PREPRINT

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, 10 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 fits extensive psychophysical data of 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 the psychophysical data ("behavioral prior"), to the point that they are indistinguishable in a model cross-validation 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. For example, contrast strongly affects perceived stimulus speed such that a drifting grating of low contrast 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 measurements with a prior preference for lower speeds [4–6]. Furthermore, previous work has shown that by embedding the Bayesian observer within a full behavioral model one can fit the model to psychophysical data and thus "reverse-engineer" the noise characteristics (i.e., likelihood) and prior expectations human subjects employed in a given speed-discrimination task [3]. 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. 7–10]. However, recovering the parameters of Bayesian observer models from behavioral data is typically difficult due to the intrinsic non-specificity of the probabilistic model formulation, which was raised as a general criticism of the Bayesian modeling approach [11, 12]. In fact, 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 resolved this potential problem by developing and validating a much better 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 [13, 14]. 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 [15–17]. It thus establishes a direct relationship between the stimulus distribution and the accuracy of neural representations in sensory systems [18–22]. Wei and Stocker [14] showed how to formulate efficient coding as an information constraint that can be embedded within the probabilistic language of the Bayesian framework. The resulting, highly constrained Bayesian observer model has proven to account for a wide range of phenomena in perception, including repulsive biases in perceived visual orientation [14, 23] and the lawful relationship between perceptual bias and discrimination threshold [24], but also in cognitive processes such as in subjective preferences judgments [25] or the representation of numbers [26].

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 the same speed discrimination data set as in the original study by Stocker and Simoncelli [3]. The data contains human psychophysical measurements over a broad range of speeds and stimulus uncertainties. We found that the new model

allowed us to reverse-engineer much more reliable and consistent estimates of individual subjects' prior beliefs about visual speed than in the original study, while still accurately accounting for their overall psychophysical behavior. Second, we then tested whether the reverse-engineered prior expectations from the psychophysical data correctly predicted the inhomogeneous 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 [27–30]. By analyzing single cell recordings of a large population of MT neurons [31], we found that the sensitivity with which visual speed is encoded in this population ("neural prior") is precisely predicted by the prior 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 (e.g. contrast-induced biases and Weber's law) to neurophysiological data (e.g. logarithmic encoding in area MT).

Results

We model speed perception as a typical two-stage, encoding-decoding process (Fig. 1A). First, for any given trial the speed v of a visual stimulus is represented by a noisy and bandwidth-limited sensory measurements m. We assume that this "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 [19, 32, 33], that is

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

This leads to a more accurate encoding of speeds for which the prior density is high, and thus ultimately determines the uncertainty (i.e., the likelihood function $p(m|v)$) the observer has about the actual stimulus speed given a particular sensory measurement. Next, 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. The latter two steps are usually termed "decoding".

An important aspect of the new model is that the prior distribution jointly determines both encoding and decoding of the Bayesian observer model [13]. Thus, both the encoding characteristics of MT neurons (Fig. 1B), and the psychophysical behavior of subjects in speed perception (Fig. 1C) should naturally be consistent with the belief of the observer about the statistical regularities of visual speed in the environment.

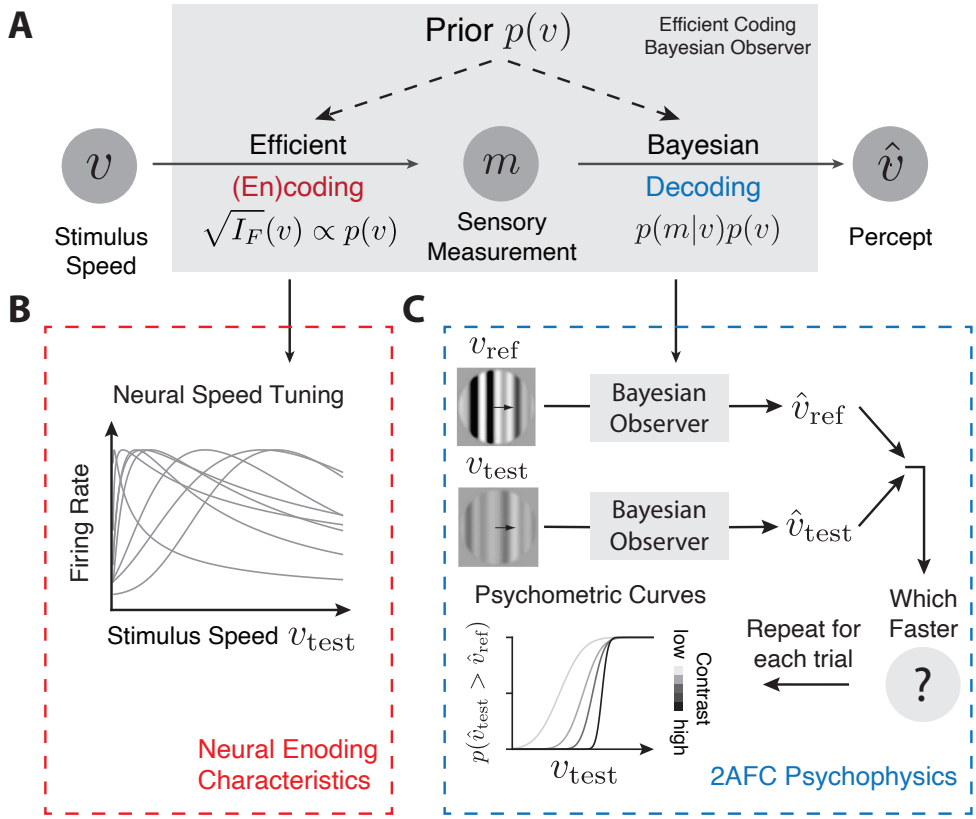


Figure 1: Observer Model. **A)** We model speed perception as an efficient encoding, Bayesian decoding process [13, 14]. Stimulus speed v is first encoded by a noisy and resource-limited sensory measurement m with an encoding accuracy that is determined by the stimulus prior $p(v)$ according to the chosen efficient coding constraint (Eq. (1)). Ultimately, a percept $\hat{v}$ 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 according to some 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 model 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 the new model to the large data set of the psychophysical speed discrimination experiment by Stocker and Simoncelli [3]. In each trial of the 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 [13]. 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 [14]. Specifically, a stimulus v is encoded with noisy sensory measurement m according to the following:

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

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 $p(v)$ 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 the 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 results [3, 8, 10], 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 were free and unconstrained parameters. Figure 2A shows the data and model fit for a few example conditions for exemplary subject #1. Overall, the model predicts psychometric curves that are similar to those from fitting a Weibull function. The overall log-likelihood of the new

