Revision

Computational Perception and Cognition Laboratory

Authors: Ling-Qi Zhang and Alan A. Stocker

Affiliation: Department of Psychology
University of Pennsylvania

Correspondence: Dr. Alan A. Stocker
Computational Perception and Cognition Laboratory
421 Goddard Laboratories
3710 Hamilton Walk
Philadelphia, PA 19104
U.S.A.

astocker@psych.upenn.edu
Phone: +1 215 573 9341

Journal: bioRxiv

Classification: Biological Sciences (Psychology/Neuroscience)

Statistics: 34 Pages, 8 Figures, 1 Table, 2 Extended Data Figures

Prior expectations in visual speed perception predict encoding characteristics of neurons in area MT

Ling-Qi Zhang and Alan A. Stocker

Department of Psychology, University of Pennsylvania

Abstract

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 due to 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 model to psychophysical data ("behavioral prior"), to the point that the two 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.

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 [1–3]. These biases and perceptual distortions are qualitatively consistent with a Bayesian observer that combines noisy sensory measurements with a prior preference for lower speeds [4–6]. 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 [3, 7]. 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 [e.g. 8–11]. However, recovering the parameters of a Bayesian observer model from behavioral data is typically difficult due to the intrinsic non-specificity of its probabilistic formulation, which has been grounds for a critical view of the Bayesian modeling approach [12, 13]. The reverse-engineered speed priors in Stocker and Simoncelli [3] indeed showed large variations among individual subjects, indicating a potential case of over-fitting due to insufficient model constraints.

In this article, we show how we resolved this potential problem by developing and validating a tightly constrained Bayesian observer model. We followed a recent proposal that imposed efficient coding as a constraint to link both the likelihood function and the prior expectations of a Bayesian observer [14, 15]. 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 [16–18]. It thus establishes a direct relationship between the stimulus distribution and the accuracy of neural representations in sensory systems [19–23]. Wei and Stocker [15] showed how to formulate efficient coding as an information constraint that can be embedded within the probabilistic language of the Bayesian framework. The resulting, tightly constrained Bayesian observer model has proven to account for a wide range of phenomena in perception, including repulsive biases in perceived visual orientation [15, 24] and the lawful relationship between perceptual bias and discrimination threshold [25], but also in more cognitive domains such as in subjective preferences judgments [26] or the representation of numbers [27, 28].

The overall goal of our current work was two-fold. First, we aimed for a quantitative validation of this new Bayesian observer model, and in particular its efficient coding constraint, in the domain of visual speed perception. We fit the model to speed discrimination data collected in the original study by Stocker and Simoncelli [3]. The data contain human psychophysical measurements

over a broad range of speeds and stimulus contrasts. We found that the new model allowed us to reverse-engineer much more reliable and consistent estimates of individual subjects' prior beliefs than the old model in the original study, while it still accurately accounted for subjects' overall psychophysical behavior. Second, based on the efficient coding hypothesis we wanted to test whether the reverse-engineered prior expectations can correctly predict the population encoding characteristics of neurons in the motion sensitive area in the primate brain. Neurons in the medial temporal (MT) area in the macaque monkey are widely recognized as directly reflecting the perceived direction and speed of visual motion of the animal [29–32]. By analyzing single cell recordings of a large population of MT neurons [33], 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 represent an important validation of the Bayesian observer constrained by efficient coding model and offer a normative framework for linking psychophysical to neurophysiological data.

Results

We model speed perception as an encoding encoding, Bayesian decoding process (Fig. 1A). On any given trial the speed v of a visual stimulus is represented by a noisy and bandwidth-limited sensory measurements m. Following Wei and Stocker [14, 15] 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 [20, 34, 35], that is

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

This constraint proposes 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 $p(m|v)$ is combined with the prior $p(v)$, resulting in the posterior $p(v|m)$. Lastly, a percept (i.e., estimate) $\hat{v}$ is computed based on the posterior and a loss function.

An important aspect of the new model is that the prior distribution jointly determines both encoding and decoding of the Bayesian observer model [14]. Thus, both the encoding characteristics of neurons representing visual speed (Fig. 1B), and the psychophysical behavior of subjects in speed perception (Fig. 1C) should be consistent with the belief of the observer about the statistical regularities of visual speed in the environment.

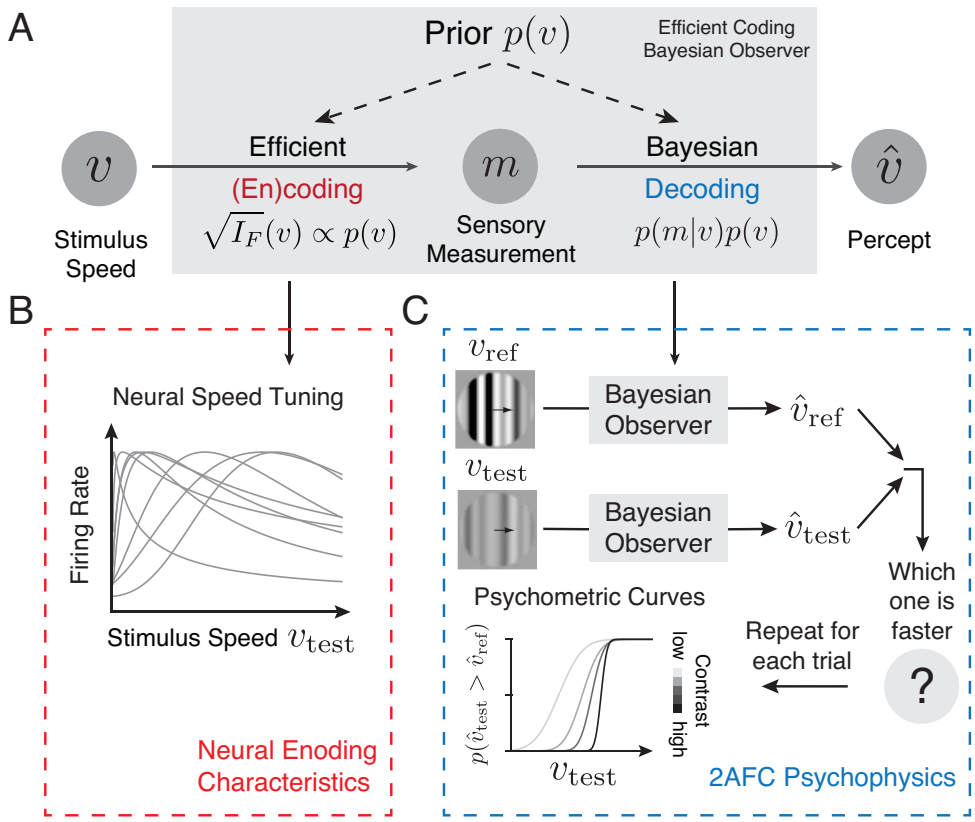


Figure 1: Bayesian observer model constrained by efficient coding. **A)** We model speed perception as an efficient encoding, Bayesian decoding process [14, 15]. Stimulus speed v is encoded by a noisy and resource-limited sensory measurement m with an encoding accuracy that is determined by the stimulus prior $p(v)$ according to an 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. Importantly, both the encoding and decoding stages are linked and jointly determined by the prior distribution over speed. **B)** Efficient coding makes predictions regarding the accuracy of neural representation of visual speed (i.e., the tuning characteristics of neurons in area MT). **C)** Embedding the Bayesian observer within a two-alternative forced choice (2AFC) decision process provides a model to predict psychophysical behavior of human subjects in a speed discrimination task.

Extracting the "behavioral prior"

We fit our model to the psychophysical speed discrimination data collected by Stocker and Simoncelli [3]. On each trial of their experiment, subjects were shown a pair of horizontally drifting gratings, consisted of a reference and a test stimulus, and were asked to choose which one was moving faster (Fig. 1C). For each combination of reference and test, 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 (see *Methods* and Stocker and Simoncelli [3] for details).

In contrast to the original model [3], the new observer model directly links the likelihood function and the prior distribution [14]. Thus perceived speed is fully determined by subjects' prior expectations and a contrast-dependent internal noise parameter $h(c)$ that determines the total amount of represented sensory information [15, 35]. Specifically, we assume that a stimulus v is encoded with noisy sensory measurement m according to the conditional distribution

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

where $F(v) = \int_{-\infty}^{v} p(v)dv$ is the cumulative density function (CDF) of v. The corresponding likelihood $p(m|v)$ is then combined with the same speed prior density $p(v)$ for the decoding of the stimulus (Fig. 1A). Our goal was to find the prior distribution $p(v)$ and the noise parameters $h(c)$ that best predicted individual subjects' perceptual behavior. In order to fit the observer model, we embedded it within a binary decision process (Fig. 1C). On each trial, speed estimates for both the reference and the test stimuli are performed, and then subjects are assumed to respond according to which estimate is faster. The full psychometric functions are predicted by marginalizing over the (unobserved) sensory measurement (see *Methods* for details).

