Behavioral/Cognitive

Predicting the Partition of Behavioral Variability in Speed Perception with Naturalistic Stimuli

Benjamin M. Chin[1] and Johannes Burge[1,2,3]

[1]Department of Psychology, [2]Neuroscience Graduate Group, and [3]Bioengineering Graduate Group, University of Pennsylvania, Philadelphia, Pennsylvania 19104

A core goal of visual neuroscience is to predict human perceptual performance from natural signals. Performance in any natural task can be limited by at least three sources of uncertainty: stimulus variability, internal noise, and suboptimal computations. Determining the relative importance of these factors has been a focus of interest for decades but requires methods for predicting the fundamental limits imposed by stimulus variability on sensory-perceptual precision. Most successes have been limited to simple stimuli and simple tasks. But perception science ultimately aims to understand how vision works with natural stimuli. Successes in this domain have proven elusive. Here, we develop a model of humans based on an image-computable (images in, estimates out) Bayesian ideal observer. Given biological constraints, the ideal optimally uses the statistics relating local intensity patterns in moving images to speed, specifying the fundamental limits imposed by natural stimuli. Next, we propose a theoretical link between two key decision-theoretic quantities that suggests how to experimentally disentangle the impacts of internal noise and deterministic suboptimal computations. In several interlocking discrimination experiments with three male observers, we confirm this link and determine the quantitative impact of each candidate performance-limiting factor. Human performance is near-exclusively limited by natural stimulus variability and internal noise, and humans use near-optimal computations to estimate speed from naturalistic image movies. The findings indicate that the partition of behavioral variability can be predicted from a principled analysis of natural images and scenes. The approach should be extendable to studies of neural variability with natural signals.

Key words: decision variable correlation; efficiency; motion energy; natural scene statistics; psychophysics; signal detection theory

Significance Statement

Accurate estimation of speed is critical for determining motion in the environment, but humans cannot perform this task without error. Different objects moving at the same speed cast different images on the eyes. This stimulus variability imposes fundamental external limits on the human ability to estimate speed. Predicting these limits has proven difficult. Here, by analyzing natural signals, we predict the quantitative impact of natural stimulus variability on human performance given biological constraints. With integrated experiments, we compare its impact to well-studied performance-limiting factors internal to the visual system. The results suggest that the deterministic computations humans perform are near optimal, and that behavioral responses to natural stimuli can be studied with the rigor and interpretability defining work with simpler stimuli.

Introduction

Human beings are adept at many fundamental sensory-perceptual tasks. A sufficiently difficult task, however, can reveal the limits of human performance. A principal aim of perception science and systems neuroscience is to determine the limits of performance, and then to determine the sources of those limits. Performance limits have been rigorously investigated with simple tasks and stimuli (Burgess et al., 1981; Pelli, 1985; Burgess and Colborne, 1988; Geisler, 1989; Dosher and Lu, 1998; Michel and Geisler, 2011; Abbey and Eckstein, 2014)

Ultimately, perception science aims to achieve a rigorous understanding of how vision works in the real world. In natural viewing, there exist at least three factors that limit performance: natural stimulus variability, suboptimal computations, and internal noise. Testing the relative importance of these sources requires two key ingredients: (1) an image-computable (images in, estimates out) ideal observer that specifies optimal performance in the task; and (2) experiments that can distinguish the behav-

Received Aug. 2, 2019; revised Nov. 12, 2019; accepted Nov. 17, 2019.

Author contributions: B.M.C. and J.B. designed research; B.M.C. and J.B. performed research; B.M.C. and J.B. analyzed data; B.M.C. and J.B. wrote the first draft of the paper; B.M.C. and J.B. edited the paper; B.M.C. and J.B. wrote the paper.

This work was supported by University of Pennsylvania startup funds to J.B.; and National Eye Institute and the Office of Behavioral and Social Sciences Research, National Institutes of Health Grant R01-EY028571 to J.B. We thank David Brainard for helpful discussions; and Josh Gold for providing comments on a draft version of the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Johannes Burge at jburge@sas.upenn.edu.

https://doi.org/10.1523/JNEUROSCI.1904-19.2019

ioral signatures of each factor. Here, we develop theoretical and empirical methods that can predict and diagnose the impact of each source in mid-level visual tasks with natural and naturalistic stimuli. We investigate the specific task of retinal speed estimation, a critical ability for estimating the motion of objects and the self through the environment.

When a pattern of light falls on the retina, millions of photoreceptors transmit information to the brain about the visual scene. This information is used to build stable representations of image and scene properties (i.e., latent variables) that are relevant for survival and reproduction, such as motion speed, 3D position, and object identity. The visual system successfully extracts these critical latent variables from local areas of natural images despite tremendous stimulus variability; infinitely many unique retinal images (i.e., light patterns) are consistent with each value of a given latent variable. Some image features that vary across different natural images are particularly informative for extracting the latent variable(s) of interest. These are the features that the visual system should encode. Many other image features carry no relevant information. These features should be ignored. (Stimulus variation unrelated to the latent variable is often referred to as "nuisance" variation.) Variation in both the relevant and irrelevant feature spaces can limit performance. But the impact of stimulus variability on performance is minimized only if all relevant features are encoded. Thus, stimulus variability can differentially impact performance depending on the quality of feature encoding.

Signal detection theory posits that sensory-perceptual performance is based on the value of a decision variable (Green and Swets, 1966). But signal detection theory does not specify how to obtain the decision variable from the stimulus. Image-computable observer models do (Adelson and Bergen, 1985; Simoncelli and Heeger, 1998; Schrater et al., 2000; Ziemba et al., 2016; Schütt and Wichmann, 2017; Fleming and Storrs, 2019). Image-computable ideal observer models specify how to optimally encode and process the most useful stimulus features (Burgess et al., 1981; Banks et al., 1987; Geisler, 1989; Burge and Geisler, 2011, 2012, 2014, 2015; Sebastian et al., 2017). Image-computable ideal observer models specify how pixels in the image should be transformed into estimates (or categorical decisions) that optimize performance in a particular task.

Ideal observers play an important role in the study of perceptual systems because they allow researchers to precisely ask, given the information available to a particular stage of processing, whether subsequent processing stages use that information as well as possible (Geisler, 1989). The explicit description of optimal processing provided by an image-computable ideal observer specifies how natural stimulus variability should propagate into the decision variable given biological constraints. Optimal processing minimizes stimulus-driven nuisance variation in the decision variable. Thus, stimulus variability and the optimal processing jointly set a fundamental limit on performance.

Human performance often tracks the pattern of ideal observer performance but rarely achieves the same absolute performance levels. It is common to attribute these discrepancies to noise, but discrepancies can also arise from systematically suboptimal computations. To what extent does each factor contribute?

Using complementary computational and experimental techniques, we answer this question for a speed discrimination task with naturalistic stimuli. We show that (1) natural stimulus variability equally impacts human and ideal performance, (2) the deterministic computations (encoding, pooling, decoding) performed by the human visual system are very nearly optimal, and (3) humans underperform the ideal near-exclusively because of stochastic internal sources of variability (e.g., late noise) and not because of a systematic misuse of the available stimulus information. The work demonstrates that, with appropriate experimental designs, image-computable ideal observer analysis can identify the reasons for human perceptual limits in visual tasks with natural and naturalistic stimuli.

Materials and Methods

Experimental design and statistical analyses. Three male human observers participated in the experiment: 2 were authors, and the third was naive to purposes of the experiment. All had normal or corrected-to-normal acuity. The research protocol was approved by the Institutional Review Board of the University of Pennsylvania and was in accordance with the Declaration of Helsinki. The study was not preregistered. All experiments were performed in MATLAB 2017a using Psychtoolbox version 3.0.12 (Brainard, 1997). Psychophysical data are presented for each individual human observer. Cumulative Gaussian fits of the psychometric functions were in good agreement with the raw data. Bootstrapped or Monte Carlo-simulated SEs or CIs are presented on all data points unless otherwise noted. Data will be made available upon reasonable request.

Equipment. Stimuli were presented on a ViewSonic G220fb 40.2 cm × 30.3 cm cathode ray tube monitor with 1280 × 1024 pixel resolution, and a refresh rate of 60 Hz. At the 92.5 cm viewing distance, the monitor subtended an FOV of 24.5 × 18.6° of visual angle. The display was linearized over 8 bits of gray level. The maximum luminance was 74 cd/m^2. The mean background gray level was set to 37 cd/m^2. The observer's head was stabilized with a chin-and-forehead rest.

Stimuli: detection experiment. Target stimuli in the detection experiment consisted of static, vertically oriented Gabor targets in cosine phase (3 and 4.5 cpd) with 1.5 octave bandwidths embedded in vertically oriented (1D) dynamic Gaussian noise that was uncorrelated in space and time. Targets subtended 1.0° of visual angle for a duration of 250 ms (15 frames at 60 Hz). Stimuli were windowed with a raised-cosine window in space and a flattop-raised-cosine window in time, exactly the same as the image movies in the speed discrimination experiment. The root-mean-squared (RMS) contrast of the target and the noise were varied independently according to the experimental design. To minimize target uncertainty, the target was presented to the subject, without noise every 10 trials.

For the detection experiment, a bit depth of more than 8 bits is required to accurately measure contrast detection thresholds. We achieved a bit depth of more than 10 bits using the LOBES video switcher (Li et al., 2003). The video switcher combines the blue channel and attenuated red channel outputs in the graphics card. Picking the right combination of blue and red channel outputs generates a precise grayscale luminance signal.

Procedure: detection experiment. Stimuli in the target detection experiment were presented using a two-interval forced choice (2IFC) procedure. On each trial, one interval contained a target plus noise, and the other interval contained noise only. The task was to select the interval containing the target. Feedback was provided. Psychometric functions were measured for each of four different RMS stimulus noise contrasts (0.00, 0.05, 0.10, 0.20) using the method of constant stimuli, with five different target contrasts per condition. Each observer completed 3200 trials in this experiment (4 noise levels × 5 target contrasts per noise level × 80 trials per target × 2 target frequencies). Each block contained 50 trials. To minimize observer uncertainty, trials were blocked by stimulus and noise contrast. The target stimulus was also presented at the beginning of each block, and then again every 10 trials, throughout the experiment.

In target detection tasks, stimulus (e.g., pixel) noise is under experimental control. Internal noise is not. Both noise types influence target detection thresholds. Target contrast power at threshold is a function of stimulus noise $C_T^2(\sigma_{pix}) \propto \sigma_{pix}^2 + \sigma_{internal}^2$ and is proportional to the sum of pixel and internal noise variances (Burgess et al., 1981); the constant of proportionality depends on the target. This fact can be leveraged to estimate the internal noise that limits detection performance. For exam-

ple, when stimulus noise and internal noise have equal variance, the squared detection threshold will be twice what it is when pixel noise is zero: $C_T^2(\sigma_{pix} = \sigma_{internal}) = 2C_T^2(\sigma_{pix} = 0)$. The amount of stimulus noise required to double thresholds is known as the equivalent input noise. The amount of internal noise that limits performance in a target detection task can therefore be estimated from the pattern of detection thresholds. The estimate of equivalent input noise from the detection experiment sets an upper bound on the amount of early noise in the human visual system (see Results).

Stimuli: speed discrimination experiment. Natural image movies were created by texture-mapping randomly selected patches of calibrated natural images onto planar surfaces, and then moving the surfaces behind a stationary 1.0° aperture. The movies were restricted to one dimension of space by vertically averaging each frame of the movie (Burge and Geisler, 2015). Each movie subtended 1.0° of visual angle. Movie duration was 250 ms (15 frames at 60 Hz). All stimuli were windowed with a raised-cosine window in space and a flattop-raised-cosine window in time. The transition regions at the beginning and end of the time window each consisted of four frames; the flattop of the window in time consisted of seven frames. Contrast was computed under the space-time window. To prevent aliasing, stimuli were low-pass filtered in space and time before presentation (Gaussian filter in frequency domain with $\sigma_{space} = 4$ cpd, $\sigma_{time} = 30$ Hz). No aliasing was visible. Training and test sets of naturalistic stimulus movies were generated. The training set had 10,500 unique stimuli (500 stimuli × 21 speeds); the test set had 61,000 unique stimuli (1000 stimuli × 61 speeds). Training stimuli were used to develop the ideal observer (see below). Test stimuli were used to evaluate the ideal and human observers in the speed discrimination experiment.

All stimuli were set to have the same mean luminance as the background and had an RMS contrast of 0.14 (equivalent to 0.20 Michelson contrast for sinewave stimuli), the modal contrast of the stimulus ensemble. The RMS contrast is given by the following:

$$C_{RMS} = \sqrt{\frac{\sum_{\mathbf{x}} \mathbf{c}^2(\mathbf{x})\mathbf{w}(\mathbf{x})}{\sum_{\mathbf{x}} \mathbf{w}(\mathbf{x})}} \quad (1)$$

where $\mathbf{c}(\mathbf{x})$ is a Weber contrast image movie, $\mathbf{w}(\mathbf{x})$ is the space-time window, and $\mathbf{x} = \{x, y, t\}$ is a vector of space-time positions. Stimuli were contrast fixed because contrast is known to affect speed percepts, and our focus was on how differences in Weber contrast patterns between stimuli impact performance rather than on how differences in overall contrast impact performance, which have already been intensively studied (Thompson, 1982; Weiss et al., 2002).

