Image-Computable Ideal Observers for Tasks with Natural Stimuli

Johannes Burge[1,2,3]

[1]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: jburge@psych.upenn.edu

[2]Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[3]Bioengineering Graduate Group, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Annu. Rev. Vis. Sci. 2020. 6:491–517

First published as a Review in Advance on June 24, 2020

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-030320-041134

Keywords

ideal observer, natural scene statistics, target detection, disparity, motion, blur

Abstract

An ideal observer is a theoretical model observer that performs a specific sensory-perceptual task optimally, making the best possible use of the available information given physical and biological constraints. An image-computable ideal observer (pixels in, estimates out) is a particularly powerful type of ideal observer that explicitly models the flow of visual information from the stimulus-encoding process to the eventual decoding of a sensory-perceptual estimate. Image-computable ideal observer analyses underlie some of the most important results in vision science. However, most of what we know from ideal observers about visual processing and performance derives from relatively simple tasks and relatively simple stimuli. This review describes recent efforts to develop image-computable ideal observers for a range of tasks with natural stimuli and shows how these observers can be used to predict and understand perceptual and neurophysiological performance. The reviewed results establish principled links among models of neural coding, computational methods for dimensionality reduction, and sensory-perceptual performance in tasks with natural stimuli.

INTRODUCTION

Ideal observer: theoretical observer that performs a specific task optimally given specified constraints

Latent variable: property of image or scene to be estimated, discriminated, detected, or categorized

Nuisance stimulus variability: stimulus variability that is irrelevant to the task

Image-computable model: model that takes pixels as inputs and provides estimates or categorical decisions as outputs

Animals evolved visual systems to perform an array of visual tasks that help them to survive and reproduce. The information available for performing these visual tasks is contained in the retinal images. But this information underdetermines the natural scenes that could have given rise to those images. Many animals, from honeybees to humans, nevertheless have the ability to obtain accurate estimates of certain behaviorally relevant properties of the environment. Gaining an understanding of the processes that support this ability is what vision science and visual neuroscience are fundamentally about (Marr 1982).

Ideal observer analysis aims to achieve a computational-level understanding of the processes that support sensory-perceptual performance. Ideal observers are theoretical observers that optimize performance in a specific task by making the best possible use of the information available for processing (see sidebar titled Intuition, Ideal Observers, and Criminal Investigators). The performance of an ideal observer serves as a principled benchmark against which to compare human performance and as a starting point for determining the factors that limit human performance (Geisler 2003, 2011). The computations that instantiate an ideal observer can also provide a normative framework for understanding the response properties of neurons that support perceptual performance in the task (Jaini & Burge 2017).

Sensory perception is hard because natural images vary in many ways that are irrelevant to any given task; many different images are associated with the same value of the task-relevant latent variable (**Figure 1*a***). (The latent variable is the property of the image or scene to be estimated, discriminated, detected, or categorized.) For example, a dog, a cat, and a mouse that are at a given distance and that are all running at the same speed cast very different images on the eyes. This irrelevant stimulus variation is called nuisance stimulus variability, and it generally harms performance. Minimizing the detrimental impact of nuisance stimulus variability on performance is achieved by encoding and processing only those stimulus features that provide information relevant to the task. Models that seek to show how sensory perception works with natural stimuli must specify which stimulus features should be encoded (i.e., which receptive fields should encode stimuli) and specify how those stimulus features should be processed.

It is thus important to distinguish ideal observers that are image computable from those that are not. Image-computable ideal observers take image pixels as input and specify explicitly how the pixels should be encoded, processed, and decoded into estimates. Model observers that are not image computable do not explicitly model how to encode stimuli and cannot directly address how to minimize the impact of nuisance stimulus variability; thus, they are not suitable for gaining a rich, detailed understanding of performance with natural stimuli.

INTUITION, IDEAL OBSERVERS, AND CRIMINAL INVESTIGATORS

Ideal observers specify the theoretically achievable performance limits in a task given a stimulus set and specified constraints. To develop intuition about ideal observers, it is useful to draw an analogy to a detective tasked with investigating a crime. The ideal detective (e.g., Sherlock Holmes) selects the most useful clues from the available evidence given the constraints of the law, properly weights and combines them, and decodes the best possible guess about the most likely culprit. The ideal observer selects the most useful features from the available stimulus given the constraints of the visual system, properly combines them, and decodes the best possible estimate of the task-relevant latent variable. Because of limits imposed by the forensic or sensory evidence, the ideal detective cannot correctly identify the culprit in every case, and the ideal observer cannot accurately estimate the latent variable from every stimulus. But both will be as accurate as possible given the constraints and the available information.