We jointly fit our model, embedded in the decision process, to all 72 conditions (i.e., psychometric functions) for each individual subject 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. Based on previous studies [3, 9, 11], we parameterized the prior distribution as a modified power-law function

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

where c_i's were free and unconstrained parameters. Figure 2A shows the data and model fit for a few example conditions for exemplary Subject 1. Fit model parameters for all subjects are listed in Table 1. Overall, the model predicts psychometric curves that are similar to those from fitting a Weibull function. The log-likelihood of the new model is close to that of separate Weibull fits to each individual conditions (Fig. 2B). Figure 2C further illustrates that the new model performs as well

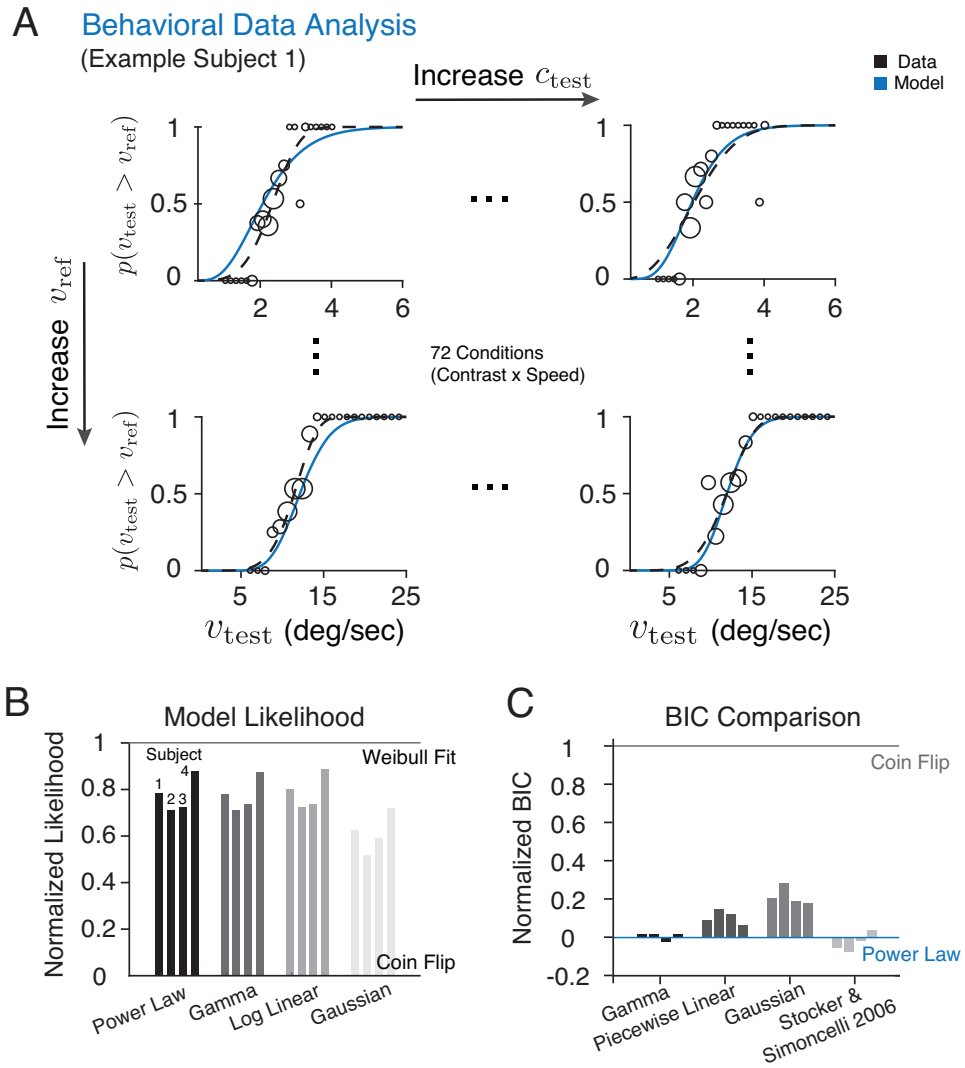


Figure 2: Extracting the "behavioral prior". **A)** We jointly fit the Bayesian observer model to psychophysical data across all contrast and reference speed conditions (72 conditions total). Shown are a few conditions for exemplary Subject 1. Circle sizes are proportional to the number of trials at that test speed. The dashed curves are Weibull fits to each conditions, and the solid blue curves represent the model prediction. See Extended Data Figure 2 - 1 and 2 - 2 for the complete set of psychometric curves and model fits to all 72 conditions. **B)** The log-likelihood values, normalized to the range set by a coin flip model (lower bound) and Weibull fits to individual psychometric curves (upper bound), of the best-fitting model for individual subjects based on four different prior parameterizations including a power-law function, Gamma distribution, piece-wise log-linear function, and a Gaussian distribution, respectively (see *Methods*). **C)** The relative BIC values for the different parameterizations as well as the original, less constrained Bayesian observer model by Stocker and Simoncelli [3]. Values are normalized to the range set by the efficient coding Bayesian model with power-law parameterization and the coin flip model (lower is better).

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

Table 1: Parameters of the fit Bayesian observer model. The best-fit parameters of the prior density $p(v; c_0, c_1, c_2) \propto (|v| + c_1)^{c_0} + c_2$ and the contrast-dependent noise parameters $\sigma = h(c)$ for each individual subject.

as the original, less constrained Bayesian observer model [3] based on the computed Bayesian Information Criterion (BIC). A more detailed model comparison based on cross-validation is provided in a later section. Importantly, the reverse-engineered prior expectations are much more consistent across individual subjects compared to the previous model [3, 11]: The exponent c_0, for example, is now close to a value of -1 for every subject. Furthermore, the contrast-dependent noise parameters 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).

In order to test the impact of our specific choice of using a power-law parameterization for the prior distribution (Eq. 3), we preformed model fits using two other parameterizations with increasing degrees of freedom (i.e., Gamma distribution and piece-wise log-linear function), and a simple Gaussian prior for comparison (see *Methods*). The model fits well for all but the Gaussian prior, resulting in similar log-likelihood values (Fig. 2B), although the BIC value is higher for the log-linear parameterization due to its large number of parameters (Fig. 2C). Crucially, however, the shapes of the resulting 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, which produced a poor model fit because it is unsuited to approximate a power-law function.

We further validated our model by comparing its predictions for contrast-induced biases and discrimination thresholds to subjects' behavior. To quantify the effect of contrast-induced 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 actual physical speed will need to be higher for the subject to match it to the higher-contrast reference. Thus, a contrast-induced slow-speed bias is manifested by a PSE ratio greater than 1 when the test contrast is lower than the reference, and vice versa. As shown in Fig. 5A, subjects clearly under-estimated the speeds of low-contrast stimuli, an effect that occurred at any contrast level and any speeds. Our new model is able to capture both

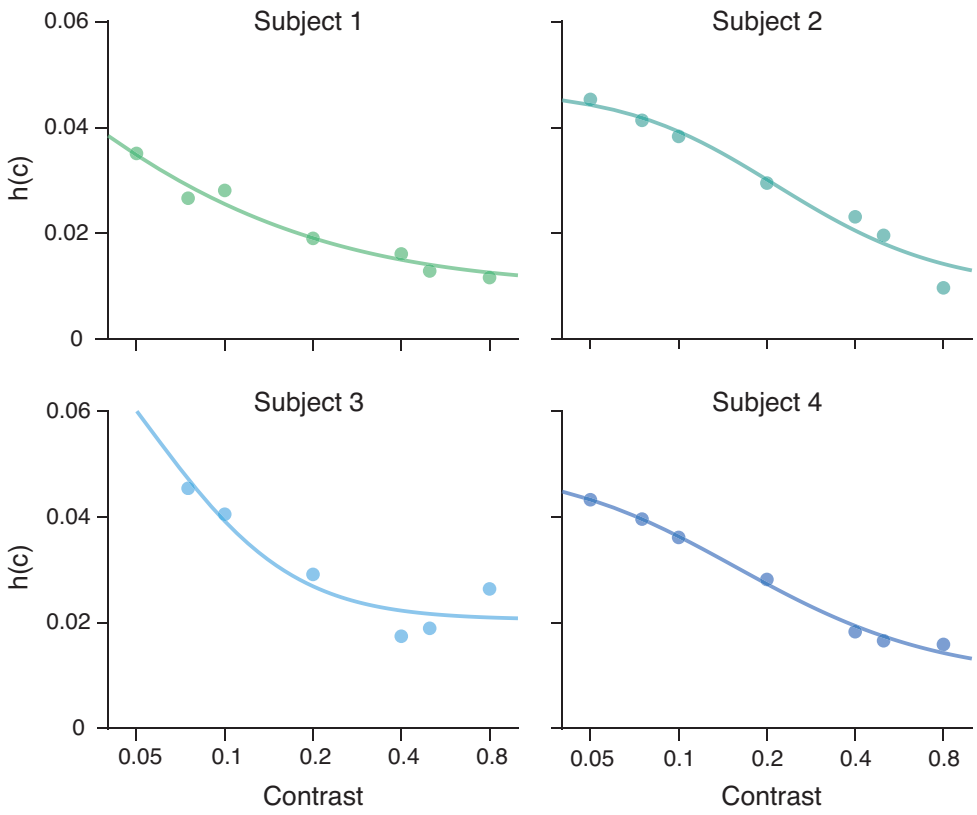


Figure 3: Contrast-dependent noise. Fit noise parameters as a function of contrast for all individual subjects. Bold lines represent parametric fits of the form $h(c) = [r_{\text{max}} c^q / (c^q + c_{50}^q) + r_{\text{base}}]^{-1/2}$ based on the contrast response function of cortical neurons [36–38].