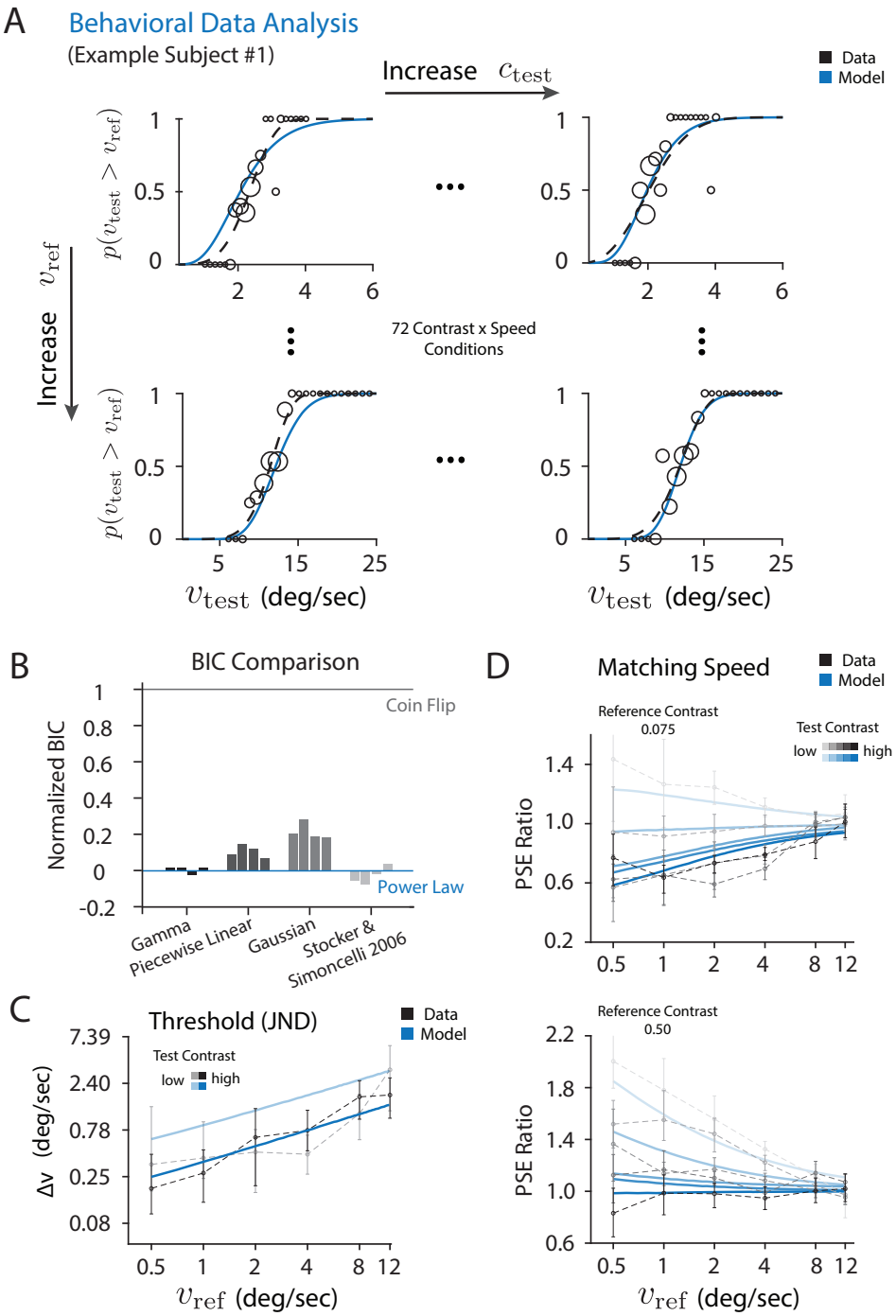


Figure 2: Extracting the "behavioral prior". **A)** We jointly fit the Bayesian observer model to psychophysical data across all contrast and reference speed conditions (72 psychometric functions). Shown are a few conditions for exemplary subject #1. Circle sizes are proportional to the number of trials at that test speed. See Supplementary Figs. 1 and 2 for the complete set of psychometric curves and model fits to all 72 conditions.

Figure 2 *(previous page)*: **B)** Goodness-of-fit comparison: Shown are the relative BIC values for individual subjects, normalized to the range set by the lower-bound value of the efficient coding Bayesian model with power-law parameterization, and an upper-bound coin flip model. Different shades correspond to three other prior parameterizations (Gamma distribution, piece-wise log-linear function, and a Gaussian distribution respectively, see *Methods*), as well as the original, less constrained Bayesian observer model by Stocker and Simoncelli [3]. We provide a more detailed model comparison with a cross-validation procedure later. **C)** Weber's law: 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). **D)** Contrast-induced slow-speed biases: Test relative to reference speeds at the point of subjective equality (PSE) extracted from the Weibull fits (black) and our model (blue). Gray 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 top panel, and 0.5 in the bottom panel. Error bars indicated $\pm$ s.d. across 500 bootstrap runs. See Supplementary Fig. 5 for the same set of threshold and bias plots for all four subjects.

model is close to that of separate Weibull fits to each individual conditions (sed Supplementary Fig. 4A). Comparing the Bayesian Information Criterion (BIC) values relevaled that the new model performed as well as the original, less constrained Bayesian observer model [3]. A more detailed model comparison with a cross-validation procedure can be found in the section below. The complete set of psychometric curves and model fits for exemplary subject #1 can be found as Supplementary Figs. 1 and 2.

More important, however, is the fact that the reverse-engineered prior expectations are much more consistent across individual subjects compare to those from the previous model [3, 10]: For all subjects, the exponent c_0, for example, is close to a value of -1 indicating much better constrained prior estimates (see Tables 1 for numerical values of all prior parameters across subjects). Furthermore, the contrast-dependent noise parameters monotonically decrease as a function of contrast as expected (Table 2) and are consistent with the functional description of the contrast response curve of cortical neurons [3, 34] (Supplementary 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 also 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 (Supplementary Fig. 4A), although the BIC value is higher for the log-linear parameterization due to its large number of parameters (Fig. 2B). Crucially, how-

ever, the shapes of the fit prior distributions are very similar across the different parameterizations, exhibiting a power-law like, slow-speed preferred distribution, with the exception of the Gaussian parameterization, which likely produced a poor model fit because it cannot well approximate a power-law function (Supplementary Fig. 4B).

Lastly, we further validated our fit model by comparing its predictions for discrimination thresholds and contrast-induced biases to subjects' behavior. Subjects' thresholds (JND) and biases (PSE) were extracted from Weibull fits to each individual psychometric curve. As shown in Fig. 2C, subjects' discrimination thresholds increase monotonically with the speed. While they follow Weber's law at higher speeds, they deviate from a constant Weber-fraction at slow speeds as has been well documented [3, 35, 36]. Furthermore, subjects clearly under-estimate the speeds of low-contrast stimuli (Fig. 2D), an effect that occurs at any contrast level and any speeds although the relative bias seems to decrease for faster stimulus speeds for this particular subject. Our new model is able to capture both the discrimination threshold behavior and the contrast-induced slow-speed biases as indicated by the fits, comparable to the original model [3]. Supplementary Fig. 5 shows the same set of threshold and bias plots for all four subjects.

Note that in previous models, the focus was on the contrast-induced biases and how they can be attributed to a slow-speed prior that shifts the percept toward slower speeds for increasing levels of stimulus uncertainty (i.e., increased sensory noise with lower contrast). 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 faster 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 of exactly -1) predicts both, the constant Weber fraction at faster speed and its deviation at slow speeds, thus closely resembling human behavior (see *Methods*). Supplementary Fig. 6 shows the numerically predicted Weber fraction based on the reverse-engineered prior for individual subjects.

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 [13, 21, 33]. Thus, if our new model is correct, then the reverse-engineered "behavioral prior" above should be a good predictor of the neural encoding characteristics of visual speed. More concretely, we expect the neural encoding accuracy, measured as the square-root of the neural population FI, to match the extracted speed prior (Eq. (1)).