The short (i.e., 250 ms) presentation duration was chosen to approximate the typical duration of a human fixation, and to reduce the possibility that large eye movements would occur while the stimulus was onscreen. For stimuli with speeds and contrasts similar to those used in this experiment, the latencies of smooth pursuit eye movements tend to be 140–200 ms (Spering et al., 2005). Saccadic latencies tend to be longer than pursuit latencies.

Procedure: speed discrimination experiment. For the speed discrimination task, data were collected using a 2IFC procedure. On each trial, a standard and a comparison image movie were presented in pseudo-random order (see below). The task was to choose the interval with the movie having the faster speed. Human observers indicated their choice via a key press. The key press also initiated the next trial. Feedback was given. A high tone indicated a correct response; a low tone indicated an incorrect response. Experimental sessions were blocked by absolute standard speed. In the same block, for example, data were collected at the −5 and 5°/s standard speeds. Movies always drifted in the same direction within a trial, but directions were mixed within a block. An equal number of left- and right-drifting movies were presented in the same block to reduce the potential effects of adaptation.

In each pass of the experiment (see below), psychometric data were measured for each of 10 standard speeds (±5, ±4, ±3, ±2, ±1°/s) using the method of constant stimuli. Seven comparison speeds were presented for each standard speed, spanning a range centered on each standard speed. Thus, across the entire experiment, observers viewed stimuli with speeds ranging from 0.25 to 8.00°/s. Each standard-comparison speed combination was presented 50 times each for a total of 3500 trials (2 directions × 5 standard speeds × 7 comparison speeds × 50 trials).

The exact same naturalistic movie was never presented twice within a pass of the experiment. Rather, movies were randomly sampled without replacement from a test set of 1000 naturalistic movies at each speed. For each standard speed, 350 "standard speed movies" were randomly selected. Similarly, for each of the seven comparison speeds corresponding to that standard, 50 "comparison speed movies" were randomly selected. Standard and comparison speed movies were then randomly paired together. This stimulus selection procedure was used to ensure that the stimuli used in the psychophysical experiment had approximately the same statistical variation as the stimuli that were used to train and test the ideal observer model. Assuming the stimulus sets are representative and sufficiently large, the stimuli presented in the experiment are likely to be representative of natural signals.

Ideal observer for speed estimation. As signals proceed through the visual system, neural states become more selective for properties of the environment, and more invariant to irrelevant features of the retinal images. The ideal observer for speed estimation computes the Bayes' optimal speed estimate from the posterior probability distribution over speed $p(X|\mathbf{R})$ given the responses $\mathbf{R}$ to a stimulus of a small population of optimal space-time receptive fields (Burge and Geisler, 2015). The receptive fields are assumed to be no larger than the stimulus (i.e., 1.0°) and to have a temporal integration period no longer than the stimulus duration (i.e., 250 ms). No restrictions were placed on the smallest size and shortest integration period of the receptive fields. The receptive fields operate on captured retinal images that include the constraints of the early visual system. The optics of the eye, the spatial sampling, wavelength sensitivity, and temporal integration of the photoreceptors, and response normalization all constrain and shape the information available for further processing. Each natural image movie was convolved with a point-spread function consistent with a 2 mm pupil, a typical size on a bright sunny day (Wyszecki and Stiles, 1982), and the chromatic aberrations of the human eye (Thibos et al., 1992). The temporal integration time of the photoreceptors was ~30 ms, consistent with direct neurophysiological measurements (Schneeweis and Schnapf, 1995). Receptive field responses were normalized consistent with standard practice (Albrecht and Geisler, 1991; Heeger, 1992; Carandini and Heeger, 2011; Burge and Geisler, 2015; Jaini and Burge, 2017; Sebastian et al., 2017; Iyer and Burge, 2019). Given the constraints imposed by natural stimulus variability and the front-end properties of the early visual system, the space-time receptive fields and the subsequent computations for decoding the speed must be optimal in order for the estimates to be considered optimal. The most useful stimulus features and the computations that optimally pool them are jointly dictated by the task and the stimuli. The receptive fields that encode the most optimal stimulus features for the task are determined via a recently developed technique called Accuracy Maximization Analysis (Geisler et al., 2009; Burge and Jaini, 2017; Jaini and Burge, 2017). Accuracy Maximization Analysis requires a labeled training set, a model of receptive field response, and a cost function but requires no parametric assumptions about the shape of the receptive fields. When the training set is representative and sufficiently large, as it is here, the learned receptive fields support equivalent performance on test and training stimulus sets.

The joint response of the set of receptive fields to each stimulus is given by $\mathbf{R} = \mathbf{f}^T(\mathbf{c} + \mathbf{n})/\|\mathbf{c} + \mathbf{n}\|$ where $\mathbf{f}$ is the set of filters, $\mathbf{c}$ is the contrast stimulus, and $\mathbf{n}$ is a sample of early noise. The optimal computations for pooling the responses of the receptive fields are specified by how the receptive field responses are distributed. The conditional receptive field responses $p(\mathbf{R}|X_k) = gauss(\mathbf{R}; \mathbf{0}, \Sigma_k)$ are mean zero and jointly Gaussian after response normalization (Burge and Geisler, 2015; Jaini and Burge, 2017). For any observed response $\mathbf{R}$, the computations that specify the likelihood $L(X_u; \mathbf{R}) = p(\mathbf{R}|X_u)$ that an observed response was elicited by a stimulus moving with speed X_u is obtained by evaluating the response in the response distribution corresponding to that speed. The responses must therefore be pooled in a weighted quadratic sum, with weights $\mathbf{w}_u$ that are given by simple functions of the covariance matrices Σ_u (Burge

and Geisler, 2015). A neuron that performs these quadratic computations outputs a response $R_u^L \propto \exp[Q_u(\mathbf{R})] = L(X_u; \mathbf{R})$ that is proportional to the likelihood that a stimulus moving at speed X_u elicited the response **R**. After response (e.g., contrast) normalization (Albrecht and Geisler, 1991; Heeger, 1992; Carandini and Heeger, 2011; Sebastian et al., 2017; Iyer and Burge, 2019), these likelihood neurons instantiate an energy-model-like hierarchical LNLN (linear, nonlinear, etc.) cascade (Adelson and Bergen, 1985; Jaini and Burge, 2017). Thus, the computations that yield likelihood neurons can be thought of as a recipe, grounded in natural image and scene statistics, for how to construct speed-tuned neurons that are maximally selective for speed and maximally invariant to natural stimulus (i.e., nuisance) variability. Similar computations yield selective invariant tuning for latent variables, such as defocus blur, binocular disparity, and 3D motion (Burge and Geisler, 2011, 2012, 2014, 2015).

To obtain the posterior probability of each speed, the likelihood must be weighted by the prior $p(X_u)$ and normalized by the weighted sum of likelihoods $\Sigma_v L(X_v;\mathbf{R})p(X_v)$. Finally, the optimal estimate must be "read out" from the posterior probability distribution. In the case of the 0, 1 cost function (i.e., L0 norm) the optimal estimate $\hat{X}_{opt} = \underset{X}{\operatorname{argmax}}\, p(X|\mathbf{R})$ is the posterior maximum. If the prior probability distribution is flat, which it is in the training and test sets, the optimal estimate is the latent variable value that corresponds to the maximum of the likelihood function (i.e., the maximum of the population response over the likelihood neurons).

Ideal, degraded, and human decision variables. The ideal decision variable for the task of speed discrimination is obtained by subtracting the optimal speed estimates corresponding to the comparison and standard stimuli as follows:

$$D_{ideal} = \hat{X}_{ideal}^{cmp} - \hat{X}_{ideal}^{std} \tag{2}$$

where $\hat{X}_{ideal}^{std}$ and $\hat{X}_{ideal}^{cmp}$ are the ideal observer estimates for the standard and comparison stimuli, respectively. The total variance of the ideal observer decision variable is $2\sigma_{ideal}^2$, where σ_{ideal}^2 is the variance of the ideal observer estimates across stimuli at a given speed. If the decision variable is greater than zero, the ideal observer responds that the comparison stimulus was faster. If the decision variable is less than zero, the ideal observer responds that the comparison stimulus was slower. Degraded observer decision variables are similarly obtained, except that the degraded observer estimates are obtained by reading out the responses of suboptimal receptive fields as well as possible.

The human decision variable is a noisy version of the ideal decision variable, under the hypothesis that human inefficiency is due only to internal sources of variability (e.g., noise). Specifically,

$$D_{human} = D_{ideal} + W \tag{3}$$

where $W \sim N(0, 2\sigma_I^2)$ is a sample of zero mean Gaussian noise, which corresponds to adding noise with variance σ_I^2 to the comparison and standard stimulus speed estimates.

Double-pass experiment. A double-pass experiment requires that each observer perform all (or a subset) of the unique trials in an experiment twice. In our experiment, each trial was uniquely identified by its standard and comparison movies. An observer completed the first pass by completing each unique trial once over 20 blocks consisting of 175 trials each. The standard speed was always constant within a block. Blocks were counterbalanced. The observer completed the second pass by completing each unique trial again over another 10 blocks. Before collecting data in the main experiment, each human observer completed multiple practice sessions to ensure that perceptual learning had stabilized. Analysis of the practice data showed no significant learning effects. Stimuli presented in practice sessions were not presented in the main experiment.

Estimating decision variable correlation. Human decision variable correlation is estimated via maximum likelihood methods from the pattern of human response agreement in the double-pass experiment. The maximum likelihood parameter estimates are those that maximize the log-likelihood of the data under a model:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, LL \tag{4}$$

where θ is a vector of model parameters describing the decision variable distribution and observer criteria across both passes of the double-pass experiment. The log-likelihood of the double-pass response data is given by the following:

$$LL = N^{--} \ln p^{--}(\theta) + N^{-+} \ln p^{-+}(\theta) + N^{+-} \ln p^{+-}(\theta) + N^{++} \ln p^{++}(\theta) \tag{5}$$

where N^{--} and N^{++} are the number of times that the observer chose standard on both passes or the comparison on both passes, respectively, and N^{-+} and N^{+-} are the number of times that the observer chose the standard on first pass and the comparison on the second and vice versa. The likelihoods of observing those samples are given by the following:

$$p^{--} = \int_{-\infty}^{c_1}\int_{-\infty}^{c_2} gauss(\mathbf{D};\mathbf{u},\Sigma) \tag{6a}$$

$$p^{-+} = \int_{-\infty}^{c_1}\int_{c_2}^{\infty} gauss(\mathbf{D};\mathbf{u},\Sigma) \tag{6b}$$

$$p^{+-} = \int_{c_1}^{\infty}\int_{-\infty}^{c_2} gauss(\mathbf{D};\mathbf{u},\Sigma) \tag{6c}$$

$$p^{++} = \int_{c_1}^{\infty}\int_{c_2}^{\infty} gauss(\mathbf{D};\mathbf{u},\Sigma) \tag{6d}$$

where **D** is the joint decision variable across passes with mean **u** and covariance Σ, and c_1 and c_2 are the observer criteria on passes 1 and 2. The mean decision variable values are set equal to the speed difference $\mu_1 = \mu_2 = X_{cmp} - X_{std}$ between the standard and comparison stimuli in each condition.

In practice, and without loss of generality, we estimate the decision variable correlation using normalized decision variables **Z**. The parameter vector for maximizing the likelihood of the normalized decision variables is $\theta = \{\rho^*, \mu_1^*, \mu_2^*, c_1^*, c_2^*\}$, where $*$ indicates that the parameter is associated with the normalized variable, and ρ is the correlation specified by the covariance Σ. The integrals in Equations 6a-d can be equivalently expressed with limits of integration $c^* = c/\sigma_{human}$ and integrand $gauss(\mathbf{Z; Mu, M\Sigma M^T})$ with normalized mean and normalized covariance as follows:

$$\mathbf{Mu} = \left[\underbrace{\mu_1^*}_{\mu_1/\sigma_{human}} \quad \underbrace{\mu_2^*}_{\mu_2/\sigma_{human}}\right]^T \tag{7a}$$

$$\mathbf{M\Sigma M}^T = \begin{bmatrix} 1 & \rho^* \\ \rho^* & 1 \end{bmatrix} \tag{7b}$$

where the normalizing matrix is $\mathbf{M} = \begin{bmatrix} 1/\sigma_{human} & 0 \\ 0 & 1/\sigma_{human} \end{bmatrix}$, and where σ_{human} is the SD of the human estimates. Normalizing the variables has the practical advantage that it converts the covariance matrix to a correlation matrix, so that it can be fully characterized with a single parameter: decision variable correlation. It also sets the normalized means equal to sensitivity d'. We fix the normalized means $\mu_1^* = \mu_2^* = d'_{human}$ to the human sensitivity measured in the discrimination experiment. We also fix the normalized criteria to $c_1^* = c_2^* = 0.0$, which is justified both by the data and the experimental design. These choices reduce the number of parameters to be estimated from five to one.