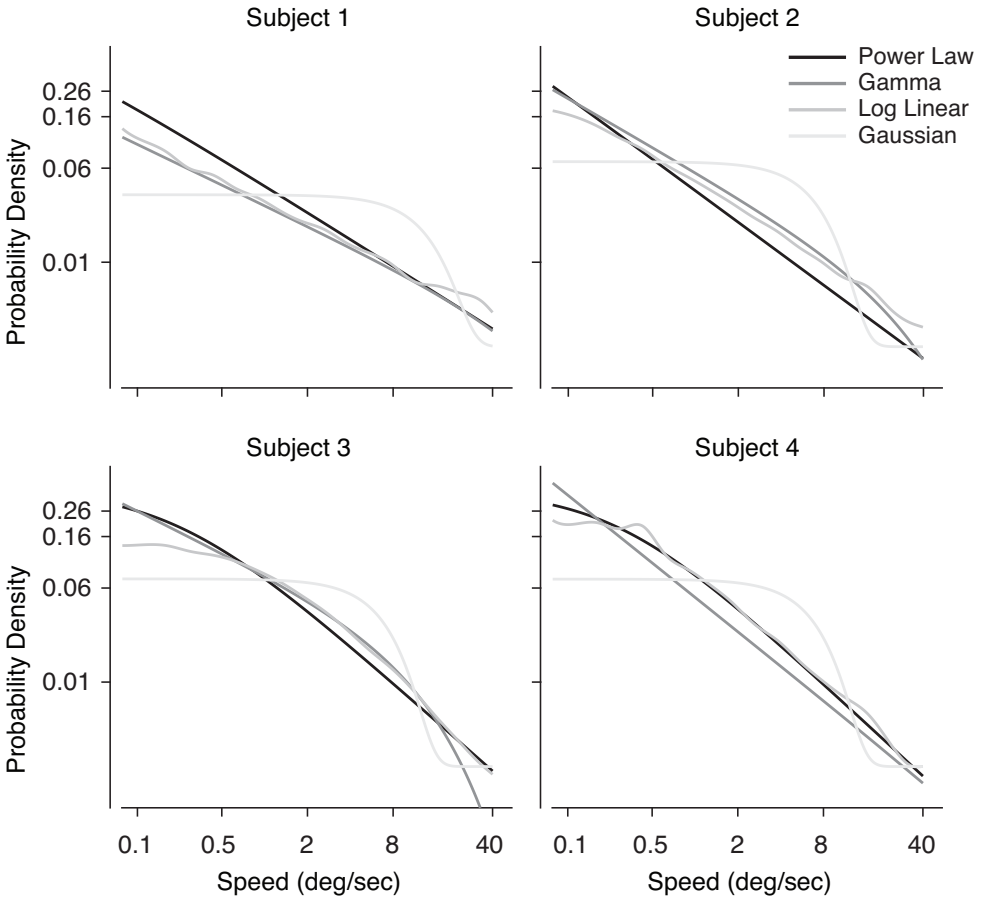


Figure 4: Effect of prior parameterization. The best-fit prior density function for each individual subject, based on four different parameterizations including a power-law function, Gamma distribution, piece-wise log-linear function, and a Gaussian distribution (which fits poorly to the data; see Fig. 2B), respectively.

the discrimination threshold behavior and the contrast-induced slow-speed biases as indicated by the fits, comparable to the original model [3]. Subjects' discrimination thresholds were measured as the minimal stimulus speed difference that can be reliably perceived, which we defined as the difference between the 50% and 75% points of the psychometric curve. As shown in Fig. 5B, all subjects' thresholds increase monotonically with speed. While they follow Weber's law at higher speeds, they deviate from a constant Weber-fraction at slow speeds, which is well documented [3, 39, 40]. Although the model predicts the overall shape of the threshold curves, the predicted higher threshold values for the low-contrast stimulus condition are not particularly evident in the data (Fig. 5B). However, previous studies have convincingly demonstrated that lower stimulus contrasts lead to higher speed discrimination thresholds under various stimulus configurations [41–43]. Thus we believe that the experimental design in Stocker and Simoncelli [3], in particular the deliberate compromise in choosing a low number of trials per condition (only 80 trials per psychometric curve) in order to test individual subjects over a large range of different contrast/speed conditions, is responsible for the non-discriminating, noisy threshold estimates. Estimation accuracy was further limited by the use of a stair-case procedure optimized for inferring the PSE rather than the slope of the psychometric curves (see Stocker and Simoncelli [3] for details). 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 contrast-dependent noise parameters explain the individual variations in threshold magnitude and bias. In addition, when the prior exponent is close to -1, the PSE ratios are mostly constant across different speed (e.g., Subject 3 and 4) [25], whereas a deviation from it predicts relative bias that decrease for higher stimulus speeds (e.g., Subject 1).

Lastly, previous models have mainly focused on the contrast-induced speed biases and how they can be attributed to a slow-speed prior that shifts the percept toward slower speeds for increasing levels of uncertainty caused by sensory noise or the ambiguity of the stimulus itself [3, 5, 6, 9, 44–46]. 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. We can analytically show that under efficient coding, the modified power-law prior (Eq. (3) with an exponent $c_0 = -1$) predicts both the constant Weber fraction at higher speed and its deviation at slow speeds, thus closely resembling human behavior (see *Methods*). In fact, predicted Weber fractions based on the reverse-engineered prior for individual subjects closely resemble previous psychophysical measurements (Fig. 6).

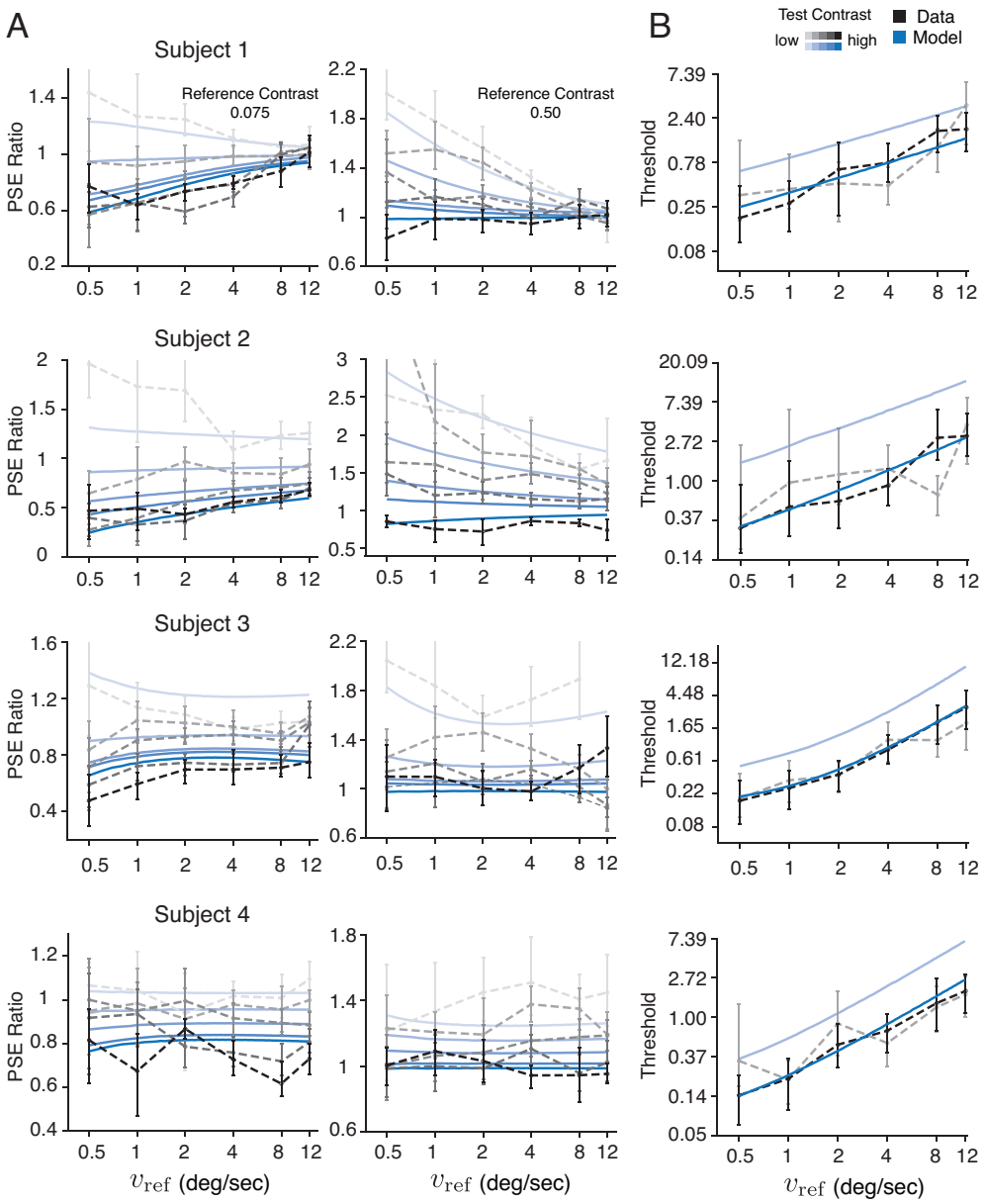


Figure 5: Predicted contrast-dependent bias and discrimination threshold. **A)** Ratio of the test relative to reference speeds at the point of subjective equality (PSE) extracted from the Weibull fits (black) and our model (blue). Shading levels correspond to different contrast levels (0.05, 0.1, 0.2, 0.4, 0.8) of the test stimulus (darker means higher contrast). The reference stimulus has a contrast of 0.075 in 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 level (0.075, 0.5), extracted from the Weibull fits (black) and our model (blue). Error bars indicate the 95% confidence interval across 500 bootstrap runs.