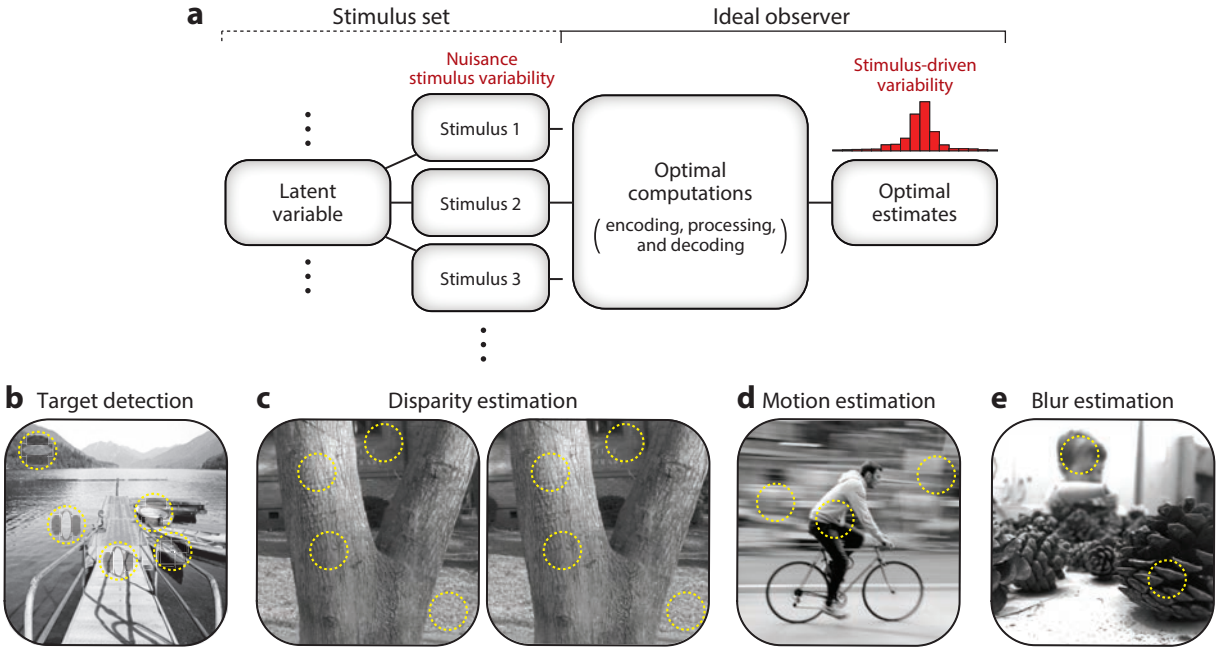

Figure 1

Image-computable ideal observer and tasks with natural stimuli. (*a*) The latent variable, the image or scene property to be estimated, can take on many values. Many different stimuli are associated with each value of the latent variable; this is called nuisance stimulus variability. The optimal computations minimize but cannot entirely eliminate the impact of nuisance stimulus variability on performance. The resulting stimulus-driven variability in the optimal estimates sets a fundamental limit on performance. (*b–e*) Four tasks with natural stimuli: (*b*) target detection, (*c*) disparity estimation, (*d*) motion estimation, and (*e*) blur estimation. In all cases, the ideal observer is constrained to use only local stimulus regions (*yellow circles*) to perform the tasks. Nuisance stimulus variability is depicted by the two stimulus regions that share the same disparity on the tree trunk in panel *c* and by the two stimulus regions that share the same speed on the left and right of the cyclist in panel *d*.

To define an image-computable ideal observer, at least three crucial ingredients are necessary: a well-defined task, a stimulus set, and a specification of the physical and biological constraints (e.g., the optics of the eye) that limit the stimulus information available for processing. The goal is to determine, from these three ingredients, the stimulus features that carry the most useful information for the task and to characterize the probabilistic relationship between those stimulus features and the task-relevant latent variable. The probabilistic relationship between the most useful stimulus features and the latent variable is the primary determinant of the optimal computations that the ideal observer performs to transform each stimulus into the optimal estimate or categorical decision.