Efficiency and early noise. Efficiency quantifies the degree to which human performance falls short of ideal performance. The exact expression for efficiency is given by the following:

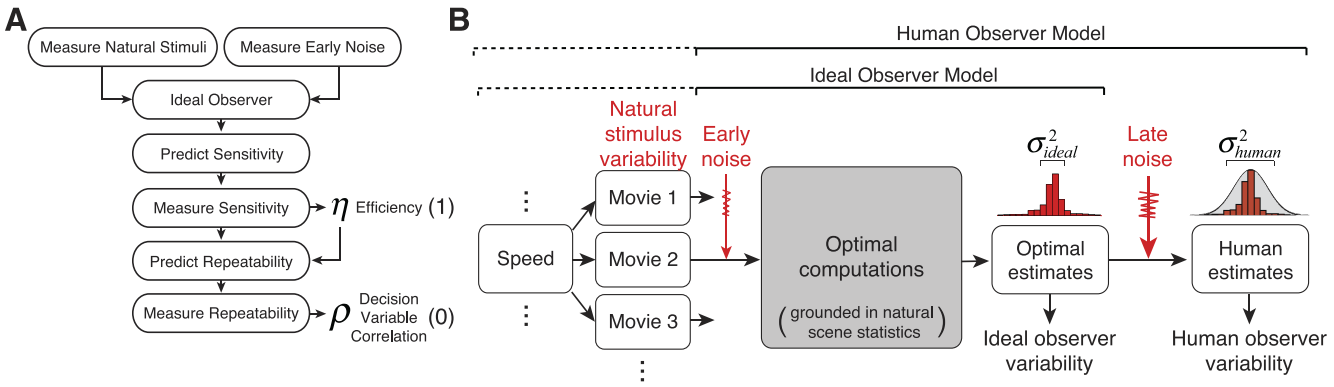

Figure 1. Plan for manuscript and ideal observer. ***A***, Plan for the manuscript. First, we measure natural stimuli and early noise to constrain an ideal observer for speed estimation. Next, we run an experiment and fit the efficiency of each human observer (one free parameter) by comparing human to ideal sensitivity. Finally, we run a double-pass experiment and show that efficiency predicts human response repeatability and decision variable correlation (zero free parameters). ***B***, Ideal observer. Speed (i.e., the latent variable) can take on one of many values. Many different image movies share the same speed. The ideal observer is defined by the optimal computations (encoding, pooling, decoding) for estimating speed with natural stimuli. The optimal computations are grounded in natural scene statistics (gray box). For each unique movie, the ideal observer outputs a point estimate of speed. The ideal observer's estimates vary across movies primarily because of natural stimulus variability, variability that is external to the observer. The degraded ideal observer is matched to overall human performance by adding late noise.

$$\eta = \left(\frac{d'_{human}}{d'_{ideal}}\right)^2 = \frac{\sigma^2_{ideal}}{\sigma^2_{human}} = \frac{\sigma^2_E + \sigma^2_{I,\,early}}{\sigma^2_{human}} \qquad (8)$$

where σ^2_{ideal} and σ^2_{human} are the variances of the ideal and human speed estimates, and σ^2_E and $\sigma^2_{I,\,early}$ are the stimulus-driven and early-noise-driven variances in the ideal speed estimates. The early-noise-driven variance in the estimates, and consequently in the decision variable, is distinct from early noise itself, which is defined in the domain of the image pixels instead of the decision variable. This is analogous to how the stimulus-driven variance in the decision variable is distinct from stimulus variability. Stimulus variability, like early noise, is defined in the domain of the image pixels and is non-zero in any set of nonidentical stimuli having the same value of the latent variable. We computed efficiency using the exact expression in Equation 8 and the approximate equality presented in the main text, which assumes that the impact of early noise on the ideal decision variable is negligible (see Results). We found that, because the maximum possible amount of early noise in the system is small (i.e., the upper bound on early noise established by the detection experiment is low), both the exact and the approximate expressions yield similar estimates of efficiency.

Results

The impact of natural stimulus variability, internal noise, and suboptimal computations can only be distinguished by combining an ideal observer with appropriate behavioral experiments. We examine how these factors impact local motion estimation, a sensory-perceptual ability that is critical for appropriate interaction with the environment (Burge et al., 2019). The plan for the manuscript is diagrammed in Figure 1*A*. First, we develop an image-computable ideal observer model of retinal speed estimation that is constrained by measurements of natural stimulus variability and early noise. Then we compare human to ideal performance with matched stimuli in two main experiments with matched stimuli. The first main experiment shows that humans track the predictions of the ideal but are consistently less sensitive: one free parameter (efficiency) accounts for the gap between human and ideal performance. We hypothesize that human inefficiency is due to stochastic internal sources of variability (e.g., late noise), and not deterministic suboptimal computations. This hypothesis predicts that natural stimulus variability should equally limit human and ideal observers. The second main experiment tests this hypothesis. Human observers viewed thousands of trials with naturalistic stimuli in which each unique trial was presented twice. In this paradigm, the repeatability of responses reveals the respective roles of stimulus- and noise-driven variability. If our hypothesis about the source of human inefficiency is correct, efficiency should predict response repeatability with zero additional free parameters. These predictions are confirmed by the experimental data.

An image-computable ideal observer for estimating retinal image speed from local regions of natural images is shown in Figure 1*B*. Given a set of stimuli, it uses the optimal computations (encoding receptive fields, pooling, decoding) for estimating speed from natural image movies (Burge and Geisler, 2015). The ideal observer thus provides a principled benchmark against which to compare human performance. The tradition in ideal observer analysis is to constrain the ideal observer by stimulus and physiological factors that can be well characterized and are known to limit the information available for subsequent processing (Geisler, 1989). Natural stimulus variability and early measurement noise are two such factors (Fig. 1*B*, red text). The optimal computations govern how these factors propagate into and determine the variance of the ideal decision variable (Fig. 1*B*). The ideal decision variable controls ideal observer performance.

Human performance is typically worse than ideal performance. To account for this performance gap, other factors must be considered. We consider suboptimal computations and internal noise, both of which have the potential to increase the variance of the human decision variable relative to the ideal. Suboptimal computations are deterministic and reflect a systematic misuse of the available stimulus information. Internal noise is random and is uncorrelated with individual stimuli; although we model it as occurring at the level of the decision variable (Fig. 1*B*), our methods do not distinguish between different stochastic internal sources of variability (see Discussion). To simultaneously determine the impact of all three factors (natural stimulus variability, suboptimal computations, and internal noise), the ideal observer must be paired with an appropriate psychophysical experiment in which each factor has a distinct behavioral signature. We perform this experiment and determine the relative importance of each factor. We find that natural stimulus variability and late noise are the primary factors limiting human performance. The impact of suboptimal computations is negligible.

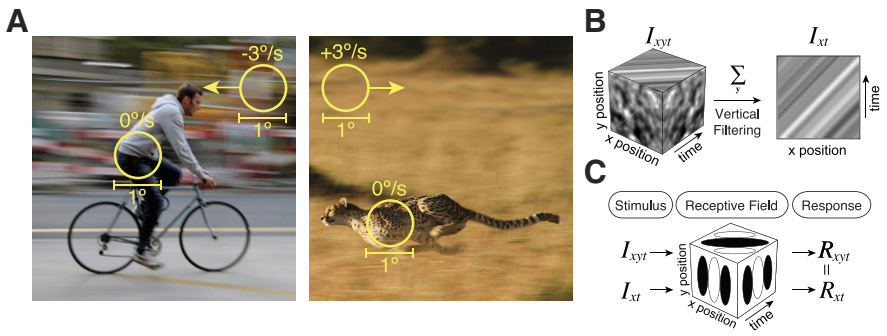

Figure 2. Naturalistic image movies and preprocessing. ***A***, Naturalistic image movies were obtained by drifting photographs of natural scenes at known speeds behind 1° apertures for 250 ms. Rotating the eye in its socket (e.g., tracking an object) creates the same pattern of motion in the stationary background. Optical properties of the eye and the temporal integration of the photoreceptors were also modeled. ***B***, Full space-time image movies (*Ixyt*) and vertically filtered space-time image movies (*Ixt*). Moving images can be represented as oriented signals in space-time. ***C***, Vertically oriented receptive fields respond identically to full space-time movies and vertically filtered movies.

Measuring natural stimuli

A fundamental problem of perception is that multiple proximal stimuli can arise from the same distal cause. This stimulus variability is an important source of uncertainty that limits human and ideal speed discrimination performance. To measure natural stimulus variability, we photographed a large number of natural scenes (Burge and Geisler, 2011, 2015), and then drifted those photographs at known speeds behind a 1° aperture, approximately the size of foveal receptive fields in early visual cortex (Gattass et al., 1981, 1988). This procedure generates motion signals that are equivalent to those obtained by rotating the eye during smooth tracking of a target (Spering et al., 2005; Osborne et al., 2007) (Fig. 2*A*). The sampled set of stimuli approximates, but almost certainly underestimates, the variability present in the natural stimulus ensemble; looming and discontinuous motions, for example, are not represented in our training set (Schrater et al., 2001; Nitzany and Victor, 2014). Thus, the forthcoming estimates of the impact of natural stimulus variability on ideal and human performance are likely to underestimate the impact of stimulus variability on human performance in natural viewing.

Movies drifted leftward or rightward with speeds ranging between 0.25 to 8.0°/s. Movies were presented for 250 ms, the approximate duration of a typical human fixation. The sampling procedure yielded tens of thousands of unique stimuli (i.e., image movies) at dozens of unique speeds. Image movies were then filtered so that only vertical orientations were present; that is, the stimuli were vertically averaged (i.e., *xt*) versions of full space-time (i.e., *xyt*) movies (Fig. 2*B*). Vertical averaging reduces stimulus complexity, but the resulting stimuli are still substantially more realistic than classic motion stimuli, such as drifting sinewaves. Furthermore, vertically oriented receptive fields respond identically to vertically averaged and original movies (Fig. 2*C*). Thus, in an individual orientation column, the filtered movies should generate the same response statistics as the full space-time movies (Burge and Geisler, 2015; Jaini and Burge, 2017). Finally, the contrasts of the vertically averaged stimuli were fixed to the modal contrast in natural scenes (see Discussion). Thus, our stimuli represent a compromise between simple and real-world stimuli, allowing us to run experiments with more natural stimuli without sacrificing quantitative rigor and interpretability. Our analysis should be generalizable to full space-time movies with more realistic forms of motion.

Measuring early noise

All measurement devices are corrupted by measurement noise. The human visual system is no exception. Early measurement noise occurs at the level of the retinal image and places a fundamental limit on how well targets can be detected. Possible sources of early noise include the Poisson variability of light itself and the stochastic nature of the photoreceptor and ganglion cell responses (Hecht et al., 1942). The ideal observer for speed discrimination should be constrained by the same early noise as the human observer if it is to provide an accurate indication of the theoretically achievable human performance limits (Fig. 1*A*).

Human observers performed a target detection task using the equivalent input noise paradigm (Burgess et al., 1981; Pelli, 1985). The task was to detect a known stationary target embedded in dynamic Gaussian white noise. On each trial, human observers viewed two stimuli in rapid succession and tried to identify the stimulus containing the target (Fig. 3*A*,*B*). The time course of stimulus presentation was identical to the forthcoming speed discrimination experiment. Figure 3*C* shows psychometric functions for target detection in 1 human observer as a function of target contrast. Each function corresponds to a different noise contrast. Detection thresholds, which are the target contrasts required to identify the target interval 76% of the time (i.e., d' of 1.0 in a 2IFC task), are shown for two different targets (3.0 and 4.5 cpd) in Figure 3*D*. Consistent with previous studies, contrast power at threshold increases linearly with pixel noise (Burgess et al., 1981; Pelli, 1985). Figure 3*E* shows the same data plotted on logarithmic axes, a common convention in the literature. There are two critical points on this function. The first is its value when pixel noise equals zero, where detection performance is limited only by internal noise. The second is at double the contrast power of the first point: the so-called "knee" of the function, where the pixel noise equals the internal noise. This level of pixel noise is known as the equivalent input noise. The knee of the function, and thus the estimate of equivalent input noise, is robust to whether or not the observer is using a detector (e.g., receptive field) that is optimal for detecting the target.

The equivalent input noise was estimated separately for each target type and human observer. Estimates were consistent across target types and were thus averaged. Noise estimates for the first, second, and third human observers are 2.5%, 2.3%, and 2.9%, respectively (Fig. 3*E*). These values are in line with previous reports (Burgess et al., 1981; Pelli, 1985; Williams, 1985).

The estimates of equivalent input noise may reflect the exact amount of early measurement noise alone (Pelli, 1991). The estimates of equivalent input noise may also reflect the combined effect of early measurement noise and noise arising at later processing (e.g., decision) stages. Regardless of which possibility is correct, the target detection experiment provides an upper bound on the amount of early noise in the human visual system. The ideal observer used in the main text is limited by early noise at this upper bound. Because the upper bound is small, early noise only weakly impacts ideal observer performance (see below).