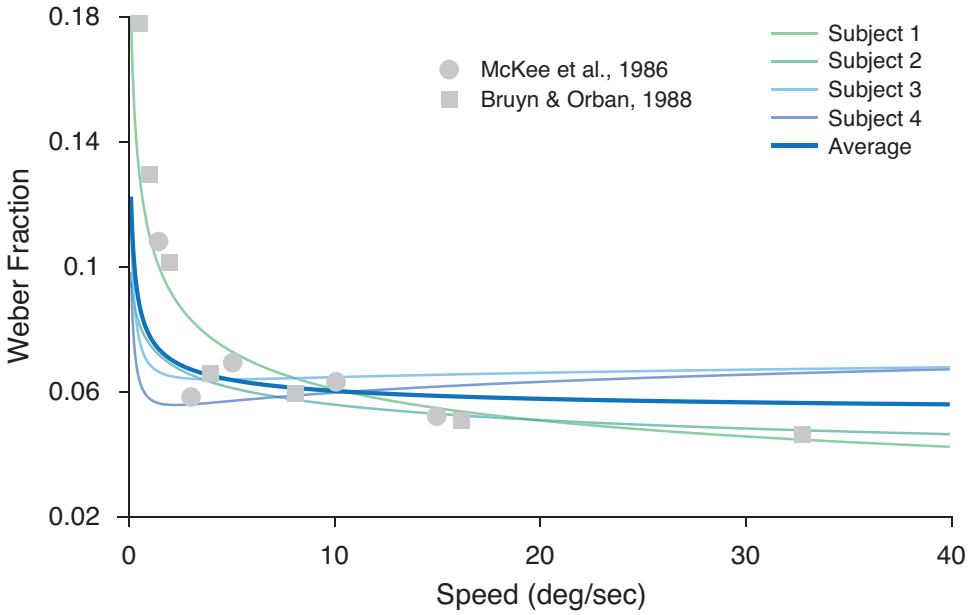


Figure 6: Predicted Weber fraction. Predicted Weber fractions $\Delta v/v$ based on the reverse-engineered behavioral priors of the four individual and the average subjects are shown in comparison to previously reported psychophysical measurements [39, 40]. Since the Weber fraction is predicted up to a constant, a scalar parameter was chosen such that the predictions are on a similar scale as the data. Note that De Bruyn and Orban [40] also found deviations from Weber's law at extremely high speeds (256 cycles/deg) which is not depicted here.

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 [14, 22, 35]. 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, widely considered the cortical area in the primate brain that represents local visual motion [29, 47]. Our analysis was based on single cell recording data from the macaque brain [33]. The data contained repeated spike counts from 480 MT neurons responding to random dot motion stimuli moving at eight different speeds (see *Methods* and Nover et al. [33] for more details). Following the original study [33], we fit a log-normal speed tuning curve model for each neuron in the data set (Fig. 7A). This tuning curve model well described the majority of the neurons (Fig. 7B). Under the assumption that neural response variability was well captured by a Poisson distribution, it is then straight-forward to compute the FI of individual neurons (Fig. 7A - see *Methods*).

Given the inherent limitations of the single-unit recorded data, assumptions about the noise and its correlation structure across the population are necessary in order to compute the population FI [48–51]. 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. 7C). 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 [51, 52], produce nearly identical estimation of the shape (i.e., normalized FI, see *Methods* Eq. (18)) of the population FI (Fig. 7D). 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 to the independent noise assumption. However, noise correlations between pairs of neurons that depend on their relative speed tuning preferences have been reported in area MT [53]. Thus, to assess the potential impact of such tuning-dependent correlations, we computed the linear population FI with a limited-range correlation model based on the relative speed preferences of individual neurons [48] (see *Methods*). We found that although again 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. 7E). The reason why these correlations seem to 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 tuning width, are mostly independent of speed preference), and are close

to uniformly tiling the logarithmic speed space, at least based on the current data set [33]. Thus, we argue that given the available evidence, estimating the shape of FI based on the independent noise assumption 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 [15, 35, 54]. Although the fit contrast-dependent noise parameters for all subjects 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, something the current data set does not provide. Limited (unpublished) neural data [55] suggests, however, that stimulus contrast indeed simply scales the population FI without changing its shape. This is intriguing given the well-documented diverse and heterogeneous effects contrast has on the shape and location of the speed tuning curves of individual neurons in area MT [55–57].

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 for which neural encoding is optimized. Figure 8A shows the extracted behavioral priors (for individual subjects and their average) 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.

In order 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 and only the contrast-dependent noise parameters $h(c)$ were free. We then used a cross-validation procedure to compare this neural observer with the unconstrained observer and the original Bayesian observer [3]. Cross-validation accounted for the difference in complexity between the models by estimating their predictive power on held-out data. As illustrated in Fig. 8B, we found that 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 prior 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 thus does not lead to over-fitting.

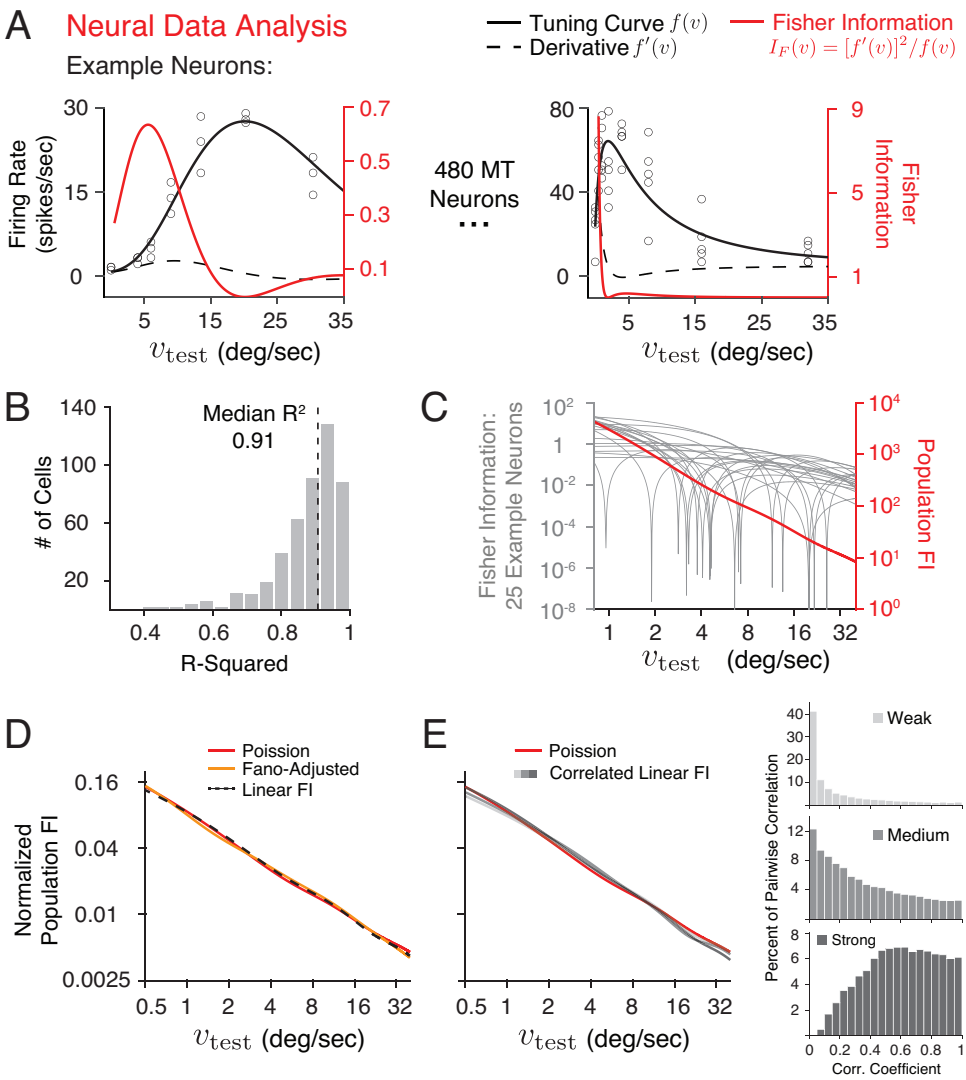


Figure 7: Extracting the "neural prior". According to the efficient coding constraint (Eq. (1)) the population Fisher information (FI) should directly reflect the prior distribution of visual speed. **A)** Two example MT neurons from the data set [33] with fit log-normal tuning curves (black, left y-axis) and their corresponding FI (red, right y-axis) assuming a Poisson noise model. **B)** A histogram of the goodness-of-fit, measured by R^2, of the log-normal tuning curve fit to individual neurons. **C)** Individual FI of 25 example neurons (gray, left y-axis), and the population FI (red, right y-axis), calculated as the direct sum of the FI over all 480 neurons in the data set. **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 limited-range correlation model [48] for three levels of correlation strength as illustrated by the histogram of pairwise correlation coefficients on the right. See *Methods* for more details.