Much of what we have learned from ideal observer analysis has had to do with relatively simple tasks and relatively simple stimuli. This review describes recent efforts to develop image-computable (i.e., pixels in, estimates out) ideal observer models for tasks with natural images. First, it provides a brief history of image-computable ideal observers in vision science to provide context for the review's emphasis on image computability. Then, it describes the basic mathematical tools required to specify an ideal observer and shows how to apply these tools to the classic task of detecting a target in noise images. The review finishes with a detailed treatment of four fundamental visual tasks with natural images: target detection, disparity estimation, speed

estimation, and defocus blur (i.e., focus error) estimation (**Figure 1*b–e***). The first task—identifying targets in natural scenes—may be the most basic of all visual tasks and is a necessary component for understanding more complex tasks like visual search (e.g., finding one's keys). The latter three tasks are all fundamental for determining the three-dimensional structure of the environment and the organism's relationship to it. In all cases, it is shown how to determine the optimal computations for the task, given the stimulus set and the constraints, that minimize the detrimental impact of nuisance stimulus variability on performance (**Figure 1*a***). The reviewed results demonstrate that natural stimulus variability shapes the optimal computations, predicts many response properties of neurons in cortex, and dictates the pattern of human performance. The work described in this review represents early steps in what will be a many-years-long effort to extend image-computable ideal observer analysis to more sophisticated tasks with natural stimuli.

Natural stimulus: photographic image or movie of a real-world scene

Natural stimulus variability: nuisance stimulus variability in natural images

BRIEF HISTORY OF IMAGE-COMPUTABLE IDEAL OBSERVERS

From their earliest application up until the late 1980s and early 1990s, essentially all ideal observer models were image computable. They markedly advanced our understanding of how certain forms of stimulus variability (e.g., Poisson noise) and pre-neural biological constraints (e.g., optics, photoreceptor pigments) shape human performance in fundamental visual tasks. Ideal observer analysis predicted the shape of functions governing target detection in white noise (Burgess et al. 1981), intensity discrimination (De Vries 1943, Rose 1948), contrast sensitivity (Banks et al. 1987), and color (wavelength) discrimination (Geisler 1989, Vos & Walraven 1972), as well as how performance in various acuity and hyperacuity tasks changes as a function of stimulus intensity and retinal eccentricity (Geisler 1984; Geisler & Davila 1985; Westheimer 1979, 1982; for a review, see Geisler 2003). These results and many others indicated that, in a wide variety of different tasks, human performance follows a pattern dictated by the physical limits of the tasks.

In the 1990s and early 2000s, image-computable ideal observer analysis fell out of favor, possibly because it was difficult to apply to the more complex tasks (e.g., stereo-depth estimation, motion estimation) that were dominating headlines in certain areas of vision science and visual neuroscience (Britten et al. 1992, DeAngelis et al. 1991, Ohzawa et al. 1990, Weiss et al. 2002). Similarly, as efforts ramped up to link natural image and scene statistics to the design and function of the visual system (Geisler 2008, Simoncelli & Olshausen 2001), widespread interest waned in the application of image-computable ideal observer analysis to the study of visual tasks with those stimuli. A view may have taken hold that the tasks and stimuli of most pressing interest to the field were beyond the reach of image-computable ideal observer analyses (see below).

Many modern Bayesian observer models have since dispensed with image computability. Rather than operate directly on an image-based representation of the stimulus, these models are scaffolded upon abstracted stimulus representations that depend on assumptions about the information that can be encoded from the stimulus. This approach has the benefit of simplifying model construction, and many important results have been obtained with it, especially in the domains of motion estimation and cue integration (Ernst & Banks 2002; Landy et al. 1995, 2011; Weiss et al. 2002). When used as a starting point for modeling and understanding human performance with simple stimuli, non-image-computable ideal observer models have proven very useful. With more complex natural stimuli, common assumptions about the encodable stimulus information are less likely to hold, and the conclusions from these modeling efforts may be called into question. Reinvigorating interest in the development of image-computable ideal observer models, and elaborating and applying these models to more complex stimuli and tasks, will be vitally important for understanding sensory-perceptual processing and performance in the years to come.

