarithmic encoding should be reflected in deviations of the statistical stimulus distributions from a power-law function. Future studies of natural stimulus statistics, modeling of psychophysical data, and neural recordings will be needed to further and more quantitatively validate the generality of this prediction.

Other recent work has used efficient coding assumptions to link perceptual discriminability to the statistical prior distribution of perceptual variables (Gu et al., 2010; Ganguli and Simoncelli, 2016; Sims, 2018). Our approach is a substantial step forward in that it embeds this link within a full behavioral observer model. Thus, rather than relying on a single summary metric of behavior (i.e., discrimination threshold), the predictions of our model are constrained by the full richness of the psychophysical data (i.e., every single datum in the set). This not only provides a much more stringent test of the observer model, but also permits more robust and precise predictions of neural coding accuracy and priors.

The presented comparison between behavioral and neural priors is limited to the extent that there were substantial experimental differences between the behavioral and neural data. For example, we compared human with nonhuman primate data and estimated the neural tuning characterization based on single-cell responses to random-dot motion rather than the broadband, drifting grating stimuli used in the psychophysical experiment. Yet, the surprisingly accurate match of the extracted neural and behavioral priors suggests that they may reflect the "true" stimulus prior, in which case these differences in stimuli and model systems should indeed matter little because the stimulus prior is largely a property of the environment and not the observer or the particular stimulus pattern. Recent studies have demonstrated that it is possible to quantitatively characterize the accuracy with which a perceptual variable is represented in the human brain using voxel-level encoding models of functional magnetic resonance imaging signals (Van Bergen et al., 2015). Future work may exploit this technique to validate and potentially refine our estimates of the neural prior in human subjects. Such work would also permit a more thorough investigation of individual differences at both behavioral and neural levels through matched task and stimulus designs.

The specific shape of the extracted neural and behavioral prior depends on the assumed efficient coding objective. The chosen efficient coding constraint (Eq. 1) results from the objective to maximize the mutual information between neural representation and stimulus (Wei and Stocker, 2016). It is possible, although unlikely given the exceptional good quantitative match, that with a different combination of efficient coding constraint and loss function, a power-law prior with a different exponent could also be consistent with both the behavioral and neural data (see Wang et al., 2012; Morais and Pillow, 2018; Rast and Drugowitsch, 2020). This is difficult to validate conclusively without access to an accurate characterization of the stimulus prior of visual speed, as the search space over all possible combinations is extensive. Previous work has shown, however, that the encoding characteristics in early visual cortex for visual stimulus variables for which good estimates of the stimulus prior exist (e.g., luminance contrast and local orientation) are closely accounted for by the mutual information maximization objective (Wang et al., 2012).

An important assumption of our observer model is that the neural and behavioral priors not only match but are also consistent with the statistical distribution of visual speeds in the natural environment ("stimulus prior"). As such, our results predict that the stimulus prior approximates a power-law distribution that lies within the range given by the neural and behavioral priors shown in Figure 8*A*. However, empirical validation of this prediction by directly measuring the stimulus prior distribution is rather challenging. Object motion but also the ego motion of the