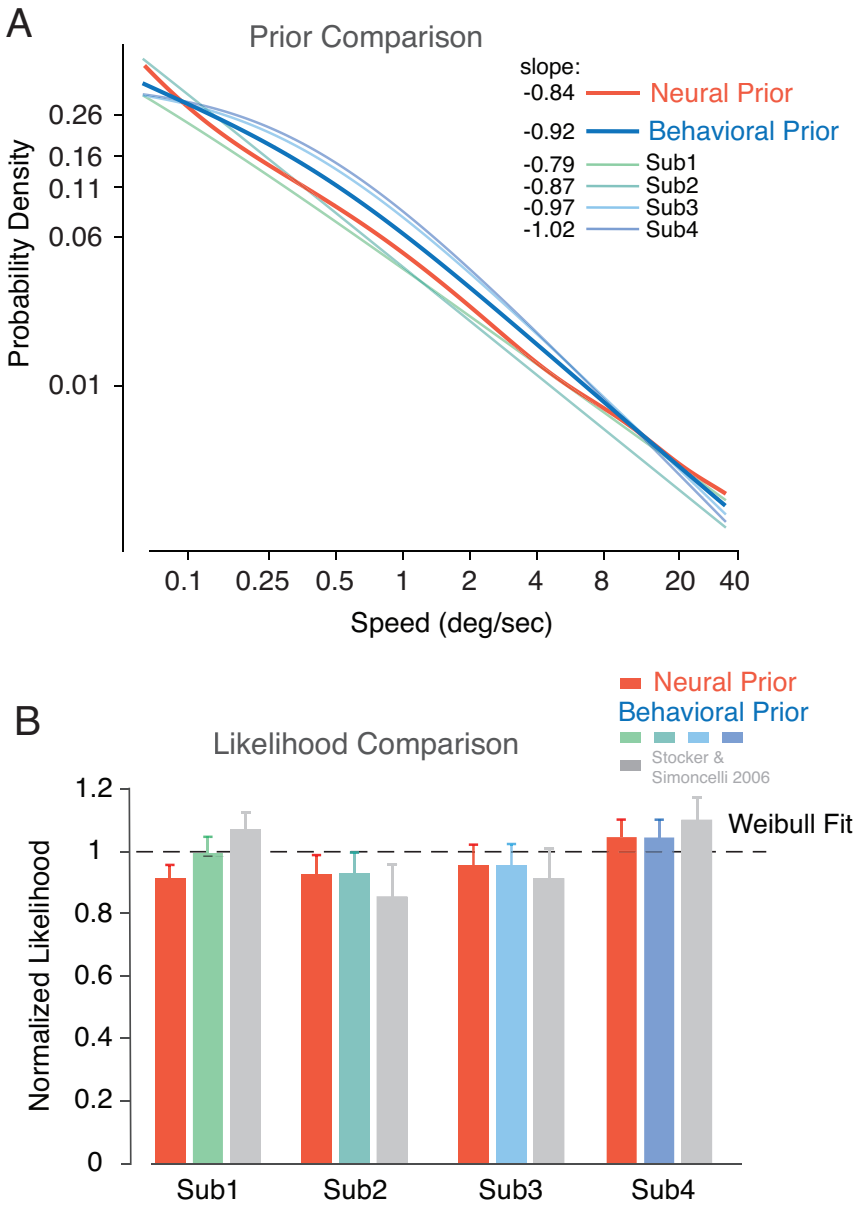


Figure 8: Comparing neural and behavioral prior. **A)** Reverse-engineered behavioral priors for individual subjects (light shades) and their average (dark blue) superimposed by the neural prior (dark red). Slope values are computed from a linear regression on the curves in log-log coordinates. **B)** The cross-validated log-likelihood of the model using the subject's best-fitting behavioral prior (blue) or the fixed neural prior (red), and the log-likelihood of the original model [3] (gray). The log-likelihood value is normalized to the range defined by a "coin-flip" model (lower bound) and the Weibull fit to each psychometric curve (upper bound). Error bars represent $\pm$ SD across 100 validation runs according to a five-fold cross-validation procedure. See *Methods* for details.

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 [58]. Neural encoding of visual speed in area MT is indeed considered logarithmic [33, 59]: When analyzed in the logarithmic speed domain, the tuning curves of MT neurons are approximately bell-shaped, scale-invariant, and tile the stimulus space with near uniform density [33, 56]. With our new model, we have demonstrated that a modified power-law prior (Eq. (3)) with an exponent of approximately -1 can well account for Weber's law behavior and the deviation from it at slow speeds. It is straightforward to show that for an exponent of exactly -1, this power-law prior also predicts the logarithmic neural encoding as specified above. Specifically, one implementation of 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 cumulative distribution function (CDF) [14, 60, 61], which is sometimes also referred to as histogram equalization [62]. With an exponent of -1, the CDF of the speed prior is exactly the logarithmic function that well-described MT tuning characteristics [33] (see *Methods* for details). 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 taken over a wide range of stimulus contrasts and speeds, which allowed us to reverse-engineer the "behavior prior" that best accounts for the psychophysical behavior of individual subjects. In addition, we analyzed the population encoding 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 expectations follow 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 to the original, more flexible Bayesian model formulation [3, 11], the added efficient coding constraint results in estimates of behavioral priors that are not only much more consistent across individual subjects but 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 [3, 5, 7, 8, 10, 11, 63] (but see [64]) by the quantitative support from electrophysiological neural data.

We also offer an explanation for when perceptual variables have a logarithmic neural representation and thus follow Weber's law [58]. According to our model, logarithmic encoding and Weber's law both directly 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 other alternative encoding solutions satisfying the efficient coding constraint are possible, see [e.g. 15]). Indeed, perceptual variables that are known to approximately follow Weber's law such as light intensity [65], numerosity [27, 28, 66], and weight [67], exhibit heavy tails in their statistical distributions under natural environmental conditions [68–71], a defining feature of a power-law function. Conversely, any deviation from Weber's law (i.e., from a constant Weber-fraction; Fig. 6) 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 [72–74]. Our approach is a substantial step forward in that it embeds this link within a full behavioral observer model. Thus rather than relying on single summary metrics of behavior (i.e. discrimination threshold or bias [54]), the predictions based on our model are constrained by the full richness of the psychophysical data (i.e., every single data point in the data 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 prior is limited to the extent that there were substantial experimental differences between the behavioral and neural data. For example, we compared human with non-human 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 suggest that they are reflecting the true stimulus prior, in

which case these differences in stimuli and model systems should indeed matter little, as the stimulus prior is a property of the environment and not the observer nor the particular stimulus pattern. Nonetheless, 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 [75]. Future work may exploit such voxel-level encoding models for human visual speed perception to validate and likely refine our estimates of the "neural prior", which would also permit an investigation of individual differences at both behavioral and neural levels through matched task and stimulus design.

The specific shape of the extracted neural and behavioral prior depends on our assumption of the objective function for efficiently representing visual speed. The efficient coding constraint (Eq. (1)) results from an attempt to maximize the mutual information between neural representation and stimulus [35]. It is possible, although unlikely given the exceptional good quantitative match, that with a different efficient coding constraint in combination with an appropriate loss function [21, 76, 77], a power-law prior with a different exponent could also be consistent with both the behavioral and neural data. 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 for visual stimulus variables such as luminance contrast and orientation for which good estimates of the stimulus prior exist, encoding characteristics in early visual cortex are closely accounted for by the mutual information maximization objective [21].

An important assumption of our observer model is that not only the neural and behavioral prior should match, but that they 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 should approximate a similar power-law distribution within the range given by the neural and behavioral priors illustrated in Fig. 8A. However, an empirical validation of this prediction by measuring the stimulus prior distribution is rather challenging. Object motion, but also the ego-motion of the 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 requires us to accurately measure and calibrate these different types of motions, as well as the algorithms used to extract the motion information. Previous studies have approximated these relative movements to various degrees and used different algorithms to extract local visual speed from spatio-temporal images, resulting in slightly different characterization of the prior distribution [78–81]. Common to all these measured stimulus priors, however, is that they have long-tail distributions with higher probabilities associated with slow speeds. Future work based on more comprehensive data [82] may provide more accurate characterizations of stimulus priors under ecologically valid, natural conditions.

We expect our model and analytic approach to be applicable to other perceptual variables and tasks that exhibit characteristic patterns of perceptual biases and discrimination thresholds, such as the estimation of local orientation [e.g. 54, 83], surface tilt [84], or heading direction [85]. Of particular interest, and posing a strong test of our model, are changes in perceptual bias and threshold that are induced by spatio-temporal context such as adaptation aftereffects or the tilt-illusion [86–88]. It is traditionally assumed that these biases are caused by a mismatch in expectation between the encoding and the decoding (i.e. the "coding catastrophe" [86]), which is in sharp contrast to the main assumption of our model. Preliminary results are promising [25, 89]. However, more quantitative analyses are necessary to test how well the framework can account for the data and what neural and behavioral prior 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.

Acknowledgments

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

Methods

Behavioral prior: Bayesian observer model constrained by efficient coding

Data

We reanalyzed the two-alternative-forced-choice (2AFC) speed discrimination data from Stocker and Simoncelli [3]. 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] deg/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 5,760 trials. We excluded one subject (labeled as Subject 3 in the original data set) that was only tested at two contrasts and two test speed levels. While we were actually able to recover a prior from this subject that was highly consistent with the rest of the subjects, we were not able to perform model comparison and cross-validation due to the low number of trials.

Model formulation

We used the Bayesian observer model constrained by efficient coding by Wei and Stocker [15] and embedded it within a decision process to predict the binary judgments in the 2AFC experiment [3]. For the main analysis, the speed prior distribution was parameterized as a modified power-law function (Eq. (3)). However, we also tested various other parameterizations (Fig. 4).