Ideal observer

An ideal observer performs a task optimally, making the best possible use of the available information given stimulus variabil-

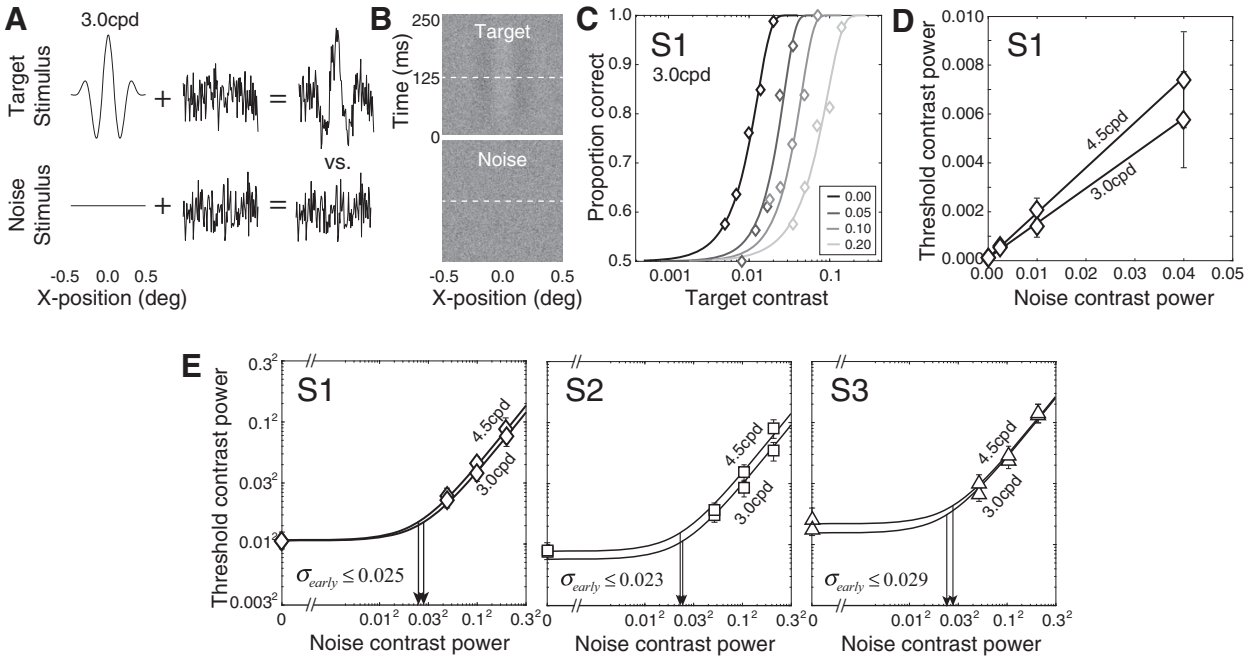

Figure 3. Measuring early noise with a target detection experiment. ***A***, Stimulus construction. On each interval, the stimulus was either a stationary target Gabor stimulus or a middle gray field corrupted by dynamic noise. ***B***, On each trial, the task was to report which of two intervals contained the target stimulus. ***C***, Psychometric functions from 1 human observer (S1) for detecting a 3 cpd target, in noise having different RMS contrasts (0.00, 0.05, 0.10, 0.20). ***D***, Threshold target contrast power for the same human observer. Thresholds increase linearly with noise contrast power. Error bars indicate 95% bootstrapped CIs; many error bars are smaller than the symbols. ***E***, Target contrast power at detection threshold plotted on a log-log axis (same data as ***D***) for all three observers. Arrows indicate the estimate of equivalent input noise.

ity and specified biological constraints. In addition to natural stimulus variability and early noise (Figs. 2, 3), we model the optics of the eye (Wyszecki and Stiles, 1982; Thibos et al., 1992), the temporal integration of photoreceptors (Schneeweis and Schnapf, 1995), and the linear filtering (Hubel and Wiesel, 1962) and response normalization (Albrecht and Geisler, 1991; Heeger, 1992; Carandini and Heeger, 2011) of cortical receptive fields. These are all well-established features of early visual processing and determine the information available for subsequent processing.

Assuming that the relevant factors have been accurately modeled, ideal observers provide principled benchmarks against which to compare human performance. Given the information available to a particular stage of processing, ideal observers allow the researcher to ask whether subsequent processing stages use that information as well as possible. Humans often track the pattern but fail to achieve the absolute limits of ideal performance. As a consequence, ideal observers often serve as principled starting points for determining additional unknown factors that cause humans to fall short of theoretically achievable performance limits.

Developing an ideal observer with natural stimuli is challenging because it is unclear *a priori* which stimulus features are most useful for the task. We find the optimal receptive fields for speed estimation using a recently developed Bayesian statistical learning method called Accuracy Maximization Analysis (Geisler et al., 2009; Burge and Jaini, 2017; Jaini and Burge, 2017). Given a stimulus set, the method learns the receptive fields that encode the most useful stimulus features for the task (Fig. 4*A*). Once the optimal features are determined, the next step is to determine how to optimally pool and decode the responses $\boldsymbol{R} = [R_1, R_2, \ldots, R_n]$ of those receptive fields where n is the total number of receptive fields. Eight receptive fields capture essentially all of the useful stimulus information; additional receptive fields provide negligible improvements in performance (Burge and Geisler, 2015).

The optimal pooling rules are specified by the joint statistics relating the latent variable and the receptive field responses (Bishop, 2006; Jaini and Burge, 2017). With appropriate response normalization, the responses across stimuli for each speed are conditionally Gaussian (Lyu and Simoncelli, 2009; Burge and Geisler, 2015; Sebastian et al., 2017; Iyer and Burge, 2019) (Fig. 4*B*). To obtain the likelihood of a particular speed, the Gaussian response statistics require that the receptive field responses to a given stimulus be pooled via weighted quadratic summation (Fig. 4*C*). The computations for computing the likelihood thus instantiate an enhanced version of the motion-energy model, indicating that energy-model-like computations are the normative computations supporting speed estimation with natural stimuli (Adelson and Bergen, 1985; Jaini and Burge, 2017). The speed tuning curves of hypothetical neurons implementing these computations are approximately log-Gaussian, similar to the approximately log-Gaussian speed tuning curves of neurons in area MT (Nover et al., 2005) (Fig. 4*D*). Finally, an appropriate readout of the population response of these hypothetical neurons is equivalent to decoding the optimal estimate from the posterior probability distribution $p(X|\mathbf{R})$ over speed (Fig. 4*E*,*F*). If a 0, 1 cost function is assumed, the latent variable value corresponding to the maximum of the posterior is the optimal estimate. We have previously verified that reasonable changes to the prior and cost function do not appreciably alter the optimal receptive fields, pooling rules, or performance (Burge and Jaini, 2017). This approach provides a recipe for how to construct neurons that are highly invariant to nuisance stimulus variability and tightly tuned to speed. It also provides a normative justification, grounded in natural scene statistics, for descriptive models proposed to account for response properties of neurons in cortex (Adelson and Bergen, 1985; Simoncelli and Heeger, 1998; Perrone and Thiele, 2001; Nover et al., 2005; Rust et al., 2006; Jaini and Burge, 2017).

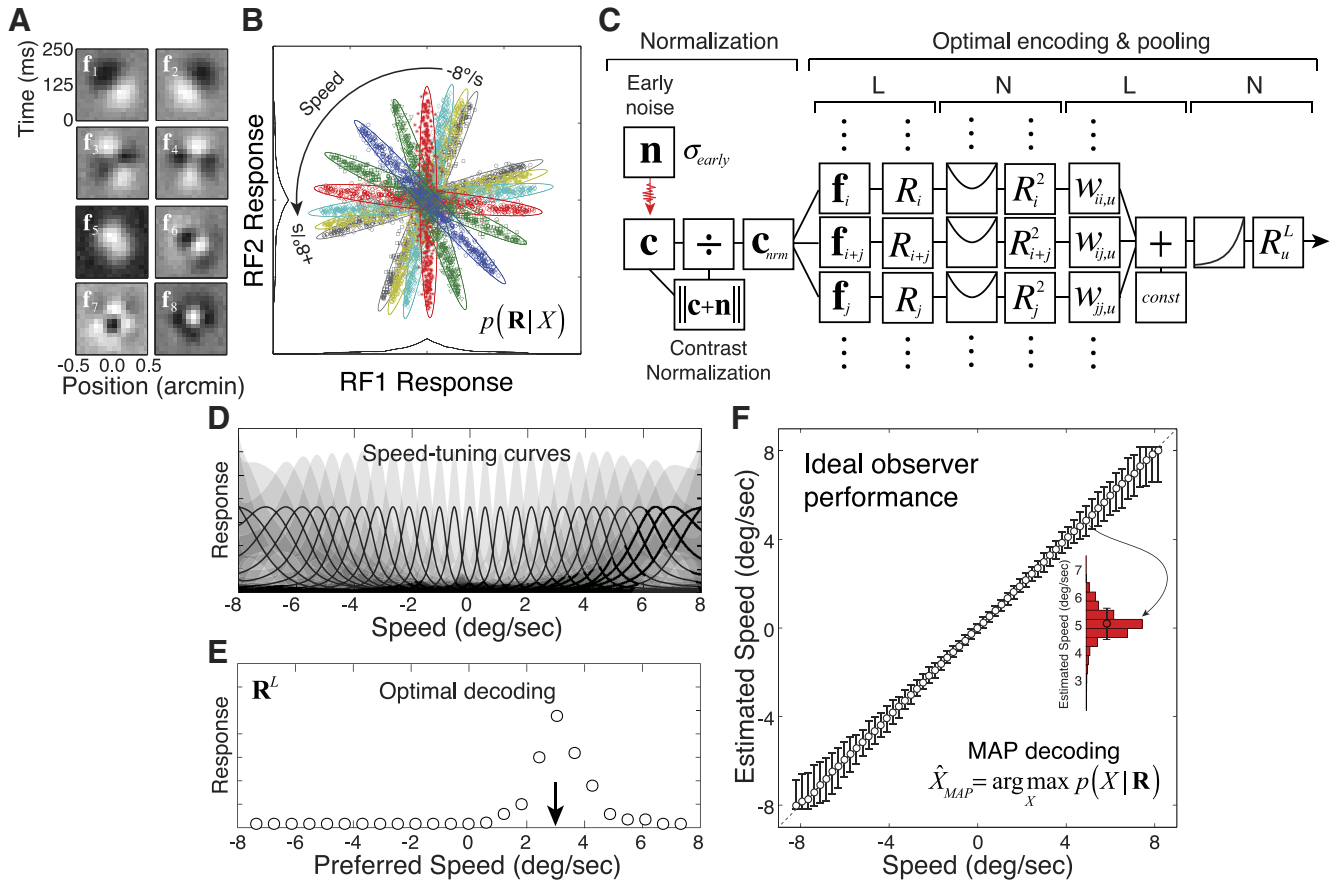

Figure 4. Ideal observer receptive fields (RFs), response distributions, computations, and estimates. ***A***, Optimal space-time RFs for speed estimation given the naturalistic stimulus set and biological constraints. ***B***, RF response distributions for RFs 1 and 2, conditioned on the speed of the image movie (colors). Each symbol represents the joint response to an individual movie. The variability of responses for each speed (color) is due to natural stimulus variability; that is, it is the nuisance stimulus variability in the feature space defined by the optimal RFs. ***C***, Computations of a hypothetical neuron implementing optimal encoding and pooling. Each noisy, contrast-normalized stimulus is processed by the optimal RFs. The responses of these RFs are pooled in a weighted quadratic sum. The weights are determined by the response covariance in ***B*** corresponding to the neuron's preferred speed. The response of this hypothetical neuron represents the likelihood that a given stimulus had its preferred speed. The optimal pooling rules thus represent an LNLN (linear, nonlinear, etc.) cascade. ***D***, Speed-tuning curves of hypothetical neurons implementing optimal encoding and pooling, whose responses represent the likelihood of each speed given a stimulus. The speed-tuning curve $\bar{R}^L(X_u)$ is the average likelihood across stimuli at each of many different speeds. Shaded regions represent ± SD CIs on response. This response variability is due to natural stimulus variability. ***E***, An arbitrary stimulus creates a population response $\mathbf{R}^L$ over hypothetical speed-tuned neurons. Optimal decoding yields the optimal estimate. ***F***, Ideal observer estimates. The optimal estimate is read out from the population of hypothetical speed-tuned neurons in ***E***, and is equivalent to reading out the posterior probability distribution $p(X|\mathbf{R})$ over speed. The variance of ideal observer speed estimates (histogram) is dominated by stimulus-driven variance.

The factors thus far described in the paper — stimulus variability and early noise, biological constraints, and the optimal computations (encoding, pooling, decoding) — all impact ideal performance in our task. Given a particular stimulus set, the only factor subject to some uncertainty is the precise amount of early noise. However, within the bound set by the detection experiment (Fig. 3), different amounts of early noise have only a minor effect on ideal performance (see below). Thus, estimates of ideal performance are set overwhelmingly by stimulus variability.