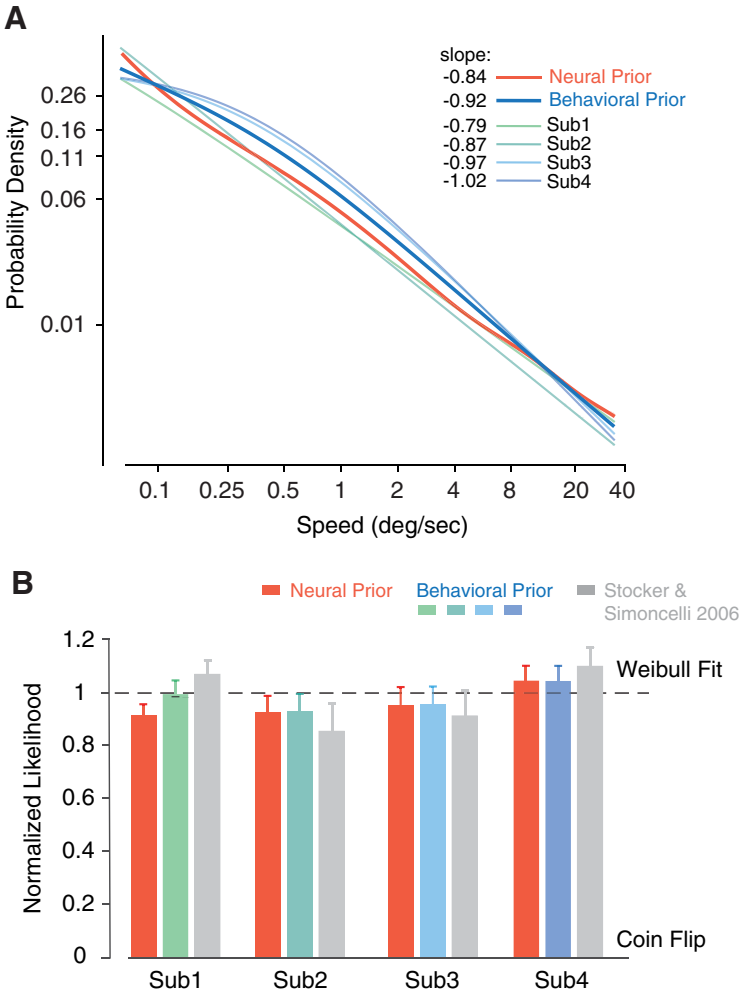

Figure 8. Comparing neural and behavioral prior. ***A***, Reverse-engineered behavioral priors of every subject and their average (dark blue) superimposed by the neural prior (red). Slope values are computed from a linear fit of the curves in log–log coordinates. ***B***, The cross-validated log-likelihood of the model using the subject's best-fitting behavioral prior (green/blue) or the fixed neural prior (red), and the log-likelihood of the original model (Stocker and Simoncelli, 2006; gray). The log-likelihood value is normalized to the range defined by a "coin-flip" model (lower bound) and a Weibull fit to each psychometric curve (upper bound). Error bars represent ±SD across 100 validation runs according to a fivefold cross-validation procedure. For details, see Materials and Methods.

observer in terms of its body, head, and eye movements all contribute to the visual motion signal. Thus, the precise characterization of the visual speed distribution would require accurate measurements and calibrations of these different types of motions, as well as of the algorithm used to extract the motion information from the visual signal. Previous studies have approximated these relative movements to various degrees and used different algorithms to extract local visual speed from spatiotemporal images, resulting in different characterization of the prior distribution (Dong and Atick, 1995; Roth and Black, 2007; Baker et al., 2011; Sinha et al., 2021). However, common to all these measured stimulus priors is that they have higher probabilities at slow speeds and form long-tailed distributions. Future work using more comprehensive data (DuTell et al., 2020) may provide a better characterization of visual speed priors under ecologically valid, natural conditions.

We expect our model and analytic approach to be applicable to any other perceptual variable and task that exhibit characteristic patterns of perceptual biases and discrimination thresholds. However, of particular interest and posing a strong test of our model are changes in perceptual bias and threshold that are induced by spatiotemporal context such as adaptation aftereffects or the tilt illusion (Clifford et al., 2007; Schwartz et al., 2007, 2009). It is traditionally assumed that these biases are caused by a mismatch in expectation between encoding and decoding (i.e., the "coding catastrophe"; Schwartz et al., 2007), which is in sharp contrast to one of the main features of our model. Preliminary results are promising (Wei et al., 2015; Wei and Stocker, 2017). However, more quantitative analyses are necessary to test how well the framework can account for the data and what neural and behavioral priors it will predict.

In summary, within the context of visual speed perception, we have demonstrated that the Bayesian observer model constrained by efficient coding has the potential to provide a unifying framework that can quantitatively link natural scene statistics with psychophysical behavior and neural representation. Our results represent a rare example in cognitive science where behavioral and neural data quantitatively match within the predictions of a normative theory.