BAYESIAN IDEAL OBSERVER MODELS

Bayesian modeling has been integral in shaping modern thinking about how to understand the computations underlying sensory perception (Knill & Richards 1996). The fundamental problem of sensory perception is to estimate behaviorally relevant image and scene properties (i.e., latent variables) from sensory stimuli. Sensory-perceptual tasks are inherently probabilistic because the relationships between sensory stimuli and behaviorally relevant latent variables are ambiguous, uncertain, and noisy. Hence, the tools of probability theory (e.g., Bayes' rule) are appropriate for the quantitative study of sensory perception. It is for these reasons that ideal observer models are fundamentally rooted in Bayesian statistical decision theory.

Consider some latent variable X that the organism is tasked with estimating from a particular sensory stimulus $\mathbf{S}$ arriving at the sense organ. (Bold symbols represent vector-valued quantities. Uppercase symbols represent random variables.) To obtain the estimate, the first crucial step is to compute the posterior probability of the latent variable X_i given the stimulus. The posterior probability of each value of the latent variable given a particular stimulus is specified by Bayes' rule:

$$p(X_i|\mathbf{S}) = \frac{p(\mathbf{S}|X_i)p(X_i)}{p(\mathbf{S})}, \quad 1.$$

where $L(X_i; \mathbf{S}) = p(\mathbf{S}|X_i)$ is the likelihood, $p(X_i)$ is the prior, and $p(\mathbf{S})$ is the probability of the particular stimulus. The last factor, a constant for each stimulus, can be computed as the sum of the likelihoods weighted by the prior probability across each possible value of the latent variable $p(\mathbf{S}) = \sum_j p(\mathbf{S}|X_j)p(X_j)$.

The optimal estimate must then be read out from the posterior. The optimal estimate is the value of X that minimizes the cost

$$\hat{X}_{opt} = \arg\min_{\hat{X}}[C(\hat{X}, X)p(X|\mathbf{S})], \quad 2.$$

where $C(\hat{X}, X)$ is a cost function that specifies the costliness of each estimation error. In general, different errors are associated with different costs, but for experiments conducted in the laboratory, it is often acceptable to assume a cost function that penalizes all errors equally. For such a cost function, Equation 2 simplifies to

$$\hat{X}_{opt} = \arg\max_{X}\left[p(X|\mathbf{S})\right], \quad 3.$$

which states that the optimal estimate $\hat{X}_{opt}$ is the most probable value of X given the stimulus.[1]

Human performance tends to change as a function of the latent variable or some other stimulus property of interest. Hence, to facilitate comparisons to human performance, it is often useful to characterize how ideal observer performance changes with the latent variable. Ideal observer performance can be compared to human performance in many ways. In some experimental paradigms, it is useful to compare optimal estimates to human estimates on a stimulus-by-stimulus basis. In other paradigms, it is more appropriate to compare how the expected value of the optimal estimate changes with the latent variable

$$\bar{\hat{X}}_{opt}(X) = E[\hat{X}_{opt}|X]. \quad 4.$$

[1]Other cost functions correspond to other optimal estimators. The squared error cost function penalizes estimation errors in proportion to the squared error. The absolute error cost function penalizes in proportion to the absolute error. The optimal estimators for these cost functions are the mean of the posterior and the median of the posterior, respectively (Bishop 2006).

In still other paradigms, it is most useful to determine the variance of the optimal estimates (i.e., the ideal decision variable) as a function of the latent variable

$$\sigma^2_{ideal}(X) = \text{var}[\hat{X}_{opt}|X]. \quad 5.$$

The variance of the optimal estimates (i.e., the stimulus-driven variation in the optimal estimates) represents the minimum possible impact of nuisance stimulus variability on performance (see **Figure 1*a***).

Compared to the ideal observer, humans tend to be inefficient. To quantify human performance relative to the ideal, it is common to compute the efficiency

$$\eta = \frac{\sigma^2_{ideal}}{\sigma^2_{human}}, \quad 6.$$

where σ^2_{human} is the variance of the human decision variable. Standard experimental procedures in psychophysics and analytical techniques from signal detection theory can be used to estimate the variance of the human decision variable (Green & Swets 1966). Efficiency quantifies how well a human or some other observer uses the available information relative to the ideal. The difference between the performance levels of the human and the ideal observers is the starting point for additional modeling efforts that seek to determine the sources of human inefficiency.