Measuring efficiency

The ideal observer benchmarks how well humans use the stimulus information available for the task. Efficiency quantifies how human sensitivity d'_{human} compares with ideal observer sensitivity d'_{ideal} and is given by the following:

$$\eta = \left(\frac{d'_{human}}{d'_{ideal}}\right)^2 = \frac{\sigma^2_{ideal}}{\sigma^2_{human}} \cong \frac{\sigma^2_E}{\sigma^2_{human}} \qquad (9)$$

where σ^2_{human} is the total variance of the human decision variable, σ^2_{ideal} is the total variance of the ideal decision variable, and σ^2_E is the stimulus-driven component of the ideal decision variable. The third approximate equality in Equation 9 assumes that stimulus-driven variability equals ideal observer variability because the impact of early noise is bounded to be small (compare Fig. 3).

To measure human sensitivity, we ran a 2IFC speed discrimination experiment. On each trial, human observers viewed two moving stimuli in rapid succession, and indicated which stimulus was moving more quickly (Fig. 5*A*). This design is similar to classic psychophysical experiments with one critical difference. Rather than presenting the same (or very similar) stimuli in each condition hundreds of times, we present hundreds of unique stimuli one time each. This stimulus variability jointly limits human and ideal performance. Human sensitivity is computed using standard expressions from signal detection theory $d'_{human} = \sqrt{2}\Phi^{-1}(PC_{human})$, where PC_{human} is the proportion of times that the comparison is chosen in a given condition in a 2IFC experiment and $\Phi^{-1}(\cdot)$ is the inverse cumulative normal. (This expression is correct assuming the observer uses the optimal criterion, an assumption that is justified by the data.)

To measure ideal sensitivity, we ran the ideal observer in a simulated experiment with the same stimuli as the human. (The

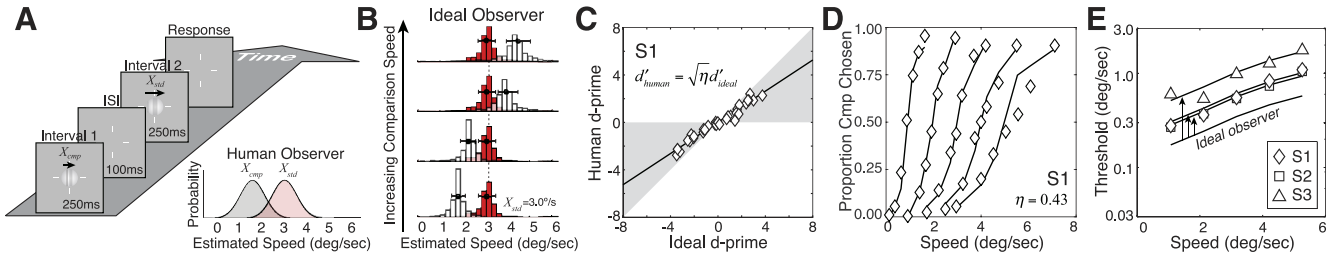

Figure 5. Measuring speed discrimination. ***A***, The task in a 2IFC experiment was to report the interval containing the faster of two naturalistic image movies. Unlike classic psychophysical studies, which present the same stimuli hundreds of times, the current study presents hundreds of unique stimuli one time each. This design injects naturalistic stimulus variability into the experiment. Human responses are assumed to be based on samples from decision variable distributions (inset). ***B***, Ideal observer estimates across hundreds of standard (red) and comparison movies (white) at one standard speed (3°/s) and four comparison speeds. ***C***, Human versus ideal observer sensitivity for all standard and comparison speeds. Shaded regions mark regions of plot where humans are less efficient than ideal but are still performing the task. For all conditions, humans are less sensitive than the ideal observer by a single-scale factor: efficiency: $d'_{human} = \sqrt{\eta} d'_{ideal}$. Negative d' values correspond to conditions in which the comparison was slower than the standard. ***D***, Psychometric functions for 1 human observer (symbols) at five standard speeds. The degraded ideal observer (solid curves) matches the efficiency of the human observer (one parameter fit to human data). ***E***, Human speed discrimination thresholds ($d' = 1.0$) as a function of standard speed for 3 human observers (symbols) on a semilog plot. The pattern of human thresholds matches ideal observer thresholds (solid curve). Vertically shifting the ideal observer thresholds by an amount set by each human's efficiency (arrows) shows degraded observer performance (solid curves, one free parameter fit per human).

ideal observer was trained on different stimuli than the human and ideal observers were tested on.) Ideal sensitivity (i.e., d') was computed directly from the distributions of ideal observer speed estimates in each condition (Fig. 5*B*). Human and ideal sensitivities across all speeds are linearly related (Fig. 5*C*). Rearranging Equation 9 shows that human sensitivity $d'_{human} = \sqrt{\eta} d'_{ideal}$ equals the ideal observer sensitivity degraded (scaled) by the square root of the efficiency. Thus, a single free parameter (efficiency) relates the pattern of human and ideal sensitivities for all conditions. The efficiencies of the first, second, and third human observers are 0.43, 0.41, and 0.17, respectively.

Transforming the sensitivity data back into percent comparison chosen shows that the details of the degraded ideal nicely account for the human psychometric functions (Fig. 5*D*). The psychometric functions can be summarized by the speed discrimination thresholds ($d' = 1.0$; 76% correct in a 2IFC task). The pattern of human and ideal thresholds match; the proportional increases of the human and ideal threshold functions with speed are the same (Fig. 5*E*). These results quantify human uncertainty σ^2_{human} and show that an ideal observer analysis of naturalistic stimuli predicts the pattern of human speed discrimination performance, and replicate our own previously published findings (Burge and Geisler, 2015).

Together, the ideal observer and speed discrimination experiment reveal the degree of human inefficiency (i.e., how far human performance falls short of the theoretical ideal). But they cannot determine the sources of this inefficiency. Humans could be inefficient because of late noise (i.e., stochastic internal sources of variability arising after early noise). Humans could also be inefficient because of fixed suboptimal computations. If inefficiency is due exclusively to late noise, stimulus variability must equally limit human and ideal observer performance. If human inefficiency is partly due to suboptimal computations, stimulus variability will cause more stimulus-driven variability in the human than in the ideal. How can human behavioral variability be partitioned to determine the sources of inefficiency in speed perception? To do so, additional experimental tools are required.

Predicting and measuring decision variable correlation

A double-pass experiment, when paired with ideal observer analysis, can determine why human performance falls short of the theoretical ideal. In a double-pass experiment (Burgess and Colborne, 1988; Gold et al., 1999; Li et al., 2006), each human observer responds to each of a large number of unique trials (the first pass), and then performs the entire experiment again (the second pass). Double-pass experiments can "unpack" each point on the psychometric function (Fig. 6*A*,*B*), providing far more information about the factors driving and limiting human performance than standard single-pass experiments. The correlation in the human decision variable across passes (decision variable correlation) is key for identifying the factors that limit performance and determine efficiency (Burgess and Colborne, 1988; Sebastian and Geisler, 2018).

The power of this experimental design is that it enables behavioral variability to be partitioned into correlated and uncorrelated factors. Factors that are correlated across passes (e.g., the stimuli) increase the correlation of the decision variable across passes. Factors that are uncorrelated across passes (e.g., internal noise) decrease decision variable correlation. If the variance of the human decision variable is dictated only by stimulus-driven variability, decision variable correlation will equal 1.0. If the variance of the human decision variable is dictated only by internal noise, decision variable correlation will equal 0.0. If both stimulus-driven variability and internal noise play a role, the correlation will have an intermediate value.

Decision variable correlation, like the decision variable itself, cannot be measured directly using standard psychophysical methods. Rather, it must be inferred from the repeatability of responses across passes in each condition. The higher the decision variable correlation, the greater the proportion of times responses agree (i.e., repeat) in a given condition (Fig. 6*B*,*C*).

In each condition, we used the pattern of response agreement to estimate decision variable correlation (Fig. 6*B*,*C*), and then plotted agreement against the proportion of times the human observer (symbols) chose the comparison stimulus as faster (Fig. 6*D*). Human response agreement implies a decision variable correlation that is significantly different from zero. For the seven conditions shown in Figure 6*D* (i.e., all comparison speeds at the 1°/s standard speed), the maximum likelihood fit of decision variable correlation across the seven comparison levels is 0.43. Thus, 43% of the total variance in the human decision variable is due to factors that are correlated across repeated presentations of the same trials.

How should the estimate of decision variable correlation be interpreted? Human decision variable correlation across passes is given by the following:

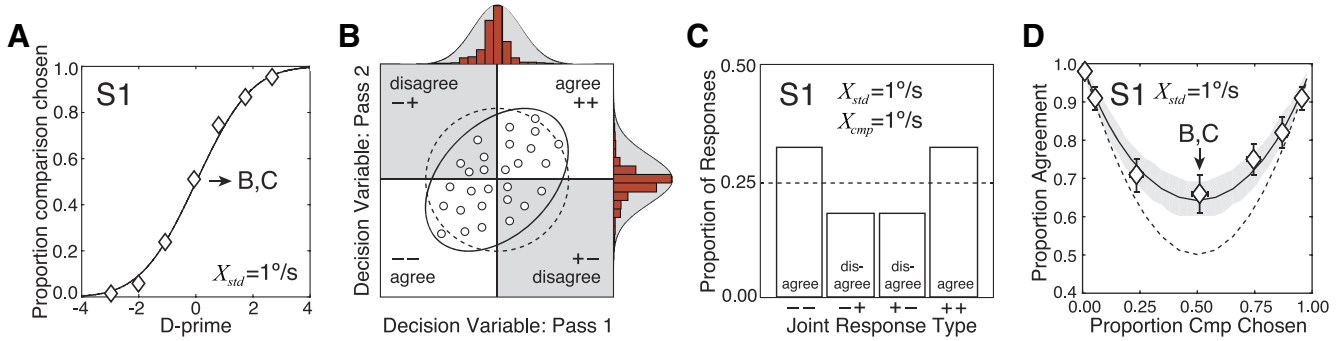

Figure 6. Decision variable correlation and response repeatability in a double-pass experiment. ***A***, Psychometric data from the first human observer and cumulative Gaussian fit plotted as proportion comparison chosen versus d' for the standard speed of 1°/s (same data as in Fig. 5*D*). ***B***, Schematic for visualizing decision variable correlation across passes when standard and comparison speeds are identical (e.g., both equal 1°/s). Samples correspond to individual double-pass trials (small circles). The value of each sample indicates the difference between the estimated speeds of the comparison and standard stimuli on each trial. Decision variable values corresponding to response agreements and disagreements fall in white and gray quadrants, respectively. Decision variable distributions with the decision variable correlation predicted by efficiency (solid ellipse) and by the null model with a decision variable correlation of zero (dashed ellipse). Decision variable correlation depends on the relative importance of correlated and uncorrelated factors across passes. Stimuli are correlated on each repeated trial of a double-pass experiment; internal noise is not. Criteria on each pass (vertical and horizontal lines, respectively) are assumed to be optimal and at zero. ***C***, Predicted response counts (bars) for each response type ($--, -+, +-, ++$) across passes (100 trials per condition) given the decision variable correlation shown in ***B***. ***D***, Proportion of trials on which responses agreed across both passes of the double-pass experiment as a function of proportion comparison chosen for 1 human observer. Agreement data (symbols) and prediction (solid curve) assuming that efficiency predicts decision variable correlation (i.e., that all human inefficiency is due to late noise). The null prediction assumes that the decision variable correlation across passes is zero (dashed curve). The agreement data are predicted directly from the efficiency of the human observer (zero free parameters). Error bars indicate 68% bootstrapped CIs on human agreement. Shaded regions represent 68% CIs from 10,000 Monte Carlo simulations of the predicted agreement data assuming 100 trials per condition.

$$\rho = \frac{\sigma_E^2}{\sigma_E^2 + \sigma_I^2} = \frac{\sigma_E^2}{\sigma_{human}^2} \tag{10}$$

where σ_E^2 is the variance of the speed estimates due to external (i.e., stimulus) factors, σ_I^2 is the variance due to internal factors (e.g., noise), and σ_{human}^2 is the total variance of the human speed estimates. Decision variable correlation is driven by stimulus variation because the stimuli are perfectly correlated across passes.

The estimated decision variable correlation is strikingly similar to the efficiency measured for each observer. Although the exact relationship between decision variable correlation and efficiency depends on the source of human inefficiency, the fact that they are similar is no accident. Under the hypothesis that all human inefficiency is due to noise (i.e., stochastic internal factors that are uncorrelated with the stimuli), stimulus variability must impact human and ideal observers identically: the stimulus-driven variance in the human speed estimates (σ_E^2 in Eq. 10) will equal the stimulus-driven variance in the ideal observer speed estimates (σ_E^2 in Eq. 9). Plugging Equation 9 into Equation 10 shows that, under the stated hypothesis, human decision variable correlation equals efficiency as follows:

$$\rho = \eta \tag{11}$$

This mathematical relationship has important consequences. It means that the estimate of human efficiency from the speed discrimination experiment (Fig. 5*C*) provides a zero-free parameter prediction of human decision variable correlation in the double-pass experiment (Fig. 6). The behavioral data confirm this prediction. Human efficiency in the discrimination experiment quantitatively predicts human response agreement in the double-pass experiment (Fig. 6*D*, symbols vs solid curve). The implication of this result is striking. It suggests that natural stimulus variability equally limits human and ideal observers and thus that the source of human inefficiency is due near-exclusively to late noise. Human speed discrimination is therefore optimal, except for the impact of late internal noise.