References

Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. Neural Comput 11:91–101.

Acharya T, Ray AK (2005) Image processing: principles and applications. New York: Wiley.

Albrecht DG, Hamilton DB (1982) Striate cortex of monkey and cat: contrast response function. J Neurophysiol 48:217–237.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366.

Baker S, Scharstein D, Lewis J, Roth S, Black MJ, Szeliski R (2011) A database and evaluation methodology for optical flow. Int J Comput Vis 92:1–31.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages (pp 217–234). In: Sensory communication (Rosenblith WA, ed). Cambridge, MA: MIT.

Blakemore MR, Snowden RJ (1999) The effect of contrast upon perceived speed: a general phenomenon? Perception 28:33–48.

Bowers JS, Davis CJ (2012) Bayesian just-so stories in psychology and neuroscience. Psychol Bull 138:389–414.

Britten K, Shadlen M, Newsome W, Movshon J (1993) Responses of neurons in macaque MT to stochastic motion signals. Vis Neurosci 10:1157–1169.

Brunel N, Nadal J (1998) Mutual information, Fisher information, and population coding. Neural Comput 10:1731–1757.

Burge J, Geisler WS (2015) Optimal speed estimation in natural image movies predicts human performance. Nat Commun 6:7900.

Champion RA, Warren PA (2017) Contrast effects on speed perception for linear and radial motion. Vision Res 140:66–72.

Cheyette SJ, Piantadosi ST (2020) A unified account of numerosity perception. Nat Hum Behav 4:1265–1272.

Clifford CW, Webster MA, Stanley GB, Stocker AA, Kohn A, Sharpee TO, Schwartz O (2007) Visual adaptation: neural, psychological and computational aspects. Vision Res 47:3125–3131.

De Bruyn B, Orban GA (1988) Human velocity and direction discrimination measured with random dot patterns. Vision Res 28:1323–1335.

Dehaene S, Mehler J (1992) Cross-linguistic regularities in the frequency of number words. Cognition 43:1–29.

Dong DW, Atick JJ (1995) Statistics of natural time-varying images. Network 6:345–358.

Dror RO, Willsky AS, Adelson EH (2004) Statistical characterization of real-world illumination. J Vis 4:821–837.

DuTell V, Gibaldi A, Focarelli G, Olshausen B, Banks M (2020) The spatiotemporal power spectrum of natural human vision. J Vis 20:1661.

Fechner GT (1860) Elemente der Psychophysik. Leipzig, Germany: Breitkopf und Härtel.

Fechner GT, Howes DH, Boring EG (1966) Elements of psychophysics. New York: Holt, Rinehart and Winston.

Ganguli D, Simoncelli E (2010) Implicit encoding of prior probabilities in optimal neural populations. Adv Neural Inf Process Syst 2010:658–666.

Ganguli D, Simoncelli EP (2014) Efficient sensory encoding and bayesian inference with heterogeneous neural populations. Neural Comput 26:2103–2134.

Ganguli D, Simoncelli EP (2016) Neural and perceptual signatures of efficient sensory coding. arXiv:1603.00058.

Gu Y, Fetsch CR, Adeyemo B, DeAngelis GC, Angelaki DE (2010) Decoding of MSTd population activity accounts for variations in the precision of heading perception. Neuron 66:596–609.

Hedges J, Stocker A, Simoncelli E (2011) Optimal inference explains the perceptual coherence of visual motion stimuli. J Vis 11(6):14, 1–16.

Heuer HW, Britten KH (2002) Contrast dependence of response normalization in area MT of the rhesus macaque. J Neurophysiol 88:3398–3408.

Horswill MS, Plooy AM (2008) Reducing contrast makes speeds in a video-based driving simulator harder to discriminate as well as making them appear slower. Perception 37:1269–1275.

Huang X, Lisberger SG (2009) Noise correlations in cortical area mt and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. J Neurophysiol 101:3012–3030.