In order to test this strong prediction, we analyzed the encoding characteristics of a large population of neurons in area MT, widely considered the cortical locus of representing local visual motion

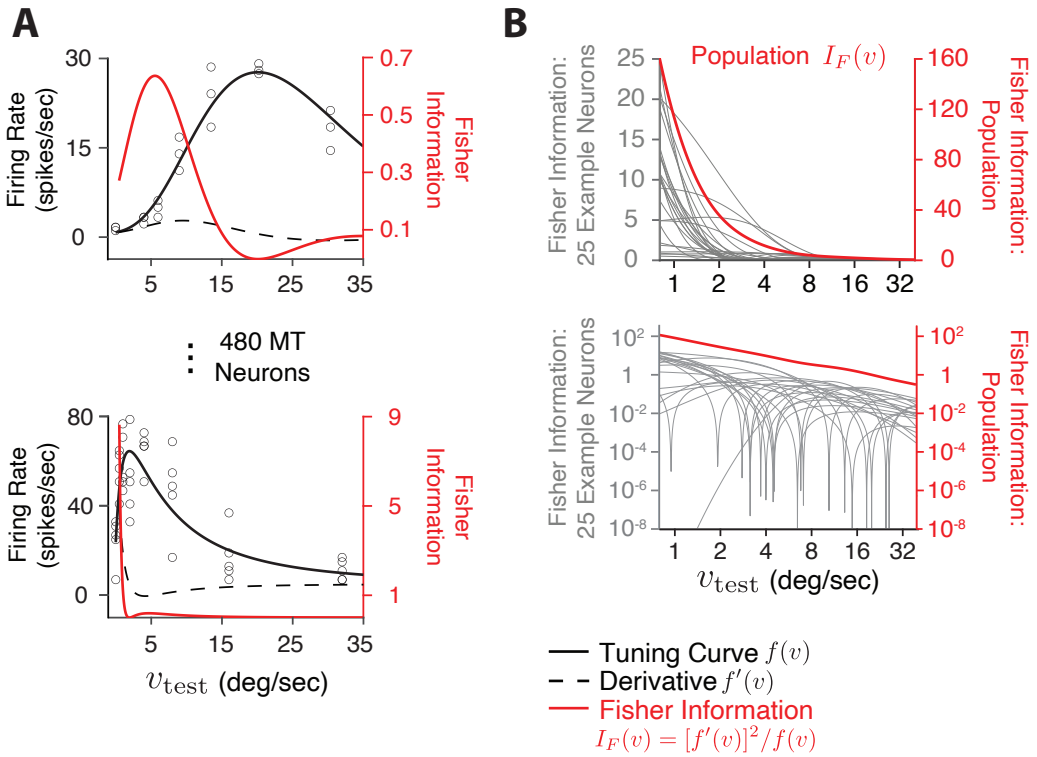


Figure 3: Extracting the "neural prior". According to the efficient coding constraint (Eq. (1) the accuracy of the neural representation of visual speed, measured by FI, should directly reflect the prior distribution of visual speed. **A)** Two example MT neurons from the data set [31] with fitted log-normal tuning curves (black, left y-axis) and their corresponding FI (red, right y-axis) assuming a Poisson noise model. **B)** 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 data set. The y-axis is on a linear scale in the upper panel, and a logarithmic scale in the lower panel.

in the primate brain [27, 37]. Our analysis was based on a data set of single cell recordings in the macaque monkey [31]. The data set consists of repeated spike counts from 480 MT neurons responding to random dot motion stimuli moving at eight different speeds (see *Methods* and Nover et al. [31] for more details). To extract the population FI, we followed the original study [31] and fit a log-normal speed tuning curve model for each neuron in the data set. We also tested an alternative tuning curve model using a Gamma distribution function and obtained similar results. Assuming that neural response variability is well captured by a Poisson distribution, it is then straight-forward to compute the FI of individual neurons (Fig. 3B - see *Methods* for details). Finally, by assuming independent noise among the neurons in the recorded population, the population FI then reflects the sum of the FI of individual neurons (Fig. 3C). The shape of the resulting population FI is very close to a power-law function: When plotted on a log-log scale it closely resembles a straight line. Although noise correlations between pairs of neurons in area MT exist [38] and noise correlation inevitably affect the population FI [39–41], our main interest is in determining the overall shape of the population FI. As long as the correlation pattern is homogeneous at the level of the efficient neural representation (i.e. in cortical space), noise correlations will not change the shape of the population FI and only contribute a constant scaling factor.

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 model [14, 33, 42]. Although the fit contrast-dependent noise parameters for all subjects are consistent with the typical contrast response function of cortical neurons (see Supplementary Fig. 3) [43, 44], 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. The analysis of an unpublished, limited neural data set [45] suggests, however, that stimulus contrast indeed just scales the population FI without changing its shape, which is interesting given the reported diverse and inhomogeneous effects contrast has on the speed tuning curves of individual neurons [46].

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 4A 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 (which corresponds to the slope in the log-log plot of Fig. 4A). Not only is the neural prior well predicted by the behavior prior but

also the difference across different subjects is small, quite in contrast to the priors extracted with the old observer model [3, 10].

To quantitatively compare the effective similarity between the behavior and neural prior, we constructed a “neural observer” model for which the prior used by the model was fixed to be the neural prior extracted from the MT data, with only the contrast-dependent noise to as the free model parameters. A cross-validation procedure then compared the log-likelihood of that neural observer, the efficient coding Bayesian observer, the original Bayesian model [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. 4B, we found that the validation performances are highly similar across all three models, and closely match to the Weibull fits. This comparison demonstrates several important aspects. First, it confirms that the behavioral and neural prior are behaviorally indistinguishable and thus effectively equivalent; if the neural data set did not exist, we would have been able to accurately predict the population encoding accuracy of MT neurons. Second, it demonstrates the excellent quality of the Bayesian observer model as its account of human behavior comes very close to that of the best possible parametric description of the data (Weibull). And finally, it shows that the complexity of our observer model is appropriate and thus does not lead to over-fitting of the data.

The particular 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 [47]. And indeed neural encoding of visual speed in area MT can be considered “logarithmic” [31, 48]: 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 nearly uniform density [31, 46]. 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 the 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 (identical tuning curves uniformly tiling the sensory space [14, 21]) of the variable of interest *transformed* by its cumulative distribution function (CDF), which is sometimes referred to as histogram equalization [13, 49, 50]. For our chosen parameterization of the power-law prior (Eq. (3)) with exponent -1, its CDF is the exact logarithmic function that best described MT tuning characteristics [31] (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 based on the efficient sensory representation of a power-law distributed perceptual variable with an exponent of -1 (Supplementary Fig. 6 shows predictions of the Weber fraction for individual