These results generalize across all conditions and human observers. Figure 7*A* shows measured response agreement versus proportion comparison chosen for the first human observer in each of the five standard speed conditions. Figure 7*B* plots measured response agreement against efficiency-predicted agreement, summarizing the agreement data for each human observer across all standard speeds; prediction uncertainty given the number of double-pass trials in each condition is shown as 95% CIs (shaded regions). The decision variable correlations that best account for the response repeatability across all conditions of the first, second, and third human observers are 0.45, 0.43, and 0.18, respectively. For the first 2 observers, stimulus-driven variance and noise variance have approximately the same magnitude. For all observers, the data are consistent with the hypothesis that decision variable correlation equals efficiency (solid curves), and data are not consistent with the null model in which decision variable correlation equals zero (dashed curves). Figure 7*C* plots decision variable correlation against efficiency for each human observer. Efficiency tightly predicts decision variable correlation for all three human observers, with zero additional free parameters.

These results must be interpreted with some caution. Uncertainty about the amount of early measurement noise can cause uncertainty about human efficiency (Eq. 8) and thus about the predicted decision variable correlation (Eq. 11). We simulated ideal observers with different amounts of early noise and computed efficiency for each human observer (Fig. 8*A*). Fortunately, the detection experiment establishes an upper bound on the amount of early noise for each human observer (compare Fig. 3), thereby constraining the uncertainty about the predicted decision variable correlation (Fig. 8*B*, red brackets). Because the upper bound on early noise is low, the maximum and minimum possible efficiencies differ by ~10%, depending on whether early noise at the upper or lower bound is assumed (Fig. 8*A*,*B*, red brackets). The measured decision variable correlation values (Fig. 8*C*) are in line with the predictions. Thus, uncertainty about the amount of early noise has only a minor impact on the interpretation of our results.

In the best performing observers, natural stimulus variability accounts for nearly half of all behavioral variability, despite the

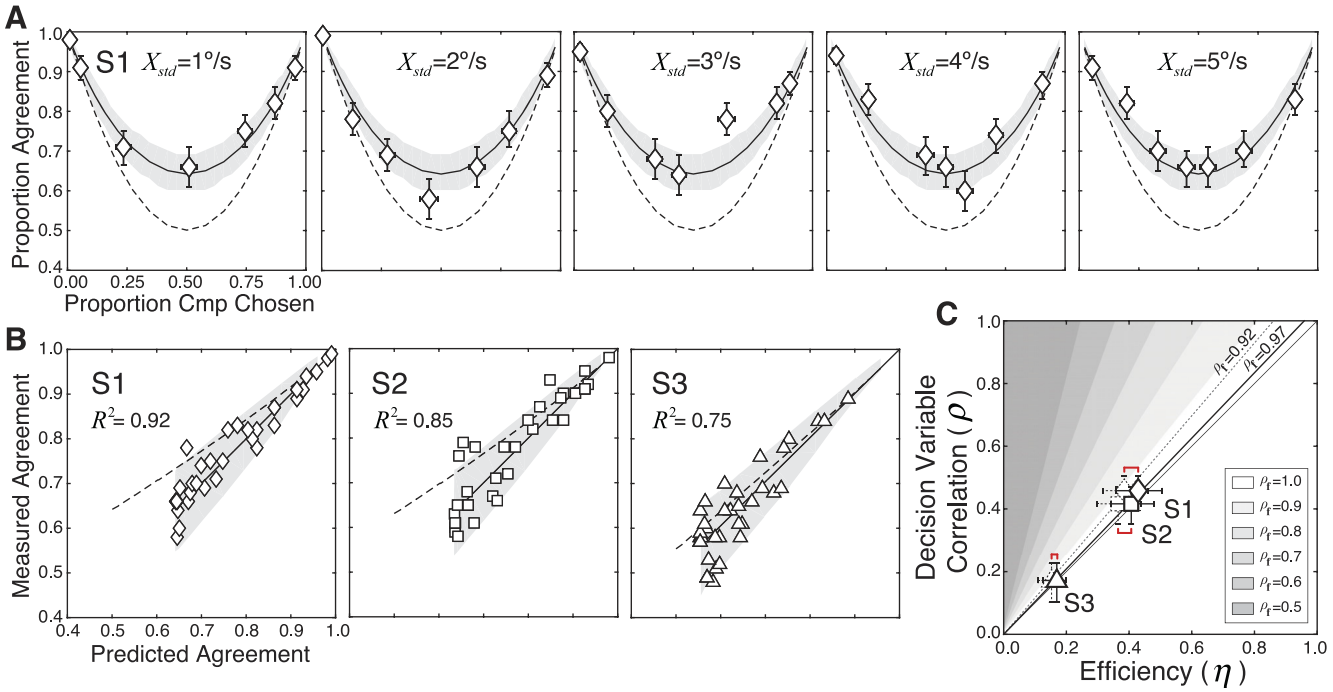

Figure 7. Predicted versus measured response agreement and decision variable correlation. ***A***, Proportion response agreement versus proportion comparison chosen for all five standard speeds (1–5°/s) for the first human observer. Human data (symbols) and predictions (curves) are shown using the same conventions as in Figure 6*D*. ***B***, Measured versus predicted response agreement for all conditions and all human observers (symbols). Human agreement equals efficiency-predicted agreement for all three human observers (solid line). Shaded regions represent 95% CIs on the prediction from 1000 Monte Carlo simulations. Efficiency-predicted agreement for the null model, which assumes decision variable correlation is zero, is also shown (dashed curve). ***C***, Decision variable correlation versus efficiency for each human observer (symbols). Human efficiency, measured in first pass of the speed discrimination experiment, tightly predicts human decision variable correlation in the double-pass experiment with zero free parameters. Error bars indicate 95% bootstrapped CIs on human efficiency and on human decision variable correlation. Shaded regions represent the expected relationship between efficiency and decision variable correlation if humans use fixed suboptimal computations (i.e., suboptimal receptive fields). Red brackets indicate uncertainty about the precise value of efficiency due to uncertainty about the precise amount of early noise (Fig. 3). Solid and dashed black lines indicate the best-fit regression lines, corresponding to receptive field correlations of 0.97 and 0.92, respectively.

fact that the naturalistic stimulus set used to probe speed discrimination performance almost certainly underestimates the importance of stimulus variability in natural viewing (see Discussion). External variability therefore shapes the optimal computations, dictates the pattern of human performance, and predicts the partition of behavioral variability (i.e., the relative importance of external and internal sources of variability). These findings motivate continued efforts to model and characterize how natural stimulus variability impacts neural and perceptual performance in natural tasks.

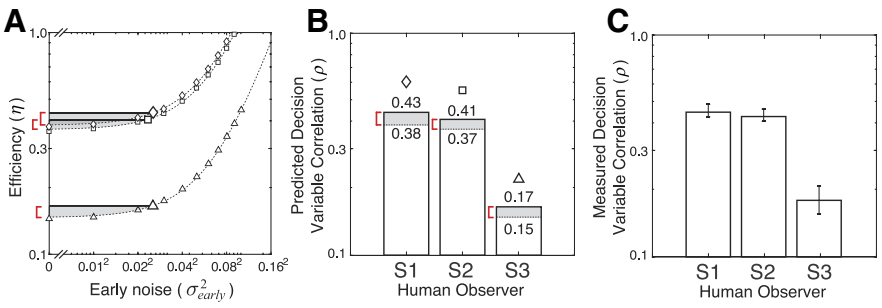

Figure 8. Early noise, efficiency, and predicted decision variable correlation. ***A***, Efficiency in speed discrimination for each human observer (symbols) changes as a function of the amount of early noise modeled in the ideal observer. If early noise is negligible, efficiency is given by $\eta = \sigma_E^2/\sigma_{human}^2$ (Eq. 9). If early noise is non-negligible, efficiency is given by $\eta = (\sigma_E^2 + \sigma_{I,\,early}^2)/\sigma_{human}^2$ (Eq. 8). Red brackets and shaded regions indicate the minimum and maximum human efficiencies, given the bound on early noise established by the detection experiment (compare Fig. 3). ***B***, Predicted decision variable correlation for each human observer given the uncertainty about human efficiency. The maximum (solid line) and minimum (dashed line) predicted decision variable correlations correspond to ideal observers having the maximum and minimum amount of early noise. The predicted decision variable correlations differ by ~10% at maximum. ***C***, Measured decision variable correlation for each human observer. Error bars indicate 95% bootstrapped CIs.

Suboptimal computations

Human efficiency equals human decision variable correlation (Figs. 7*C*, 8*B*,*C*). To confidently conclude from this result that human inefficiency is almost entirely due to noise (i.e., stochastic internal sources of variability), we must rule out the possibility that suboptimal computations can produce the same result. How do fixed suboptimal computations impact the relationship between efficiency and decision variable correlation? To answer this question, one must determine how suboptimal computations impact the stimulus-driven component of the decision variable. To do so, we analyzed the estimates of a degraded observer that suboptimally encodes stimulus features (Burgess et al., 1981; Dosher and Lu, 1998; Neri and Levi, 2006; Sebastian and Geisler, 2018). If the wrong features are encoded, informative features may be missed, irrelevant features may be processed, and the variance of the stimulus-driven component of the decision variable may be increased relative to the ideal.

To create suboptimal feature encoders (i.e., suboptimal receptive fields), we corrupted the optimal receptive fields with fixed

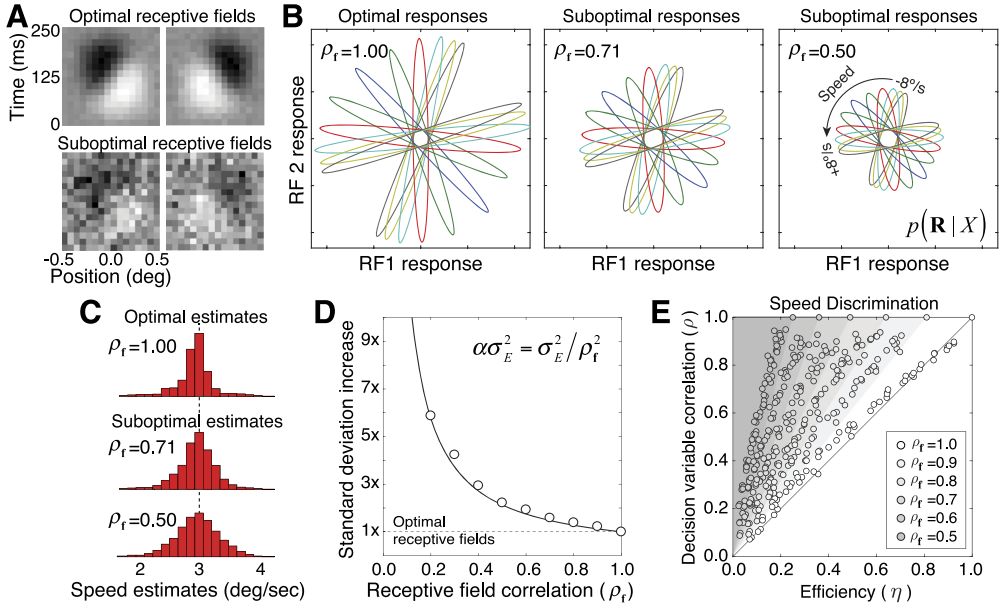

Figure 9. Relationship between suboptimal receptive fields and stimulus-driven variability in degraded observers. ***A***, Optimal receptive fields (top; also see Fig. 4*A*) and suboptimal receptive fields from the degraded observer (bottom); only the first two receptive fields of each observer are shown. To obtain a suboptimal receptive field with a particular receptive field correlation ρ_f, we added fixed samples of Gaussian white noise to the corresponding optimal receptive field. The variance of the corrupting noise is given by $\sigma^2_{corrupt} = ((1/\rho_f^2) - 1)/N$, where N is the number of pixels defining each receptive field. ***B***, Impact of suboptimal receptive fields on the conditional response distributions $p(\mathbf{R}|X)$. As the receptive fields become more suboptimal, the response distributions (colored ellipses) more poorly distinguish different values of the latent variable (colors). ***C***, Effect of suboptimal receptive fields on degraded observer speed estimates for movies drifting at one speed (3°/s). As receptive field correlation decreases, the stimulus-driven variance of the estimates increases because informative stimulus features are not encoded and uninformative features are. ***D***, The proportional increase of stimulus-driven SD for degraded versus the ideal observer estimates, assuming that the degraded observer has no late internal noise. Symbols plot the mean result from 100 Monte Carlo simulations. The stimulus-driven variance of the speed estimates increases with the squared inverse of receptive field correlation. ***E***, Relationship between decision variable correlation and efficiency for degraded observers with different combinations of fixed suboptimal computations (i.e., receptive field correlations; gray levels) and internal noise. Points indicate mean decision variable correlation and mean efficiency from 100 Monte Carlo simulations of each degraded observer.

samples of Gaussian white noise (Fig. 9*A*). Receptive field correlation (i.e., cosine similarity) quantifies the degree of suboptimality $\rho_f = \mathbf{f}_{opt}^T \mathbf{f}_{subopt} / (\|\mathbf{f}_{opt}\| \, \|\mathbf{f}_{subopt}\|)$ where $\mathbf{f}_{opt}$ and $\mathbf{f}_{subopt}$ are the optimal and suboptimal receptive fields, respectively. Compared with the responses of the optimal receptive fields, the responses of these suboptimal receptive fields segregate less well as a function of speed (Fig. 9*B*). We generated degraded observers with suboptimal receptive fields having different receptive field correlations and examined estimation performance (Fig. 9*C*). We found that the stimulus-driven variance $\alpha\sigma_E^2$ of the degraded observer estimates is a scaled version of the ideal stimulus-driven variance; the scale factor $\alpha = 1/\rho_f^2$ is equal to the squared inverse of receptive field correlation (Fig. 9*D*). Thus, suboptimal receptive fields systematically increase the variance of the stimulus-driven component of the decision variable.