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:

$$p(m|v) = \mathcal{N}(m; \mu = F(v), \sigma^2 = h^2(c)) \qquad \text{(same as Eq. (2))}$$

where $F(v) = \int_{-\infty}^{v} p(v)dv$ is the cumulative density function (CDF) of v. 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

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

which is proportional to $p(v)$. The total amount of encoding resource is measured by $\int \sqrt{I_F(v)}dv$ which evaluates to $1/h(c)$. In order to handle the unbounded nature of a magnitude variable such as speed (as compare to circular variables), we added a small constant ($2.5 * 10^{-3}$) to $p(v)$ such that its CDF does 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

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

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

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

With an L_0 loss function [3], our final estimate $\hat{v}$ of the stimulus v is given as

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

It defines the optimally decoded stimulus $\hat{v}$ given m. The estimate is a deterministic function of a particular m (implicit in the likelihood function $l(v)$), thus we can explicitly express the dependency as $\hat{v}(m)$. However, in a psychophysical experiment, we cannot observe m directly. This requires us to marginalize over m, thus

$$p(\hat{v}|v) = \int \hat{v}(m)p(m|v)dm \; . \tag{8}$$

Furthermore, in the 2AFC speed discrimination experiment, subjects report a binary decision and not the estimate $\hat{v}$. We assume subjects make binary decisions (i.e., which one is faster) by internally comparing their estimate $\hat{v}_r$ of the reference stimulus v_r with their estimate $\hat{v}_t$ of the test stimulus v_t. Across many repeated trials, subject responses can be modeled as binomial distribution, with the probability of test stimuli being faster $p(v_t > v_r)$ as

$$p(v_t > v_r) = \int_{-\infty}^{+\infty} p(\hat{v_t}|v_t) \int_{-\infty}^{\hat{v_t}} p(\hat{v_r}|v_r)dv_r dv_t \; . \tag{9}$$

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

$$\mathcal{L} = \sum_{i=1}^{N} \{k_i \log[p(v_{it} > v_{ir})] + (1 - k_i) \log[1 - p(v_{it} > v_{ir})]\} \; . \tag{10}$$

We find the best-fitting model parameter c_0, c_1, c_2 and $h(c)$ by maximizing $\mathcal{L}$ using MATLAB's *fminsearchbnd* algorithm. Note that the model is highly constrained: For each subject, we are fitting a single three-parameter prior distribution plus one scalar noise parameter $h(c)$ for each of the 7 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 two alternative parameterizations as follows:

- 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] deg/s. Each corresponding $p(v^*_{1:18})$ value is a free prior parameter; prior density values are linearly interpolated between those values.

As a baseline 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 discrimination threshold (Δ_v) is inversely proportional to the square-root of FI [25, 90], thus

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

According to our 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 specific case of the modified power-law prior with an exponent of -1 (Eq. (3) we find

$$\Delta_v \propto (v + c_0) \ . \tag{13}$$

With $c_0 = 0$ (i.e., a strict power-law with exponent -1) we obtain $v \propto \Delta_v$, which is the definition of Weber's law (i.e., a constant Weber fraction). For non-zero c_0 the Weber fraction changes to

$$\frac{\Delta_v}{v} \propto (1 + \frac{c_0}{v}) \ . \tag{14}$$

At high speeds, $\frac{c_0}{v} \approx 0$ and thus the Weber fraction is constant. At low speeds, $\frac{c_0}{v} \to \infty$ causing $\frac{\Delta_v}{v}$ to increase. Both effects are characteristic features of human speed discrimination behavior [3, 39, 40] (Fig. 6).

Our efficient coding constraint implies that the stimulus is transformed according to the CDF of $p(v)$ (see Eq. (2)). For a power-law prior with exponent -1 its CDF is

$$\int z_1(v + c_0)^{-1} dv = z_1 \log(v + c_0) + z_2 \ , \tag{15}$$

which is the logarithmic transformation that has been previously used for describing the speed tuning of MT neurons [33].

Neural prior: MT encoding analysis

Data

We reanalyzed the electrophysiological recording data from Nover et al. [33]. 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] deg/s. Stimulus location, direction, size and disparity were individually optimized for each neuron. Every stimulus speed was presented three to seven times. We considered a neuron's single-trial response as the mean firing rate over the entire stimulus duration (1.5 s). We analyzed a total of 480 neurons.

Population Fisher information

Following Nover et al. [33], each neuron's firing rates as a function of stimulus speed were fit with a Gaussian tuning curve in log-speed

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

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 fitting procedure assuming a Poisson model for firing rate variability produced very similar results.

We computed population FI for different assumptions about neural response variability and its pair-wise noise correlation 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

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

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

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

As in Nover et al. [33], 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, obtained by linear regression of the neuron's firing rate

variance against firing rate mean, thus

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

Finally, we computed the linear Fisher information [51, 52]

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

where Σ is the noise correlation matrix and is the identity matrix for the independent noise case. To understand the effect of tuning preference-dependent noise correlations observed in area MT [53], we adopted the limited-range correlation model

$$\begin{aligned} \Sigma_{ij} &= \sigma^2 \rho_{ij} \\ \rho_{ij} &= \exp(-|\Delta_{ij}|/L) \ , \end{aligned} \tag{21}$$

where Δ_{ij} is difference in log-speed preference between neuron i and neuron j, σ^2 is the noise variance, and L is the overall correlation strength [48]. For the simulations shown in Fig. 7E, 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 five-fold cross-validation procedure: The trial data for each condition were first randomly and equally divided into 5 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 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 best fit prior for that run. For the neural prior condition, we assumed the prior to be 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 [3], and of individual Weibull fits for every condition. Log-likelihood values in Fig. 8B were normalized to the range set by an upper bound given by the values of the Weibull fits, and a lower bound represented by the log-likelihoods of a coin-flip model for the decision (i.e. a model with a fixed decision probability of 0.5).

Code and Data Availability

Data and analysis code are available through GitHub:
https://github.com/cpc-lab-stocker/Speed_Prior_2021

References

[1] Peter Thompson. Perceived rate of movement depends on contrast. *Vision research*, 22(3): 377–380, 1982.

[2] Leland S Stone and Peter Thompson. Human speed perception is contrast dependent. *Vision research*, 32(8):1535–1549, 1992.

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

[4] Eero Simoncelli. *Distributed analysis and representation of visual motion*. PhD thesis, MIT, Dept. of Electrical Engineering, Cambridge, MA, 1993.

[5] Yair Weiss, Eero P Simoncelli, and Edward H Adelson. Motion illusions as optimal percepts. *Nature neuroscience*, 5(6):598–604, 2002.

[6] A.A. Stocker. *Analog VLSI Circuits for the Perception of Visual Motion*. John Wiley & Sons Ltd., Chichester, May 2006. 242 pages. Hardcover.

[7] A.A. Stocker and E.P. Simoncelli. Constraining a Bayesian model of human visual speed perception. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems NIPS 17*, pages 1361–1368, Cambridge, MA, December 2004. MIT Press.

[8] A.E. Welchman, J.M. Lam, and H.H. Bülthoff. Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceeding of the National Academy of Sciences of the U.S.A.*, 105(33):12087–12092, August 2008.

[9] J.H. Hedges, A.A. Stocker, and E.P. Simoncelli. Optimal inference explains the perceptual coherence of visual motion stimuli. *Journal of Vision*, 11(6):1–16, May 2011.

[10] Grigorios Sotiropoulos, Aaron R. Seitz, and Peggy Series. Contrast dependency and prior expectations in human speed perception. *Vision Research*, 97:16 – 23, 2014. doi: 10.1016/j.visres.2014.01.012.

[11] Matjaž Jogan and Alan A Stocker. Signal integration in human visual speed perception. *Journal of Neuroscience*, 35(25):9381–9390, 2015.

[12] Matt Jones and Bradley C Love. Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and brain sciences*, 34(4):169, 2011.

[13] Jeffrey S Bowers and Colin J Davis. Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, 138(3):389, 2012.

[14] Xue-Xin Wei and Alan A Stocker. Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference. *Advances in neural information processing systems*, 25:1304–1312, 2012.

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

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

[17] Simon B Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforschung*, 36(910-912):51, 1981.

[18] Suva Roy, Na Young Jun, Emily L Davis, John Pearson, and Greg D Field. Inter-mosaic coordination of retinal receptive fields. *Nature*, pages 1–5, 2021.

[19] R. Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, 1988.

[20] M. D. McDonnell and N. G. Stocks. Maximally informative stimuli and tuning curves for sigmoidal rate-coding neurons and populations. *Physical Review Letters*, 101(5):058103, 2008.

[21] Z. Wang, A.A. Stocker, and D.D. Lee. Optimal neural tuning curves for arbitrary stimulus distributions: Discrimax, Infomax and minimum L_p loss. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems NIPS 25*, pages 2177–2185. MIT Press, May 2012.

[22] Deep Ganguli and Eero P Simoncelli. Efficient sensory encoding and bayesian inference with heterogeneous neural populations. *Neural computation*, 26(10):2103–2134, 2014.

[23] Thomas E Yerxa, Eric Kee, Michael R DeWeese, and Emily A Cooper. Efficient sensory coding of multidimensional stimuli. *PLoS computational biology*, 16(9):e1008146, 2020.