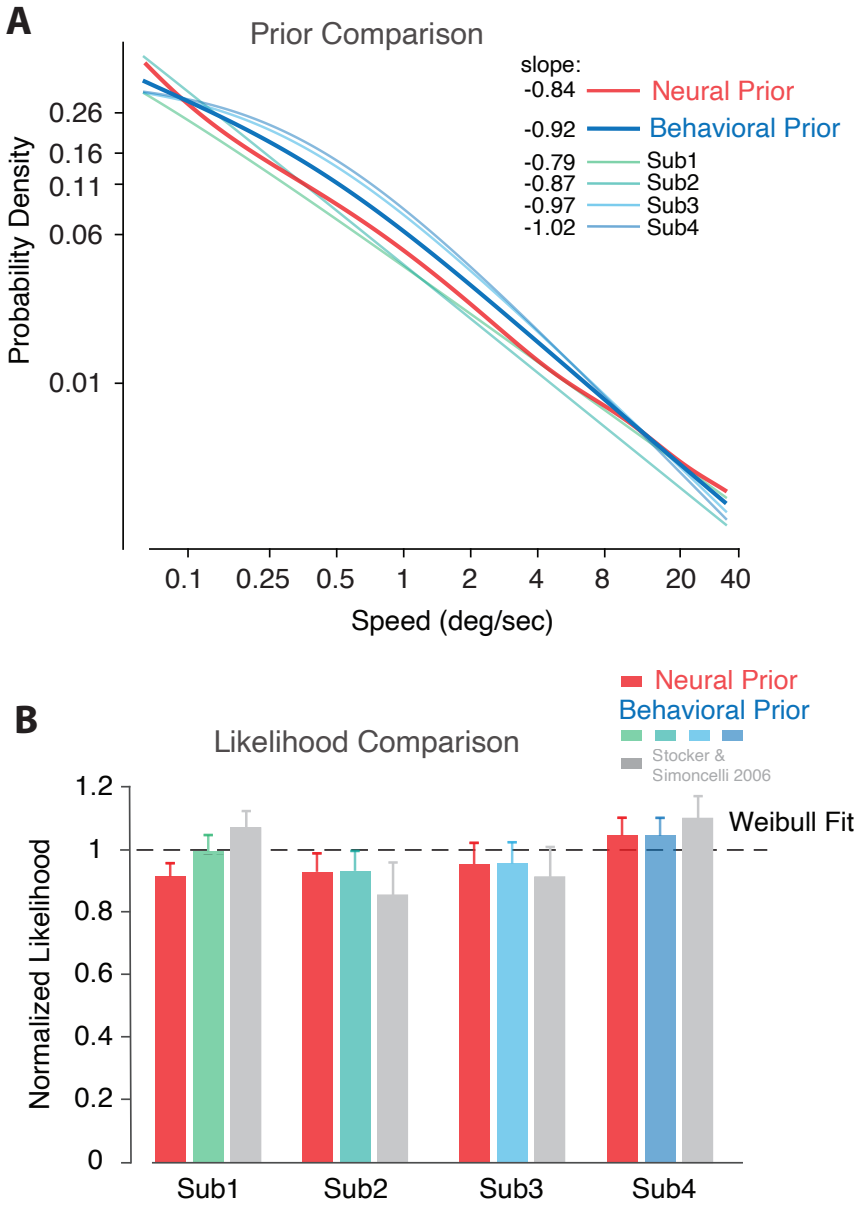


Figure 4: Comparing neural and behavioral prior. **A)** Reverse-engineered behavioral priors for individual subjects (light shades) and their average (dark blue) superimposed by the neural prior (dark red). Slope values are computed from a linear fit to the curves in the log-log domain. **B)** The cross-validated 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 [3], less constrained Bayesian model (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 $\pm$ s.d. across 100 validation runs. For each run, the models were fit to randomly sampled 80% and then validated against the remaining 20% of the speed discrimination data according to a five-fold cross-validation procedure. See *Methods* for details.

subjects' priors).

Discussion

We presented a Bayesian observer model constrained by efficient coding for human visual speed perception. We fit this model to an existing data set consisting of human 2AFC speed discrimination measurements over a wide range of stimulus contrast and speed, which allowed us to reverse-engineer the "behavior prior" that best accounts for the psychophysical behavior of 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 MT population.

Specifically, 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, 10], the added efficient coding constraint of the new formulation led to estimates of behavioral priors that were not only much more consistent across individual subjects but also were remarkably well matched by the population encoding characteristics of neurons in the motion sensitive area MT in the primate brain. Our work provides further evidence for the slow-speed prior interpretation of motion illusions [3, 6, 7, 10, 51] (but see [52]), which, in addition, is substantially strengthened by the quantitative support from the analysis of electrophysiological recordings.

Our study also offers an explanation for when perceptual variables have a logarithmic neural representation and thus follow Weber's law [47]. According to our model, logarithmic encoding and Weber's law are both a direct normative prediction based on the efficient representation of a perceptual variable with a prior distribution following a power-law function. We predict that perceptual variables that conform to Weber's law have distributions that follow a power-law function with an exponent of approximately -1 and are logarithmically encoded in the brain (although other alternative encoding solutions that satisfy the efficient coding constraint are possible, see [e.g. 14]). Indeed, perceptual variables that are known to follow approximately Weber's law such as light intensity [53], numerosity [54, 55], and weight [56], exhibit heavy tails in their statistical distributions

under natural environmental conditions [57–60], a defining feature of a power-law function. Consequently, any deviation from Weber's law (i.e., from a constant Weber-fraction; see Supplementary Fig. 6) and the logarithmic encoding should be reflected in deviations of the statistical stimulus distributions from a strict power-law function. Future studies of natural stimulus statistics, modeling of psychophysical data, and brain recordings will be needed to further and more quantitatively validate the generality of this prediction.

Our results demonstrated that it is possible to make quantitative predictions about the coding accuracy of neural population representations based on a behavioral theory and psychophysical data. Beginning with Fechner's seminal work of relating neural (logarithmic) encoding to perceptual discriminability of magnitude variables [47], more recent studies extended, within the context of efficient coding, this link from discriminability to the statistical prior distributions of perceptual variables [42, 61, 62]. 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 or bias), 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 more stringent test of the model but also permits more precise and consistent predictions of neural coding accuracy and priors.

Our comparison between behavioral and neural prior is limited to the extent that there were substantial experimental differences in the two data sets we considered. For example, we compared human with non-human primate data and estimated the neural tuning characterization based on single cell recordings using random-dot motion (and not grating stimuli as in the psychophysical experiment). Yet, the surprisingly accurate match of the priors indicates that these differences seem to matter little for how visual speed is neurally represented (at least at the level of the population FI) or perceptually processed. Which in turn can be taken as evidence that these priors are indeed representing the stimulus prior, as the latter is a property of the environment and not the observer. Nonetheless, recent studies have demonstrated that it is possible to quantitatively characterize the accuracy (i.e. uncertainty) with which a perceptual variable is represented in the human brain using voxel-level encoding models of fMRI signals. Using this technique, it has been shown, for example, that the well-known difference in perceptual precision between cardinal and oblique orientations matches their differences in neural encoding accuracy [63]. Similarly, others have investigated the neural basis of bias in motion perception based on such neuroimaging techniques [64]. Future work is needed to validate and maybe refine our estimates of the "neural prior" by using such voxel-level encoding models for human visual speed perception, which would also permit an investigation of individual perceptual differences at both behavioral and neural levels through matched task and stimulus design.

Note, however, the specific shape of the extracted neural and behavioral prior depends on our assumption of the objective function for efficiently representing visual speed. The particular form of the efficient coding constraint we currently adopted (Eq. (1)) results from maximizing the mutual information between the representation and stimulus [21, 33]. 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 [20, 65, 66], 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 is extensive. Previous work has shown, however, that for visual stimulus variables for which good estimates of the stimulus prior exist such as luminance contrast and orientation, encoding characteristic in early visual cortex are at minimum very closely described by the mutual information maximization objective [20].