Hürlimann F, Kiper D, Carandini M (2002) Testing the Bayesian model of perceived speed. Vision Res 42:2253–2257.

Jogan M, Stocker AA (2015) Signal integration in human visual speed perception. J Neurosci 35:9381–9390.

Jones M, Love BC (2011) Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. Behav Brain Sci 34:169.

Kanitscheider I, Coen-Cagli R, Kohn A, Pouget A (2015) Measuring Fisher information accurately in correlated neural populations. PLoS Comput Biol 11:e1004218.

Kohn A, Coen-Cagli R, Kanitscheider I, Pouget A (2016) Correlations and neuronal population information. Annu Rev Neurosci 39:237–256.

Krekelberg B, van Wezel R, Albright T (2006) Interactions between speed and contrast tuning in the middle temporal area: implications for the neural code for speed. J Neurosci 26:8988–8998.

Lakshminarasimhan KJ, Petsalis M, Park H, DeAngelis GC, Pitkow X, Angelaki DE (2018) A dynamic Bayesian observer model reveals origins of bias in visual path integration. Neuron 99:194–206.

Laughlin SB (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturforschung C Biosci 36:910–912.

Linsker R (1988) Self-organization in a perceptual network. Computer 21:105–117.

McDonnell MD, Stocks NG (2008) Maximally informative stimuli and tuning curves for sigmoidal rate-coding neurons and populations. Phys Rev Lett 101:058103.

McKee SP, Silverman GH, Nakayama K (1986) Precise velocity discrimination despite random variations in temporal frequency and contrast. Vision Res 26:609–619.

Morais M, Pillow JW (2018) Power-law efficient neural codes provide general link between perceptual bias and discriminability. Adv Neural Inf Process Syst 31:5071–5080.

Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A (2014) Information-limiting correlations. Nat Neurosci 17:1410–1417.