If humans are well modeled by a degraded observer with both suboptimal receptive fields and late noise, the total variance of the human estimates is given by $\sigma^2_{human} = \alpha\sigma_E^2 + \sigma_I^2$. Replacing terms in Equations 9 and 10 and performing some simple algebra show that the relationship between efficiency and decision variable correlation is given by the following:

$$\rho = \alpha\eta = \frac{\eta}{\rho_f^2} \qquad (12)$$

Thus, with suboptimal computations (i.e., receptive fields), decision variable correlation will be systematically larger than efficiency by the inverse square of receptive field correlation. (When receptive field correlation equals 1.0, Eq. 12 reduces to Eq. 11.) For example, if receptive field correlation is 0.5, decision variable correlation is 4× higher than efficiency. We verified the relationship between decision variable correlation and efficiency by simulating degraded observers with different levels of suboptimal computations and late noise (Fig. 9*E*). As predicted by Equation 12, the more suboptimal the computations (i.e., receptive field correlations), the more decision variable correlation exceeds efficiency. We reanalyzed our results in the context of Equation 12, comparing the behavioral data with the predictions of various degraded observer models. For all three observers, decision variable correlation is larger than efficiency by ~5%, corresponding to a receptive field correlation of 0.97 (Fig. 7*C*). (These numbers assume an ideal observer with early noise set to the upper bound established by the detection experiment; Fig. 3). If no early noise is assumed, then decision variable correlation exceeds efficiency by 15%, corresponding to a receptive field correlation of 0.92 (Fig. 7*C*). Thus, no more than 15% of human inefficiency can be attributed to fixed suboptimal computations.

The simulations just described only consider the potential impact of fixed suboptimal computations that are linear. We cannot definitively rule out nonlinear suboptimal computations that leave stimulus-driven variability unchanged while selectively amplifying the impact of early noise, making amplified early noise indistinguishable from late noise. However, such computations are highly unlikely, given current knowledge of early visual processing. More importantly, suboptimal computations that selectively amplify early noise will not alter the predicted relationships between efficiency and decision variable correlation.

Thus, our results imply that the deterministic computations performed by the human visual system in speed estimation are very nearly optimal. Although natural stimulus (i.e., nuisance)

variability is a major and unavoidable factor that limits performance in natural viewing, its impact is minimized as much as possible by the computations performed by the visual system.

Stimulus variability and behavioral variability

In this paper, we have shown that natural stimulus variability limits behavioral performance and drives response repeatability. Thus, reducing stimulus variability should increase sensitivity (i.e., improve behavioral performance) but decrease response repeatability. To test this prediction, we ran a new speed discrimination experiment using drifting random-phase sinewave gratings (Fig. 10). A stimulus set composed of drifting sinewaves has less variability than the set of naturalistic stimuli used in the main experiment. As predicted, with sinewave stimuli human sensitivity improves (Fig. 10*A*), responses become less repeatable (Fig. 10*B*), and decision variable correlation is lower (Fig. 10*C*). Interestingly, reducing stimulus variability affects decision variable correlation in the third human observer less than it does in the first two. This is the expected pattern of results given that the third observer (S3) had low decision variable correlation with naturalistic stimuli and was thus already dominated by internal noise (Figs. 7*C*, 8*C*). However, not all of the results were quite as expected. We anticipated that decision variable correlation would equal 0.0 for all three human observers with sinewave stimuli. But decision variable correlation exceeded 0.0 for each observer. What accounts for this discrepancy? We have ruled out commonly considered trial order effects (e.g., feedback-based effects) as the cause (Laming, 1979), but we are unsure of the cause. Whatever the case, with reduced stimulus variability, internal noise, which is uncorrelated across stimulus repeats, becomes the dominant source of variability limiting performance in all human observers.

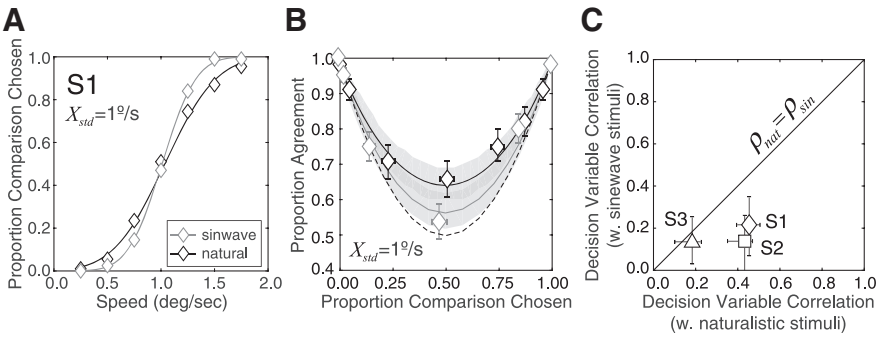

Figure 10. Effects of reducing stimulus variability. ***A***, Speed discrimination psychometric functions for the first human observer with naturalistic stimuli (black curve) and drifting sinewave stimuli (gray curve) for a 1°/s standard speed. Sinewave stimuli can be discriminated more precisely. ***B***, Proportion response agreement versus proportion comparison chosen for naturalistic stimuli (black) and artificial stimuli (gray) for the same human observer. ***C***, Decision variable correlation with artificial stimuli versus decision variable correlation with naturalistic stimuli for each human observer (symbols). Error bars indicate 95% bootstrapped CIs. Decision variable correlation is consistently lower when artificial stimuli are used.

Discussion

Simple stimuli and/or simple tasks have dominated behavioral neuroscience because of the need for rigor and interpretability in assessing stimulus influences on neural and behavioral responses. The present experiments demonstrate that, with appropriate techniques, the required rigor and interpretability can be obtained with naturalistic stimuli. We have shown that image-computable ideal observers can be fruitfully combined with human behavioral experiments to reveal the factors that limit behavioral performance in mid-level tasks with naturalistic stimuli. In particular, an image-computable ideal observer, constrained by the same factors as the early visual system, predicts the pattern of human speed discrimination performance with naturalistic stimuli (Burge and Geisler, 2015). Perhaps more remarkably, human efficiency in the task predicts human decision variable correlation in a double-pass experiment without free parameters, a result that holds only if the deterministic computations performed by humans are very nearly optimal.

Limitations and future directions

One limitation of our approach, which is common to most psychophysical approaches, is that it cannot pinpoint the processing stage or brain area at which the limiting source of internal variability arises. Although we model it as noise occurring at the level of the decision variable, it could also occur at the encoding receptive field responses, the computation of the likelihood, the readout of the posterior into estimates, the placement of the criterion at the decision stage, or some combination of the above. We believe we have ruled out the possibility that the noise limiting speed discrimination is early (Fig. 3). But we cannot distinguish among other stochastic sources of internal variability. These issues are probably best addressed with neurophysiological methods. Similarly, our approach cannot distinguish between different types of fixed suboptimal computations. We modeled them by degrading each in the set of optimal receptive fields. But an array of computations that make fixed suboptimal use of the available stimulus information could have similar effects.

Another potential issue is that eye movements were not controlled, raising the concern that human and ideal observers were not on equal footing. If eye movements are stimulus independent, they could manifest like internal noise, and decrease decision variable correlation (Rolfs, 2009; Kowler, 2011). On the other hand, if different eye movements are reliably elicited by different stimuli with the same speed (Turano and Heidenreich, 1999; Rucci and Poletti, 2015), they could manifest like suboptimal computations, and increase decision variable correlation. However, we believe that the steps we took to minimize the possible impact of uncontrolled eye movements are likely to have been largely successful. First, stimuli were presented for 250 ms, the approximate duration of a typical fixation, and our stimuli were above half-maximal contrast for only ~200 ms. Under stimulus conditions (i.e., speeds and contrasts) similar to ours, smooth pursuit eye movements have a latency of 140–200 ms (Spering et al., 2005). Thus, if large eye movements occurred, it is likely that they would have occurred only in the last fraction of the trial. Second, numerous reports indicate that, when estimating motion, humans and other primates tend to weight stimulus information more heavily at the beginning than at the end of trial (Yates et al., 2017). Thus, the portion of the trial in which the eyes are most likely to have been stable is the portion that is most likely to have contributed to the speed estimate. Finally, fixational eye movements (i.e., drift, microsaccades, tremor) are likely to have contributed to our estimate of early measurement noise, and thus would have equivalently impacted both human and ideal performance. Still, given that eye movements can impact speed percepts under certain conditions (Turano and Heidenreich, 1999;

Freeman et al., 2010; Goettker et al., 2018), this issue should be examined rigorously in future experiments.

There are many other possible directions for future work. First, there is a well-established tradition of examining how changing overall contrast impacts speed-sensitive neurons and speed perception (Thompson, 1982; Schrater et al., 2000; Weiss et al., 2002; Priebe et al., 2003; Priebe and Lisberger, 2004; Jogan and Stocker, 2015; Gekas et al., 2017). All stimuli in the current experiment were fixed to the most common contrast in the natural image movie set. As overall contrast is reduced, speed-sensitive neurons respond less vigorously, and moving stimuli are perceived to move more slowly (Thompson, 1982; Weiss et al., 2002; Priebe et al., 2003). It is widely believed that these effects occur because the visual system has internalized a prior for slow speed (Weiss et al., 2002). In the current manuscript, rather than covering well-trodden ground, we have focused on quantifying how image structure (i.e., the pattern of contrast) impacts speed estimation and discrimination. Thus, our results likely underestimate the impact of stimulus variability on ideal and human performance in natural viewing. The approach advanced in this manuscript can be generalized to examine how changes in overall contrast impact human and ideal performance. The role of stimulus variability has not been examined in this context, and may make an interesting topic for future work. Experiments should also be performed with full space-time (i.e., *xyt*) movies, with stimuli containing looming and discontinuous motion (Schrater et al., 2001; Nitzany and Victor, 2014). Finally, these same methods could be applied to a host of other tasks in vision and in other sensory modalities. New databases of natural images and natural sounds with corresponding groundtruth information about the distal scenes will significantly aid these efforts (Adams et al., 2016; Burge et al., 2016; Traer and McDermott, 2016).

Sources of performance limits

Efforts to determine the dominant factors that limit performance span research from sensation to cognition. The conclusions that researchers have reached are as diverse as the research areas in which the efforts have been undertaken. Stimulus noise (Hecht et al., 1942), physiological optics (Banks et al., 1987), internal noise (Burgess et al., 1981; Pelli, 1985, 1991; Williams, 1985), suboptimal computations (Dosher and Lu, 1998; Beck et al., 2012; Drugowitsch et al., 2016), trial-sequential dependences (Laming, 1979), and various cognitive factors (Tversky and Kahneman, 1971) have all been implicated as the dominant factors that limit performance. What accounts for the diversity of these conclusions? We cannot provide a definitive answer. The relative importance of these factors is likely to depend on several things.

Evolution has pushed sensory-perceptual systems toward the optimal solutions for tasks that are critical for survival and reproduction. Humans are more likely to be assessed as optimal when visual systems are probed with stimuli that they evolved to process in tasks that they evolved to perform. In target detection tasks, for example, humans become progressively more efficient as stimuli become more natural (Banks et al., 1987; Abbey and Eckstein, 2014; Sebastian et al., 2017). Conversely, when stimuli and tasks bear little relation to those that drove the evolution of the system, the computations are less likely to be optimal. A new framework, a science of tasks, would be useful to help reconcile these disparate findings.

Image-computable ideal observers

Ideal observer analysis has a long history in vision science and systems neuroscience. In conjunction with behavioral experiments, image-computable ideal observers have shown that human light sensitivity is as sensitive as allowed by the laws of physics (Hecht et al., 1942), that the shape of the human contrast sensitivity function is dictated by the optics of the human eye (Banks et al., 1987), and that the pattern of human performance in a wide variety of basic psychophysical tasks can be predicted from first principles (Geisler, 1989).