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 approximates a similar power-law distribution that falls within the range given by the neural and behavioral priors illustrated in Fig. 4A. Validating this prediction by measuring the stimulus prior distribution, however, turns out to be 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 retinal motion signal, that ultimately is the input to the visual system. Thus the precise characterization of the retinal speed distribution requires us to accurately measure and calibrate these different types of motions, and the algorithms used to extract the motion information from some camera input. Previous studies have approximated these relative movements to various degree and precision, and also used different algorithms to extract local visual speed from spatio-temporal image streams, resulting in slightly different characterization of the prior distribution [67–70]. Common to all these measured stimulus priors, however, is that they have long-tail distributions with higher probability associated with slow speeds. Future work based on more comprehensive data sets [71] 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. 42, 72], surface tilt [73], or heading direction [74] to just name a few. 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, for example, adaptation aftereffects or the tilt-illusion [75–77]. It is commonly assumed that these biases are caused by a mismatch in expectation between the encoding and the decoding side (i.e. the "coding catastrophe" [75]), which is in stark contrast to the main assumption of our model. Preliminary

results are promising [24, 78]. However, more quantitative analyses are necessary to test how well the framework can account for the data and what neural, behavioral, and statistical 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, successful 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 data set on MT neurons, Benjamin Chin for his work on an earlier implementation of the model, 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 set

We reanalyzed the two alternative forced choice (2AFC) speed discrimination data from Stocker and Simoncelli [3]. We omit a detailed description of the psychophysical experiment as this can be found in the original article. Briefly, 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 the six different 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. 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 used the data of four of the five subjects as the fifth subject (labeled as subject #3 in the original dataset) only provided data at two contrasts and two test speeds. 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 [14] and embedded it within a decision process to predict the binary judgment of the subject in the 2AFC experiment [3]. For most of our analysis, we assumed a parametric description of the speed prior distribution as a modified power-law function $p(v) \propto (v + c_1)^{c_0} + c_2$ (Eq. (3)). However, we also tested various other parameterizations (see Supplementary Fig. 4).

Encoding can be described by the conditional probability $p(m|v)$; it determines how stimulus speed v is transformed probabilistically into a noise sensory measurement m. We can satisfy the efficient coding constraint $\sqrt{I_F}(v) \propto p(v)$ (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 $1/h(c)$.

Note that in practice, to handle the unbounded nature of a magnitude variable such as speed (as compare to circular variables), a small constant ($2.5 * 10^{-3}$) was added to $p(v)$ when computing the CDF such that it 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, thus we can explicitly express the dependency as $\hat{v}(m)$. However, in a psychophysi-

cal 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 *fminsearchbnd*. 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 distribution, 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 prior-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. According to our efficient coding constraint (Eq. (1)), the prior directly determines FI of the encoding. In addition, it has been shown that FI ($\sqrt{I_F}$) is inversely proportional to discrimination threshold (Δ_v) [24, 79]:

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

Substituting $\sqrt{I_F}$ with $p(v)$ (Eq. (1)), we already find

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

This equation allows us to predict threshold numerically (up to a scale factor) for any given prior. If we consider the specific case of the modified power-law prior with an exponent of -1, namely $p(v) \propto (v + c_0)^{-1}$, we get

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

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

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

when $v \gg c_0$, we have $\frac{c_0}{v} \approx 0$, thus the Weber fraction is a constant. As $v \to 0$, $\frac{c_0}{v} \to \infty$, causing $\frac{\Delta_v}{v}$ to increase. Both are determining features of the speed discrimination behavior in human subjects [3, 35, 36], also see Supplementary Fig. 6.

Furthermore, 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 a (modified) logarithmic transformation that has been previously used for describing the speed tuning of MT neurons [31].

Neural prior: MT encoding analysis

Data set

We reanalyzed the electrophysiological recording data from Nover et al. [31]. All experimental details can be found in the original article. Briefly, neurons in area MT of the macaque 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

Each neuron was fit with a speed tuning curve. Previous result [31] suggests that MT speed tuning can be well described by a log-Gaussian function

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

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 rate using MATLAB *fmincon*. Fitting the parameters with maximum likelihood assuming a Poisson model for firing rate variability produces very similar results.

In order to assess the population FI for encoding speed, we assume that the response variability of each neuron is well-described by a Poisson process, with noise being independent among the recorded population. The population FI is then calculated as

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

We extract the "neural prior" (the prior that corresponds to the measured MT encoding precision assuming efficient encoding) by computing the normalized square-root of FI, thus

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

We repeated the above analysis using a Gamma distribution function as alternative tuning-curve model and obtained similar results.

Cross-validation

Quantitative comparison between neural and behavioral prior was based on cross-validation statistics. We performed a five-fold cross-validation procedure: The dataset within each condition is first randomly and equally divided into 5 groups. For each group, the model was fitted on the remaining groups as the training set, and evaluated on that group as the validation set. Model validation

performance was measured as the log-likelihood of the validation data given the fit model. The entire procedure was repeated 20 times, resulting in 100 estimates of model validation likelihood. For the behavioral prior we considered the full observer model using the best fit prior for that run. For the neural prior we assumed the prior to be fixed and equal to the prior extracted from population FI, with only the contrast-dependent noise parameters being fit. The same procedure was used to compute the validation likelihood of the original, less constrained Bayesian model [3], and the Weibull fits to each psychometric curve. This likelihood of the Weibull fits served as an upper bound of model performance. Log-likelihood values in Fig. 4B were normalized to the range set by this upper bound and a lower bound represented by a coin-flip model for the decision (i.e. constant decision probability of 0.5). We report the median and $\pm$ s.d. of the normalized validation log-likelihood values across the 100 estimates.

Code and Data Availability

Data and analysis code for our paper 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] Alan Alfred Stocker. *Constraint Optimization Networks for Visual Motion Perception - Analysis and Synthesis*. PhD thesis, Dept. of Physics, Swiss Federal Institute of Technology ETH, Zürich, Switzerland, March 2002.

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

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

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

[13] 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.

[14] 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.

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

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

[17] 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.

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

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

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

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

[34] 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.

[35] 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.

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

[37] 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.

[38] 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.

[39] 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.

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

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

[42] 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.

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

[44] 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.

[45] 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.

[46] 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.

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

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

[49] 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.

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

[51] 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.

[52] 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.

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

[54] 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.

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

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

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

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

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

[60] 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.

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

[62] 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.

[63] 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.

[64] Brett Vintch and Justin L Gardner. Cortical correlates of human motion perception biases. *Journal of Neuroscience*, 34(7):2592–2604, 2014.

[65] 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.

[66] 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.

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

[68] 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.

[69] 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.

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

[71] 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.

[72] 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.

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

[74] 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.