[24] R Taylor and PM Bays. Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *Journal of Neuroscience*, 2018. doi: 10.1523/JNEUROSCI.1018-18.2018.

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

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

[27] Samuel J Cheyette and Steven T Piantadosi. A unified account of numerosity perception. *Nature Human Behaviour*, 4(12):1265–1272, 2020.

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

[29] S.M. Zeki. Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology, London*, 236:549–573, 1974.

[30] William T Newsome and Edmond B Pare. A selective impairment of motion perception following lesions of the middle temporal visual area (mt). *Journal of Neuroscience*, 8(6):2201–2211, 1988.

[31] J Anthony Movshon and William T Newsome. Visual response properties of striate cortical neurons projecting to area mt in macaque monkeys. *Journal of Neuroscience*, 16(23):7733–7741, 1996.

[32] N.J. Priebe, C.R. Cassanello, and S.G. Lisberger. The neural representation of speed in macaque area MT/V5. *The Journal of Neuroscience*, 23(13):5650–5661, July 2003.

[33] Harris Nover, Charles H Anderson, and Gregory C DeAngelis. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *Journal of Neuroscience*, 25(43):10049–10060, 2005.

[34] N. Brunel and J.P. Nadal. Mutual information, Fisher information, and population coding. *Neural Computation*, 10(7):1731–1757, Oct 1998.

[35] Xue-Xin Wei and Alan A Stocker. Mutual information, Fisher information, and efficient coding. *Neural computation*, 28(2):305–326, 2016.

[36] G. Sclar, J. Maunsell, and P. Lennie. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, 30(1):1–10, 1990.

[37] Duane G Albrecht and David B Hamilton. Striate cortex of monkey and cat: contrast response function. *Journal of neurophysiology*, 48(1):217–237, 1982.

[38] Hilary W Heuer and Kenneth H Britten. Contrast dependence of response normalization in area mt of the rhesus macaque. *Journal of neurophysiology*, 88(6):3398–3408, 2002.

[39] Suzanne P McKee, Gerald H Silverman, and Ken Nakayama. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision research*, 26(4):609–619, 1986.

[40] Bart De Bruyn and Guy A Orban. Human velocity and direction discrimination measured with random dot patterns. *Vision research*, 28(12):1323–1335, 1988.

[41] Steven C Panish. Velocity discrimination at constant multiples of threshold contrast. *Vision research*, 28(2):193–201, 1988.

[42] Kathleen Turano and Allan Pantle. On the mechanism that encodes the movement of contrast variations: velocity discrimination. *Vision Research*, 29(2):207–221, 1989.

[43] Mark S Horswill and Annaliese M Plooy. Reducing contrast makes speeds in a video-based driving simulator harder to discriminate as well as making them appear slower. *Perception*, 37(8):1269–1275, 2008.

[44] F. Hürlimann, D. Kiper, and M. Carandini. Testing the Bayesian model of perceived speed. *Vision Research*, 42:2253–2257, 2002.

[45] Bas Rokers, Jacqueline M Fulvio, Jonathan W Pillow, and Emily A Cooper. Systematic misperceptions of 3-d motion explained by bayesian inference. *Journal of vision*, 18(3):23–23, 2018.

[46] Kaushik J Lakshminarasimhan, Marina Petsalis, Hyeshin Park, Gregory C DeAngelis, Xaq Pitkow, and Dora E Angelaki. A dynamic bayesian observer model reveals origins of bias in visual path integration. *Neuron*, 99(1):194–206, 2018.

[47] K. Britten, M. Shadlen, W. Newsome, and J.A. Movshon. Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, 10:1157–1169, 1993.

[48] Larry F Abbott and Peter Dayan. The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101, 1999.

[49] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–366, 2006.

[50] Rubén Moreno-Bote, Jeffrey Beck, Ingmar Kanitscheider, Xaq Pitkow, Peter Latham, and Alexandre Pouget. Information-limiting correlations. *Nature neuroscience*, 17(10):1410–1417, 2014.

[51] Adam Kohn, Ruben Coen-Cagli, Ingmar Kanitscheider, and Alexandre Pouget. Correlations and neuronal population information. *Annual review of neuroscience*, 39:237–256, 2016.

[52] Ingmar Kanitscheider, Ruben Coen-Cagli, Adam Kohn, and Alexandre Pouget. Measuring fisher information accurately in correlated neural populations. *PLoS computational biology*, 11(6):e1004218, 2015.

[53] Xin Huang and Stephen G. Lisberger. 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. *Journal of Neurophysiology*, 101(6):3012–3030, 2009. doi: 10.1152/jn.00010.2009.

[54] Jean-Paul Noel, Ling-Qi Zhang, Alan A. Stocker, and Dora E. Angelaki. Individuals with autism spectrum disorder have altered visual encoding capacity. *PLOS Biology*, 19(5):1–21, 05 2021. doi: 10.1371/journal.pbio.3001215.

[55] Alan A Stocker, Najib Majaj, Chris Tailby, J Anthony Movshon, and Eero P Simoncelli. Decoding velocity from population responses in area MT of the macaque. *Journal of Vision*, 9 (8):741–741, 2009.

[56] C.C. Pack, J.N. Hunter, and R.T. Born. Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology*, 93:1809–1815, 2005.

[57] B. Krekelberg, R.J.A. van Wezel, and T. Albright. Interactions between speed and contrast tuning in the middle temporal area: Implications for the neural code for speed. *Journal of Neuroscience*, 26:8988–8998, August 2006.

[58] Gustav Theodor Fechner. *Elemente der psychophysik*, volume 2. Breitkopf u. Härtel, 1860.

[59] Johannes Burge and Wilson S Geisler. Optimal speed estimation in natural image movies predicts human performance. *Nature communications*, 6(1):1–11, 2015.

[60] D. Ganguli and E.P. Simoncelli. Implicit encoding of prior probabilities in optimal neural populations. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Adv. Neural Information Processing Systems 23*, volume 23, pages 658–666, Cambridge, MA, December 2010. MIT Press.

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

[62] Tinku Acharya and Ajoy K Ray. *Image processing: principles and applications*. John Wiley & Sons, 2005.

[63] Irene Senna, Cesare V. Parise, and Marc O. Ernst. Hearing in slow-motion: Humans underestimate the speed of moving sounds. *Scientific Reports*, 5(1):14054, 2015. doi: 10.1038/srep14054.

[64] Reuben Rideaux and Andrew E Welchman. But still it moves: static image statistics underlie how we see motion. *Journal of Neuroscience*, 40(12):2538–2552, 2020.

[65] Michel Treisman. Noise and weber's law: The discrimination of brightness and other dimensions. *Psychological review*, 71(4):314, 1964.

[66] Andreas Nieder and Earl K Miller. Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1):149–157, 2003.

[67] Gustav Theodor Fechner, Davis H Howes, and Edwin Garrigues Boring. *Elements of psychophysics*, volume 1. Holt, Rinehart and Winston New York, 1966.

[68] Ron O Dror, Alan S Willsky, and Edward H Adelson. Statistical characterization of real-world illumination. *Journal of Vision*, 4(9):11–11, 2004.

[69] Stanislas Dehaene and Jacques Mehler. Cross-linguistic regularities in the frequency of number words. *Cognition*, 43(1):1–29, 1992.

[70] Steven T Piantadosi and Jessica F Cantlon. True numerical cognition in the wild. *Psychological science*, 28(4):462–469, 2017.

[71] Megan AK Peters, Jonathan Balzer, and Ladan Shams. Smaller= denser, and the brain knows it: natural statistics of object density shape weight expectations. *PloS one*, 10(3):e0119794, 2015.

[72] Yong Gu, Christopher R Fetsch, Babatunde Adeyemo, Gregory C DeAngelis, and Dora E Angelaki. Decoding of mstd population activity accounts for variations in the precision of heading perception. *Neuron*, 66(4):596–609, 2010.

[73] Deep Ganguli and Eero P Simoncelli. Neural and perceptual signatures of efficient sensory coding. *arXiv preprint arXiv:1603.00058*, 2016.

[74] Chris R Sims. Efficient coding explains the universal law of generalization in human perception. *Science*, 360(6389):652–656, 2018.

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

[76] Michael Morais and Jonathan W Pillow. Power-law efficient neural codes provide general link between perceptual bias and discriminability. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 5071–5080. Curran Associates, Inc., 2018.

[77] Luke Rast and Jan Drugowitsch. Adaptation properties allow identification of optimized neural codes. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances*

in Neural Information Processing Systems, volume 33, pages 1142–1152. Curran Associates, Inc., 2020.

[78] Stefan Roth and Michael J Black. On the spatial statistics of optical flow. *International Journal of Computer Vision*, 74(1):33–50, 2007.

[79] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *International journal of computer vision*, 92(1):1–31, 2011.

[80] Shiva R Sinha, William Bialek, and Rob R De Ruyter Van Steveninck. Optimal local estimates of visual motion in a natural environment. *Physical review letters*, 126(1):018101, 2021.

[81] Dawei W Dong and Joseph J Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–358, 1995.

[82] Vasha DuTell, Agostino Gibaldi, Giulia Focarelli, Bruno Olshausen, and Marty Banks. The spatiotemporal power spectrum of natural human vision. *Journal of Vision*, 20(11):1661–1661, 2020.

[83] V. de Gardelle, S. Kouider, and J. Sackur. An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. *Journal of Vision*, 10(10), 2010.