To develop an image-computable ideal observer, it is critical to have a characterization of the task-relevant stimulus statistics. Obtaining such a characterization has been out of reach for all but the simplest tasks with the simplest stimuli. The vision and systems neuroscience communities have traditionally focused on understanding how simple forms of stimulus variability (e.g., Poisson or Gaussian white noise) impact performance (Hecht et al., 1942; Burgess et al., 1981; Pelli, 1985; Banks et al., 1987; Frechette et al., 2005). The impact of natural stimulus variability, the variation in light patterns associated with different natural scenes sharing the same latent variable values, has only recently begun to receive significant attention (Geisler and Perry, 2009; Burge and Geisler, 2011, 2012, 2014, 2015; Kane et al., 2011; Sebastian et al., 2015, 2017; Schütt and Wichmann, 2017; Kim and Burge, 2018; Sinha et al., 2018).

Many impactful ideal observer models developed in recent years are not image-computable (Landy et al., 1995; Ernst and Banks, 2002; Weiss et al., 2002; Stocker and Simoncelli, 2006; Burge et al., 2010; Wei and Stocker, 2015). The weakness of these models is that they do not explicitly specify the stimulus encoding process, and therefore make assumptions about the information that stimuli provide about the task-relevant variable (e.g., the likelihood function in the Bayesian framework). Consequently, these models cannot predict directly from stimuli how nuisance stimulus variability will impact behavioral variability, or explain how information is transformed as it proceeds through the hierarchy of visual processing stages. Image-computable models are thus necessary to achieve the goal of understanding how vision works with real-world stimuli. The current work represents an important step in that direction.

Impact on neuroscience

Behavioral and neural responses both vary from trial to trial, even when the value of the latent variable (e.g., speed) is held constant. In many classic neurophysiological experiments, stimulus variability is eliminated by design, and experimental distinctions are not made between the latent variable of interest (e.g., orientation) and the stimulus (e.g., an oriented Gabor) used to probe neural response. Such experiments are well suited for quantifying how different internal factors impact neural variability. Indeed, it has recently been shown that, under these conditions, neural variability can be partitioned into two internal factors: a Poisson point process and systemwide gain fluctuations (Goris et al., 2014). This approach provides an elegant account of a widely observed phenomenon ("super-Poisson variability") (Tomko and Crapper, 1974; Tolhurst et al., 1981, 1983) that had previously resisted rigorous explanation. However, the designs of these classic experiments are unsuitable for estimating the impact of stimulus variability on neural response.

In the real world, behavioral variability is jointly driven by external and internal factors. Our results show that both factors place similar limits on performance. A full account of neural encoding and decoding must include a treatment of all significant sources of response variability. Partitioning the impact of realistic forms of stimulus variability from internal sources of neural variability will be an important next step for the field.

References

Abbey CK, Eckstein MP (2014) Observer efficiency in free-localization tasks with correlated noise. Front Psychol 5:345.

Adams WJ, Elder JH, Graf EW, Leyland J, Lugtigheid AJ, Muryy A (2016) The Southampton-York Natural Scenes (SYNS) dataset: statistics of surface attitude. Sci Rep 6:35805.

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284–299.

Albrecht DG, Geisler WS (1991) Motion selectivity and the contrast-response function of simple cells in the visual cortex. Vis Neurosci 7:531–546.

Banks MS, Geisler WS, Bennett PJ (1987) The physical limits of grating visibility. Vision Res 27:1915–1924.

Beck JM, Ma WJ, Pitkow X, Latham PE, Pouget A (2012) Not noisy, just wrong: the role of suboptimal inference in behavioral variability. Neuron 74:30–39.

Bishop CM (2006) Pattern recognition and machine learning. New York: Springer.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Burge J, Geisler WS (2011) Optimal defocus estimation in individual natural images. Proc Natl Acad Sci U S A 108:16849–16854.

Burge J, Geisler (2012) Optimal defocus estimates from individual images for autofocusing a digital camera, Proc. SPIE 8299, Digital Photography VIII, 82990E. Available at https://doi.org/10.1117/12.912066.

Burge J, Geisler WS (2014) Optimal disparity estimation in natural stereo images. J Vis 14:2.

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

Burge J, Jaini P (2017) Accuracy maximization analysis for sensory-perceptual tasks: computational improvements, filter robustness, and coding advantages for scaled additive noise. PLoS Comput Biol 13: e1005281.

Burge J, Fowlkes CC, Banks MS (2010) Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. J Neurosci 30:7269–7280.

Burge J, McCann BC, Geisler WS (2016) Estimating 3D tilt from local image cues in natural scenes. J Vis 16:2.

Burge J, Rodriguez-Lopez V, Dorronsoro C (2019) Monovision and the misperception of motion. Curr Biol 29:2586–2592.e4.

Burgess AE, Colborne B (1988) Visual signal detection: IV. Observer inconsistency. J Opt Soc Am A 5:617–627.

Burgess AE, Wagner RF, Jennings RJ, Barlow HB (1981) Efficiency of human visual signal discrimination. Science 214:93–94.

Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62.

Dosher BA, Lu ZL (1998) Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proc Natl Acad Sci U S A 95:13988–13993.

Drugowitsch J, Wyart V, Devauchelle AD, Koechlin E (2016) Computational precision of mental inference as critical source of human choice suboptimality. Neuron 92:1398–1411.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Fleming RW, Storrs KR (2019) Learning to see stuff. Curr Opin Behav Sci 30:100–108.

Frechette ES, Sher A, Grivich MI, Petrusca D, Litke AM, Chichilnisky EJ (2005) Fidelity of the ensemble code for visual motion in primate retina. J Neurophysiol 94:119–135.

Freeman TC, Champion RA, Warren PA (2010) A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. Curr Biol 20:757–762.

Gattass R, Gross CG, Sandell JH (1981) Visual topography of V2 in the macaque. J Comp Neurol 201:519–539.

Gattass R, Sousa AP, Gross CG (1988) Visuotopic organization and extent of V3 and V4 of the macaque. J Neurosci 8:1831–1845.

Geisler WS (1989) Sequential ideal-observer analysis of visual discriminations. Psychol Rev 96:267–314.

Geisler WS, Perry JS (2009) Contour statistics in natural images: grouping across occlusions. Vis Neurosci 26:109–121.

Geisler WS, Najemnik J, Ing AD (2009) Optimal stimulus encoders for natural tasks. J Vis 9:17.1–16.

Gekas N, Meso AI, Masson GS, Mamassian P (2017) A normalization mechanism for estimating visual motion across speeds and scales. Curr Biol 27:1514–1520.e3.

Goettker A, Braun DI, Schütz AC, Gegenfurtner KR (2018) Execution of saccadic eye movements affects speed perception. Proc Natl Acad Sci U S A 115:2240–2245.

Gold JM, Bennett PJ, Sekuler AB (1999) Signal but not noise changes with perceptual learning. Nature 402:176–178.

Goris RL, Movshon JA, Simoncelli EP (2014) Partitioning neuronal variability. Nat Neurosci 17:858–865.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.

Hecht S, Shlaer S, Pirenne MH (1942) Energy, quanta, and vision. J Gen Physiol 25:819–840.

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Vis Neurosci 9:181–197.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154.

Iyer A, Burge J (2019) The statistics of how natural images drive the responses of neurons. J Vis 19:4.

Jaini P, Burge J (2017) Linking normative models of natural tasks to descriptive models of neural response. J Vis 17:16.

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

Kane D, Bex P, Dakin S (2011) Quantifying "the aperture problem" for judgments of motion direction in natural scenes. J Vis 11:3.

Kim S, Burge J (2018) The lawful imprecision of human surface tilt estimation in natural scenes. eLife 7:e31448.

Kowler E (2011) Eye movements: the past 25 years. Vision Res 51:1457–1483.

Laming DR (1979) Choice reaction performance following an error. Acta Psychol 43:199–224.

Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Res 35:389–412.

Li RW, Klein SA, Levi DM (2006) The receptive field and internal noise for position acuity change with feature separation. J Vis 6:311–321.

Li X, Lu ZL, Xu P, Jin J, Zhou Y (2003) Generating high gray-level resolution monochrome displays with conventional computer graphics cards and color monitors. J Neurosci Methods 130:9–18.

Lyu S, Simoncelli EP (2009) Modeling multiscale subbands of photographic images with fields of gaussian scale mixtures. IEEE Trans Pattern Anal Mach Intell 31:693–706.

Michel M, Geisler WS (2011) Intrinsic position uncertainty explains detection and localization performance in peripheral vision. J Vis 11:18.

Neri P, Levi DM (2006) Receptive versus perceptive fields from the reverse-correlation viewpoint. Vision Res 46:2465–2474.

Nitzany EI, Victor JD (2014) The statistics of local motion signals in naturalistic movies. J Vis 14:4.

Nover H, Anderson CH, DeAngelis GC (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. J Neurosci 25:10049–10060.

Osborne LC, Hohl SS, Bialek W, Lisberger SG (2007) Time course of precision in smooth-pursuit eye movements of monkeys. J Neurosci 27:2987–2998.

Pelli DG (1985) Uncertainty explains many aspects of visual contrast detection and discrimination. J Opt Soc Am A 2:1508–1532.

Pelli DG (1991) Noise in the visual system may be early. In: Computational models of visual processing (Landy MS, Movshon JA, eds), pp 147–152. Cambridge: Massachusetts Institute of Technology.

Perrone JA, Thiele A (2001) Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat Neurosci 4:526–532.

Priebe NJ, Lisberger SG (2004) Estimating target speed from the population response in visual area MT. J Neurosci 24:1907–1916.

Priebe NJ, Cassanello CR, Lisberger SG (2003) The neural representation of speed in macaque area MT/V5. J Neurosci 23:5650–5661.

Rolfs M (2009) Microsaccades: small steps on a long way. Vision Res 49: 2415–2441.

Rucci M, Poletti M (2015) Control and functions of fixational eye movements. Annu Rev Vis Sci 1:499–518.

Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nat Neurosci 9:1421–1431.

Schneeweis DM, Schnapf JL (1995) Photovoltage of rods and cones in the macaque retina. Science 268:1053–1056.
Schrater PR, Knill DC, Simoncelli EP (2000) Mechanisms of visual motion detection. Nat Neurosci 3:64 – 68.
Schrater PR, Knill DC, Simoncelli EP (2001) Perceiving visual expansion without optic flow. Nature 410:816 – 819.
Schütt HH, Wichmann FA (2017) An image-computable psychophysical spatial vision model. J Vis 17:12.
Sebastian S, Geisler WS (2018) Decision-variable correlation. J Vis 18:3.
Sebastian S, Burge J, Geisler WS (2015) Defocus blur discrimination in natural images with natural optics. J Vis 15:16.
Sebastian S, Abrams J, Geisler WS (2017) Constrained sampling experiments reveal principles of detection in natural scenes. Proc Natl Acad Sci U S A 114:E5731–E5740.
Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vision Res 38:743–761.
Sinha SR, Bialek W, de Ruyter van Steveninck R (2018) Optimal local estimates of visual motion in a natural environment. arXiv: 1812.11878, 1–6.
Spering M, Kerzel D, Braun DI, Hawken MJ, Gegenfurtner KR (2005) Effects of contrast on smooth pursuit eye movements. J Vis 5:455–465.
Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9:578–585.
Thibos LN, Ye M, Zhang X, Bradley A (1992) The chromatic eye: a new reduced-eye model of ocular chromatic aberration in humans. Appl Opt 31:3594–3600.
Thompson P (1982) Perceived rate of movement depends on contrast. Vision Res 22:377–380.
Tolhurst DJ, Movshon JA, Thompson ID (1981) The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. Exp Brain Res 41:414 – 419.
Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res 23:775–785.
Tomko GJ, Crapper DR (1974) Neuronal variability: non-stationary responses to identical visual stimuli. Brain Res 79:405–418.
Traer J, McDermott JH (2016) Statistics of natural reverberation enable perceptual separation of sound and space. Proc Natl Acad Sci U S A 113:E7856–E7865.
Turano KA, Heidenreich SM (1999) Eye movements affect the perceived speed of visual motion. Vision Res 39:1177–1187.
Tversky A, Kahneman D (1971) Belief in the law of small numbers. Psychol Bull 76:105–110.
Wei XX, Stocker AA (2015) A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. Nat Neurosci 18: 1509–1517.
Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598 – 604.
Williams DR (1985) Visibility of interference fringes near the resolution limit. J Opt Soc Am A 2:1087–1093.
Wyszecki G, Stiles W (1982) Color science: concepts and methods. In: Quantitative data and formulas. New York: Wiley.
Yates JL, Park IM, Katz LN, Pillow JW, Huk AC (2017) Functional dissection of signal and noise in MT and LIP during decision-making. Nat Neurosci 20:1285–1292.
Ziemba CM, Freeman J, Movshon JA, Simoncelli EP (2016) Selectivity and tolerance for visual texture in macaque V2. Proc Natl Acad Sci U S A 113:E3140–E3149.