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

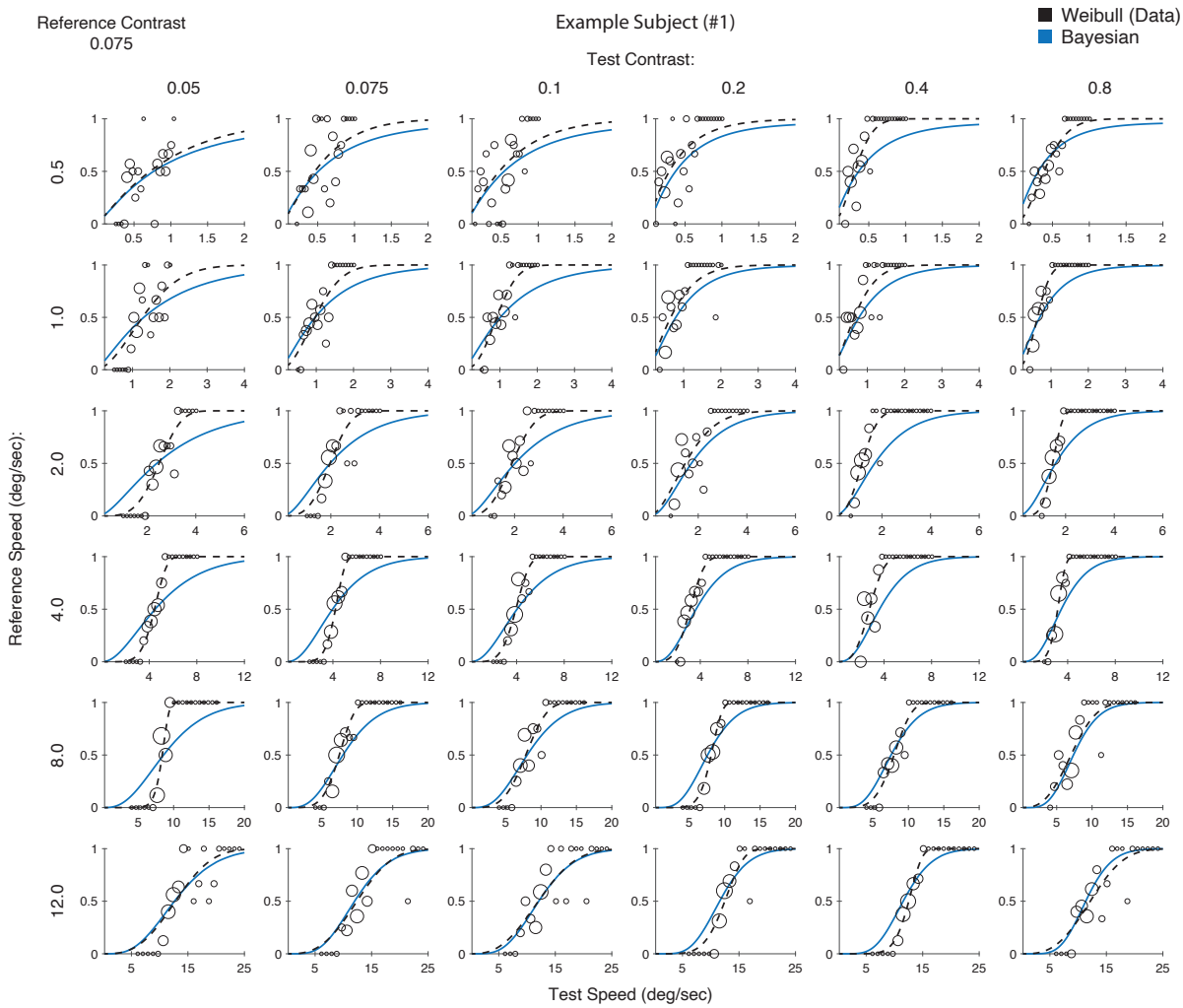
[76] Odelia Schwartz, Terrence J Sejnowski, and Peter Dayan. Perceptual organization in the tilt illusion. *Journal of Vision*, 9(4):19–19, 2009.

[77] 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.

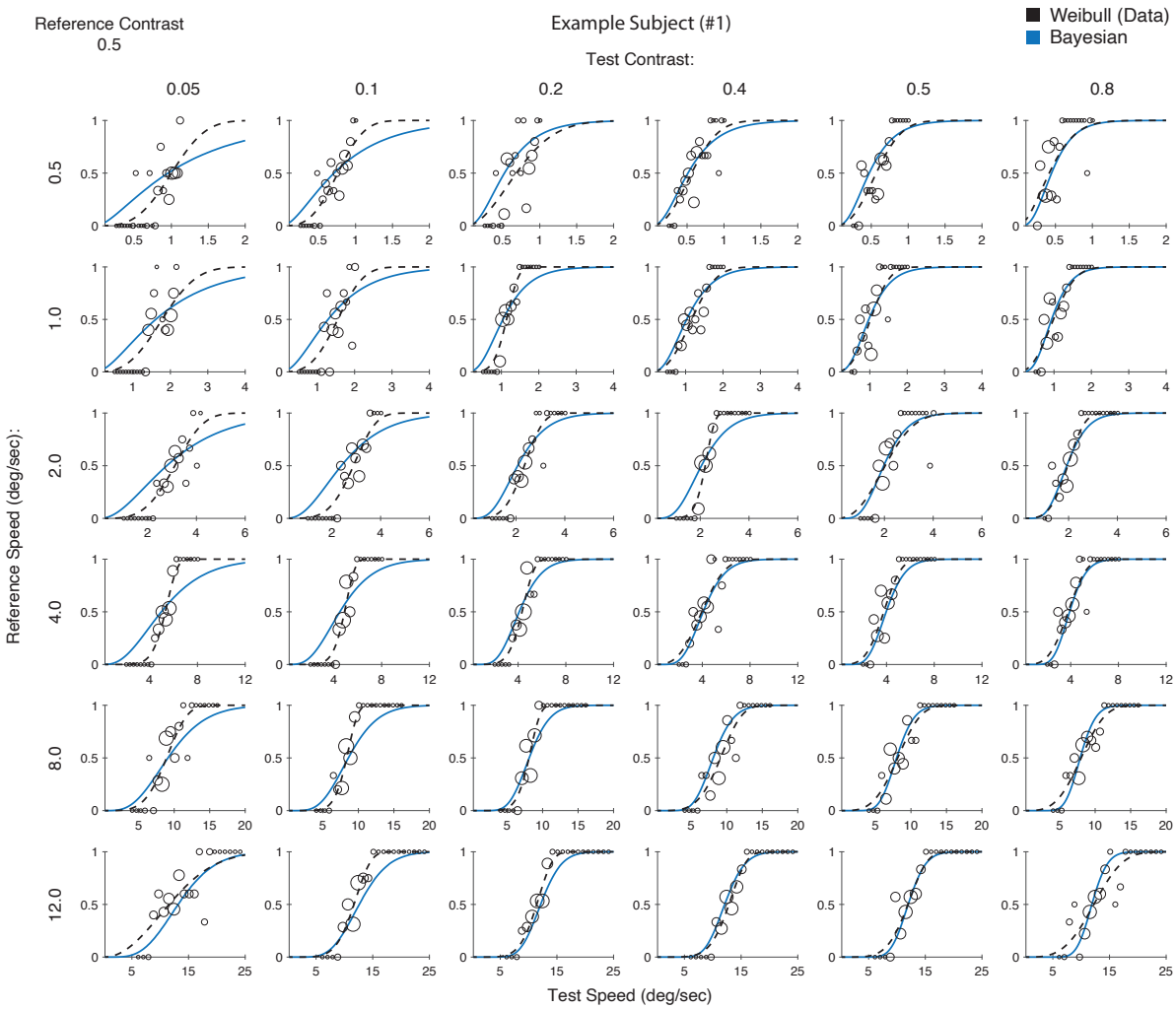
[78] X.-X. Wei, P. Ortega, and A.A. Stocker. Perceptual adaptation: Getting ready for the future. In *Vision Science Society VSS conference*, May 2015. doi: doi:10.1167/15.12.388.