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

[85] B. T. Crane. Direction specific biases in human visual and vestibular heading perception. *PLOS ONE*, 7(12):1–15, 12 2012. doi: 10.1371/journal.pone.0051383.

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

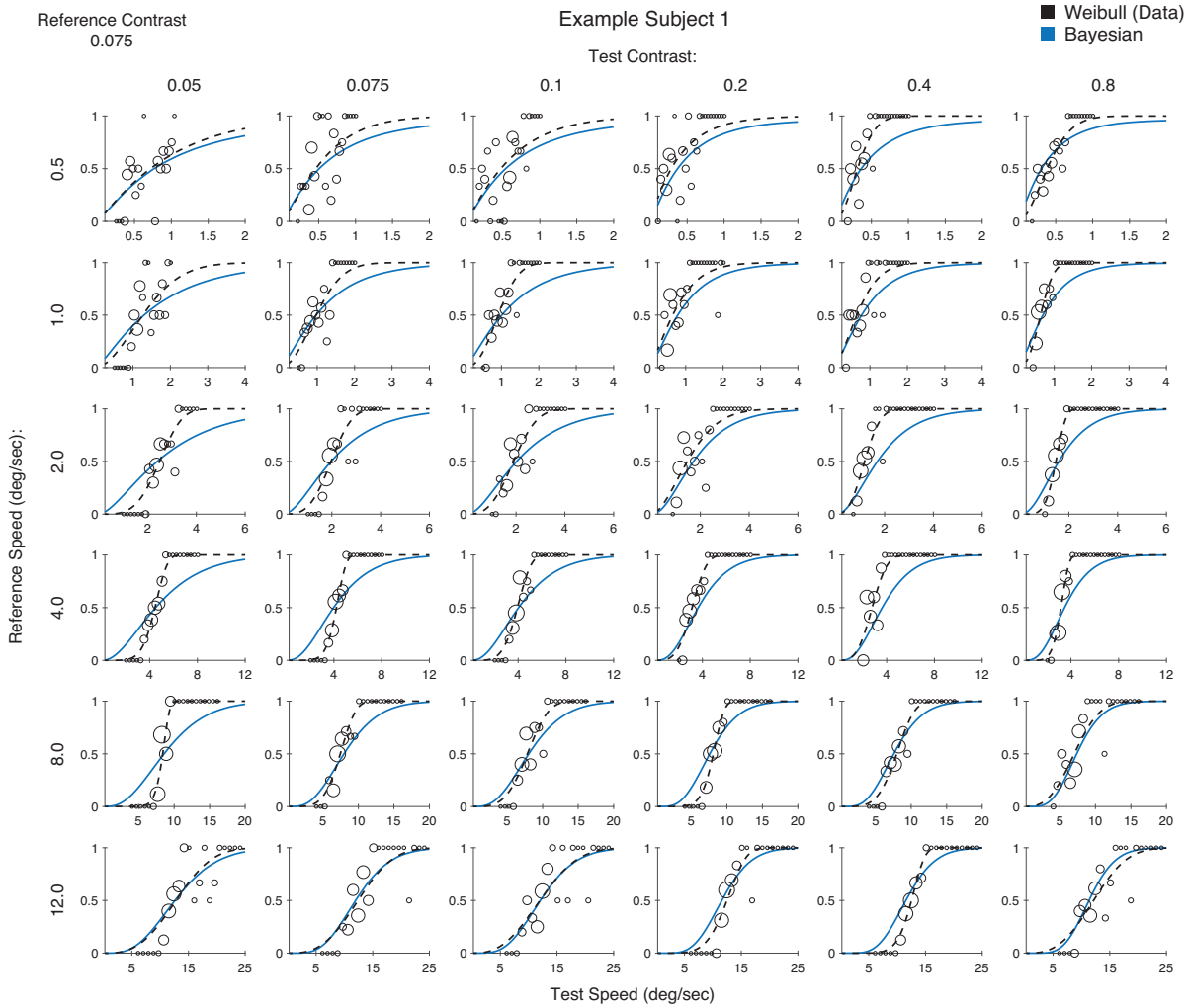
[87] Odelia Schwartz, Terrence J Sejnowski, and Peter Dayan. Perceptual organization in the tilt illusion. *Journal of Vision*, 9(4):19–19, 2009.

[88] Colin WG Clifford, Michael A Webster, Garrett B Stanley, Alan A Stocker, Adam Kohn, Tatyana O Sharpee, and Odelia Schwartz. Visual adaptation: Neural, psychological and computational aspects. *Vision research*, 47(25):3125–3131, 2007.

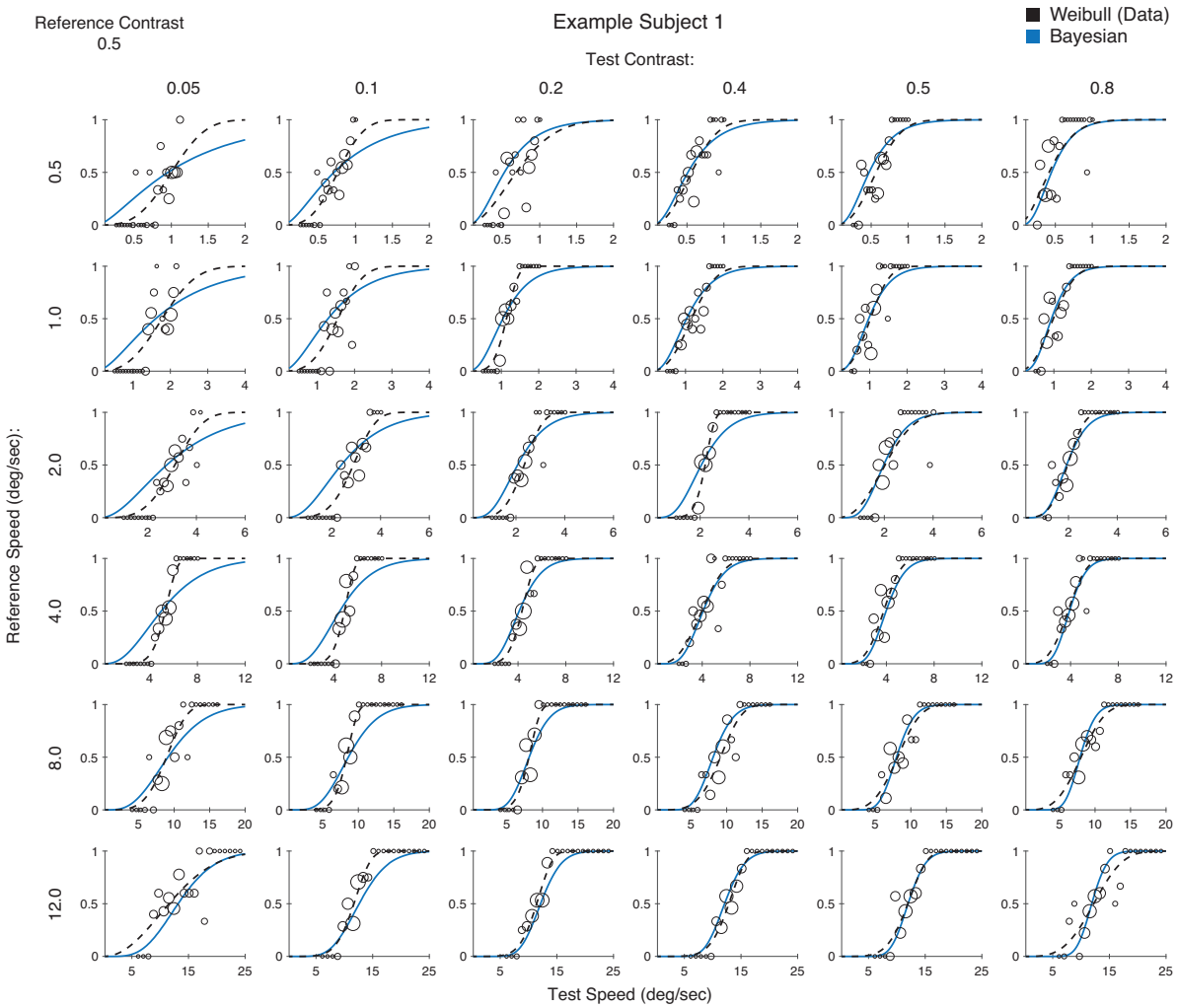
[89] X.-X. Wei, P. Ortega, and A.A. Stocker. Perceptual adaptation: Getting ready for the future. In *Vision Science Society VSS conference*, May 2015. doi: doi:10.1167/15.12.388.

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

Extended Data Figures



Extended Data Figure 2 - 1: Psychometric curves across all conditions for exemplary Subject 1 (low-contrast reference). Contrast of the reference stimulus was 0.075. Each row corresponds to a difference reference speed as indicated on the left, and each column corresponds to a different test contrast as indicated on top. The dashed black curve represents a Weibull fit to the data of each individual condition, and the blue curve is the Bayesian model fit to the entire data set. The size of the circle is proportional to the number of trials at that value.



Extended Data Figure 2 - 2: Psychometric curves across all conditions for exemplary subject 1 (high-contrast reference). Contrast of the reference stimulus was 0.5. Each row corresponds to a difference reference speed as indicated on the left, and each column corresponds to a different test contrast as indicated on top. The dashed black curve represents a Weibull fit to the data of each individual condition, and the blue curve is the Bayesian model fit to the entire data set. The size of the circle is proportional to the number of trials at that value.