Movshon JA, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J Neurosci 16:7733–7741.
Newsome WT, Pare EB (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). J Neurosci 8:2201–2211.
Nieder A, Miller EK (2003) Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. Neuron 37:149–157.
Noel JP, Zhang LQ, Stocker AA, Angelaki DE (2021) Individuals with autism spectrum disorder have altered visual encoding capacity. PLoS Biol 19: e3001215.
Nover H, Anderson CH, DeAngelis GC (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. J Neurosci 25:10049–10060.
Pack C, Hunter J, Born R (2005) Contrast dependence of suppressive influences in cortical area MT of alert macaque. J Neurophysiol 93:1809–1815.
Panish SC (1988) Velocity discrimination at constant multiples of threshold contrast. Vision Res 28:193–201.
Peters MA, Balzer J, Shams L (2015) Smaller = denser, and the brain knows it: natural statistics of object density shape weight expectations. PLoS One 10:e0119794.
Piantadosi ST, Cantlon JF (2017) True numerical cognition in the wild. Psychol Sci 28:462–469.
Polanía R, Woodford M, Ruff CC (2019) Efficient coding of subjective value. Nat Neurosci 22:134–142.
Prat-Carrabin A, Woodford M (2021) Efficient coding of numbers explains decision bias and noise. bioRxiv. doi:10.1101/2020.02.18.942938.
Priebe N, Cassanello C, Lisberger S (2003) The neural representation of speed in macaque area MT/V5. J Neurosci 23:5650–5661.
Rast L, Drugowitsch J (2020) Adaptation properties allow identification of optimized neural codes. Adv Neural Inf Process Syst 33:1142–1152.
Rideaux R, Welchman AE (2020) But still it moves: static image statistics underlie how we see motion. J Neurosci 40:2538–2552.
Rokers B, Fulvio JM, Pillow JW, Cooper EA (2018) Systematic misperceptions of 3-D motion explained by Bayesian inference. J Vis 18(3):23, 1–23.
Roth S, Black MJ (2007) On the spatial statistics of optical flow. Int J Comput Vision 74:33–50.
Roy S, Jun NY, Davis EL, Pearson J, Field GD (2021) Inter-mosaic coordination of retinal receptive fields. Nature 592:409–413.
Schwartz O, Hsu A, Dayan P (2007) Space and time in visual context. Nat Rev Neurosci 8:522–535.
Schwartz O, Sejnowski TJ, Dayan P (2009) Perceptual organization in the tilt illusion. J Vis 9(4):19, 1–20.
Sclar G, Maunsell J, Lennie P (1990) Coding of image contrast in central visual pathways of the macaque monkey. Vision Res 30:1–10.
Senna I, Parise CV, Ernst MO (2015) Hearing in slow-motion: humans underestimate the speed of moving sounds. Sci Rep 5:14054.
Seriès P, Stocker AA, Simoncelli EP (2009) Is the homunculus "aware" of sensory adaptation? Neural Comput 21:3271–3304.
Simoncelli E (1993) Distributed analysis and representation of visual motion. PhD thesis, MIT.
Sims CR (2018) Efficient coding explains the universal law of generalization in human perception. Science 360:652–656.
Sinha SR, Bialek W, Van Steveninck RRDR (2021) Optimal local estimates of visual motion in a natural environment. Phys Rev Lett 126:018101.
Sotiropoulos G, Seitz AR, Series P (2014) Contrast dependency and prior expectations in human speed perception. Vision Res 97:16–23.
Stocker A (2006) Analog VLSI circuits for the perception of visual motion. Chichester, UK: Wiley.
Stocker A, Simoncelli E (2004) Constraining a Bayesian model of human visual speed perception. Adv Neural Inf Process Syst 2004:1361–1368.
Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9:578–585.
Stocker AA, Majaj N, Tailby C, Movshon JA, Simoncelli EP (2009) Decoding velocity from population responses in area MT of the macaque. J Vis 9:741.
Stone LS, Thompson P (1992) Human speed perception is contrast dependent. Vision Res 32:1535–1549.
Taylor R, Bays P (2018) Efficient coding in visual working memory accounts for stimulus-specific variations in recall. J Neurosci 38:7132–7142.
Thompson P (1982) Perceived rate of movement depends on contrast. Vision Res 22:377–380.
Treisman M (1964) Noise and Weber's law: the discrimination of brightness and other dimensions. Psychol Rev 71:314–330.
Turano K, Pantle A (1989) On the mechanism that encodes the movement of contrast variations: velocity discrimination. Vision Res 29:207–221.
Van Bergen RS, Ma WJ, Pratte MS, Jehee JF (2015) Sensory uncertainty decoded from visual cortex predicts behavior. Nat Neurosci 18:1728–1730.
Wang Z, Stocker A, Lee D (2012) Optimal neural tuning curves for arbitrary stimulus distributions: Discrimax, Infomax and minimum L_p loss. Adv Neural Inf Process Syst 3:2177–2185.
Wang Z, Stocker A, Lee D (2016) Efficient neural codes that minimize L_p reconstruction error. Neural Comput 28:2656–2686.
Wei XX, Stocker AA (2012) Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference. Adv Neural Inf Process Syst 25:1304–1312.
Wei XX, Stocker AA (2015) A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. Nat Neurosci 18:1509–1517.
Wei XX, Stocker AA (2016) Mutual information, Fisher information, and efficient coding. Neural Comput 28:305–326.
Wei XX, Stocker AA (2017) Lawful relation between perceptual bias and discriminability. Proc Natl Acad Sci U S A 114:10244–10249.
Wei XX, Ortega P, Stocker A (2015) Perceptual adaptation: getting ready for the future. J Vis 15:388.
Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598–604.
Welchman A, Lam J, Bülthoff H (2008) Bayesian motion estimation accounts for a surprising bias in 3D vision. Proc Natl Acad Sci U S A 105:12087–12092.
Yerxa TE, Kee E, DeWeese MR, Cooper EA (2020) Efficient sensory coding of multidimensional stimuli. PLoS Comput Biol 16:e1008146.
Zeki S (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol 236:549–573.
Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370:140–143.