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

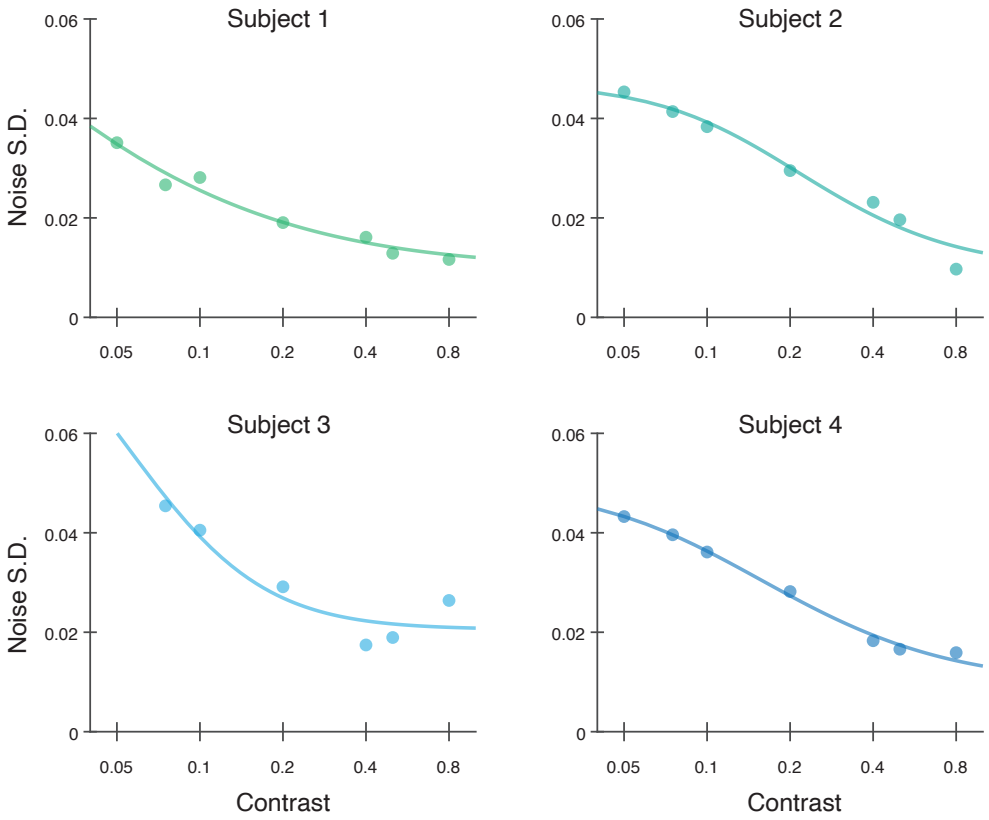
Supplementary information



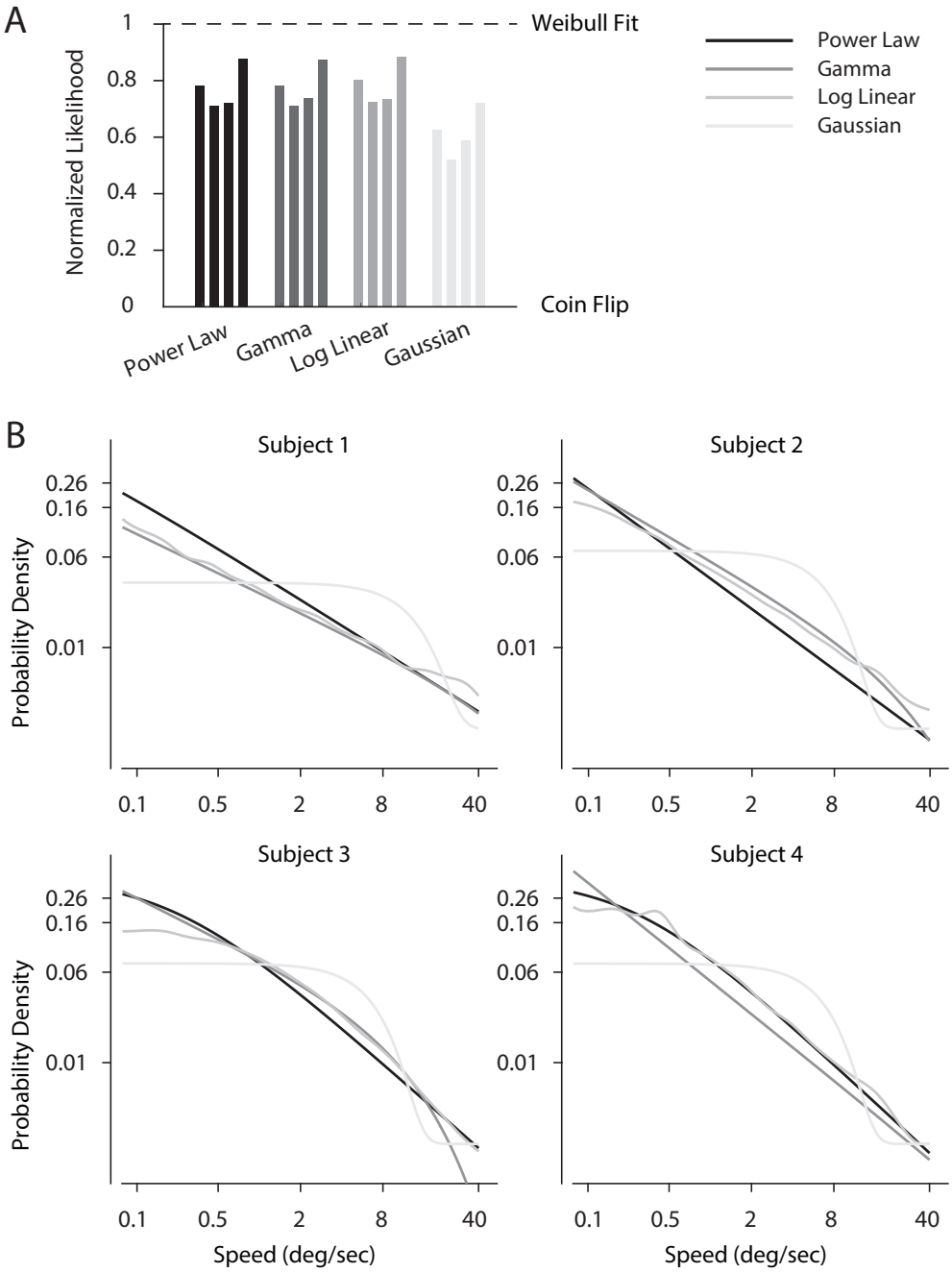
Supplementary Figure 1: Psychometric curves for low contrast reference stimuli for an example subject (#1). The contrast of the reference stimuli is 0.075. Each row corresponds to a difference reference speed as indicated on the left, and each column corresponds to a different test contrast as indicated on top. The dashed black curve is the Weibull fit to each 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.



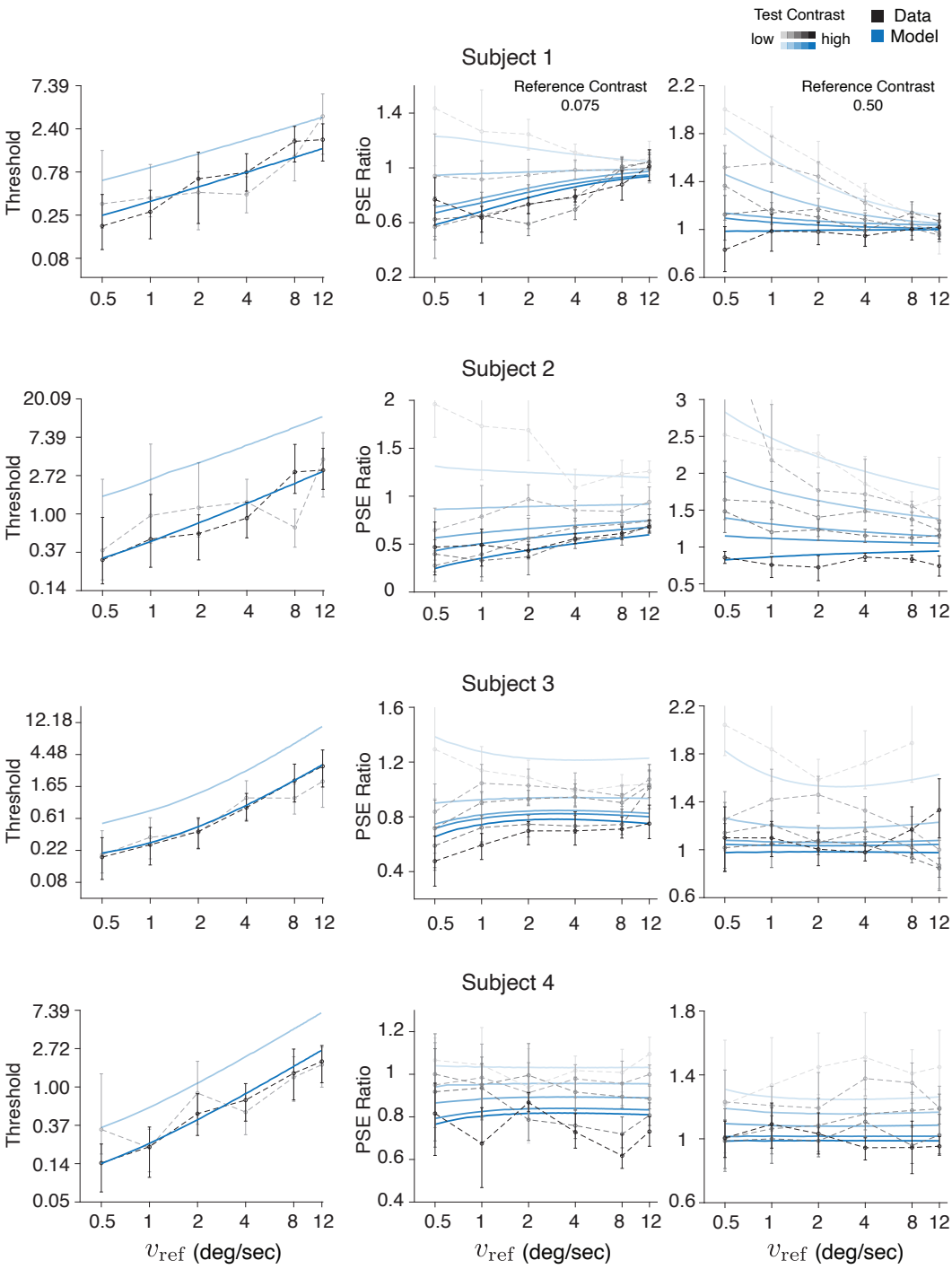
Supplementary Figure 2: Psychometric curves for high contrast reference stimuli for an example subject (#1). The contrast of the reference stimuli is 0.5. Each row corresponds to a difference reference speed as indicated on the left, and each column corresponds to a different test contrast as indicated on top. The dashed black curve is the Weibull fit to each 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.



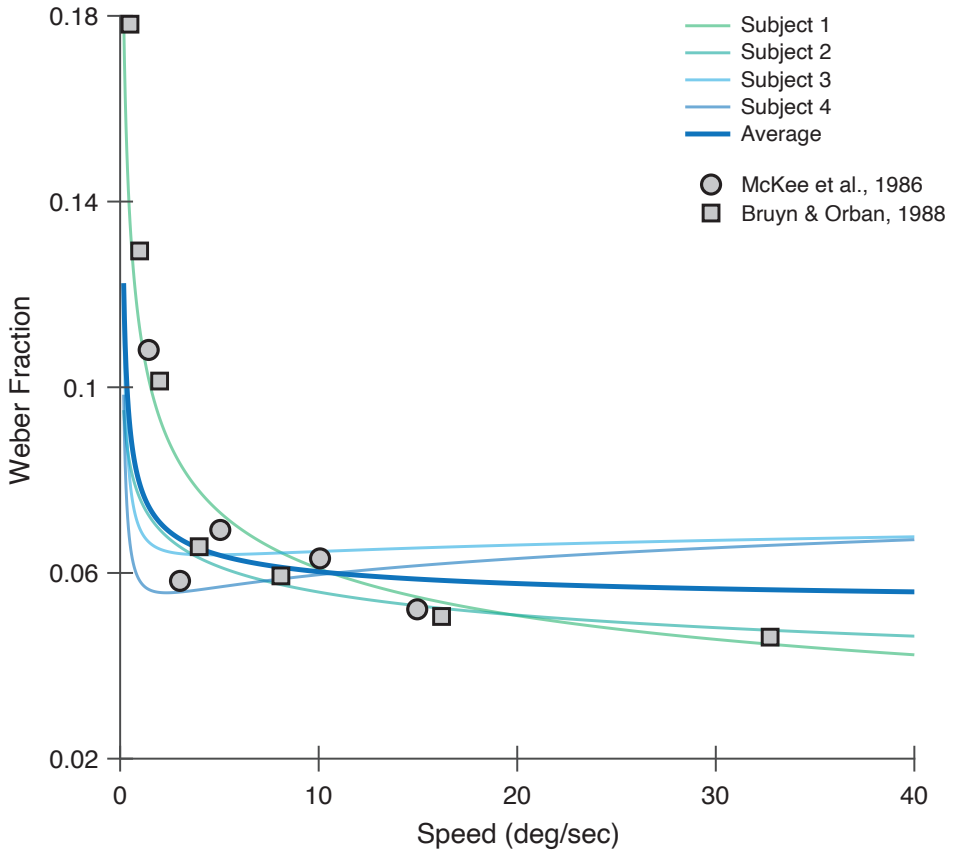
Supplementary Figure 3: Contrast-dependent noise. Noise parameters as a function of contrast for all individual subjects. We also fit a parametric function of the form $h(c) = [r_{\text{max}} c^q / (c^q + c_{50}^q + r_{\text{base}})]^{-1/2}$ based on the contrast response function of cortical neurons [3, 34].



Supplementary Figure 4: Prior parameterization. **A)** The normalized log-likelihood values and **B)** the probability density function for the best-fitted prior for each individual subjects, based on four different parameterizations including a power-law function, Gamma distribution, piece-wise log-linear function, and a Gaussian distribution respectively (see *Methods*).



Supplementary Figure 5: Weber's law-like discrimination thresholds (left column) and contrast-induced, relative biases (middle column - contrast of reference stimulus 0.075; right column - contrast of reference stimulus 0.5) for individual subjects extracted from the Weibull fits (black) and our model (blue). Note that whether these relative biases are converging or diverging depends on the exponent c_0 of the prior; with an exponent of exactly -1 these ratios are constant across speed (i.e. parallel horizontal lines) [24]. Error bars indicated $\pm$ s.d. across 500 bootstrap runs.



Supplementary Figure 6: Predicted Weber fraction. The Weber fraction $\Delta v/v$, predicted based solely on the estimated behavioral prior for four individual and the average subjects, is shown in comparison to previously reported psychophysical measurements [35, 36]. A scalar parameter was chosen such that the predictions are on a similar scale as the data. Note that De Bruyn and Orban [36] also found deviations from Weber's law at extremely high speeds (256 cycles/deg) which is not depicted here.

Subject	c_0	c_1	c_2
Sub1	-0.790	0.003	$6.15 * 10^{-5}$
Sub2	-0.867	0.002	10^{-8}
Sub3	-1.045	0.220	$1.3 * 10^{-5}$
Sub4	-1.097	0.248	$7 * 10^{-6}$

Table 1: Prior Parameters of the model. The fit parameters of the prior $p(v; c_0, c_1, c_2) \propto (|v| + c_1)^{c_0} + c_2$ of the Bayesian observer model for each individual subject.

Subject	$h(0.05)$	$h(0.075)$	$h(0.10)$	$h(0.20)$	$h(0.40)$	$h(0.50)$	$h(0.80)$
Sub1	0.035	0.027	0.028	0.019	0.016	0.013	0.012
Sub2	0.045	0.041	0.038	0.030	0.023	0.020	0.010
Sub3	0.061	0.045	0.041	0.029	0.017	0.019	0.026
Sub4	0.043	0.040	0.036	0.028	0.018	0.017	0.016

Table 2: Noise Parameters of the Model. The fit contrast-dependent noise parameters of the Bayesian observer model for each individual subject.