As stated above, specification of an image-computable ideal observer depends critically on the ability to characterize the probability distributions relating the stimuli and the latent variable. This fact helps emphasize why it is difficult to develop ideal observers with natural stimuli. Despite decades of work, there is still no satisfactory statistical model of natural images. Some notable successes have led to valuable methods for the synthesis of natural-looking textures (Portilla & Simoncelli 2000), but a full model of natural images has proven elusive. The absence of a statistical model of natural stimuli would seem to pose an insurmountable barrier to the development of ideal observers with such stimuli. Recent advances, however, present a way forward.

IDEAL OBSERVERS FOR TARGET DETECTION

Target Detection in White Noise Images

Consider the relatively simple task of detecting a known target $\mathbf{t}$ in white noise (e.g., television snow). In this case, the shape of the optimal encoding receptive field can be derived analytically (**Figure 2*a***). In white Gaussian noise, the optimal receptive field—the most useful stimulus feature to encode—is shaped $\mathbf{f}_{opt} \propto \mathbf{t}$, exactly like the target (Burgess et al. 1981). Given the optimal receptive field, the statistical relationship between the receptive field response and the task-relevant latent variable can be determined directly. Remarkably, the analysis indicates that the response $R = \mathbf{f}^T_{opt}\mathbf{S}$ of the optimal receptive field to any particular stimulus conveys *all* of the useful information in that stimulus about whether the target is present or absent (**Figure 2*b***). Thus, the posterior probability of the latent variable given a particular stimulus is equal to the posterior probability of the latent variable given the response of the optimal receptive field to the stimulus (**Figure 2*c***):

$$p(X_i|\mathbf{S}) = p(X_i|R) = \frac{p(R|X_i)p(X_i)}{p(R)}. \quad 7.$$

This powerful result shows that a single stimulus feature—the feature selected for by the optimal receptive field—is sufficient to support optimal performance in this task. Thus, a high-dimensional

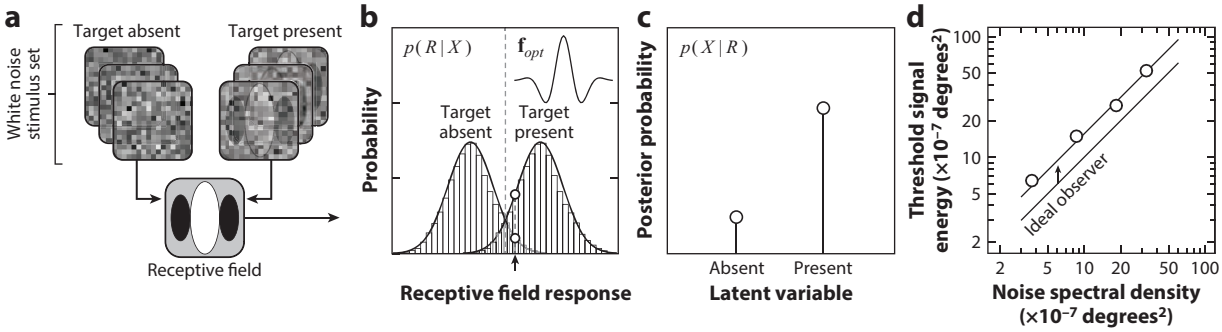

Figure 2

Image-computable ideal observer for target detection in white Gaussian noise. (*a*) The stimulus set is comprised of white noise samples with and without an added target. The optimal receptive field is shaped like the target. (*b*) Receptive field response distributions to target-absent and target-present stimuli. The likelihood that a particular indicated response (*arrow*) was elicited by a target-absent stimulus or a target-present stimulus is represented by the bottom and top circles, respectively. The inset shows a cross-section of the optimal receptive field. (*c*) Posterior probability distribution associated with the observed response indicated in panel *b*. (*d*) Human (*symbols*) and ideal (*line*) observer target detection thresholds as a function of noise spectral density (i.e., noise variance). Scaling the ideal observer thresholds by a single efficiency parameter (*arrow*) nicely accounts for the human data. Panel *d* adapted with permission from Burgess et al. (1981).

stimulus (i.e., multiple pixels) can be characterized by a one-dimensional (i.e., scalar) response without *any* loss of task-relevant information.

The equations underlying the ideal observer indicate that its target-detection thresholds in white noise increase in proportion to the noise variance. Human thresholds follow a similar pattern (**Figure 2*d***). Within the context of the ideal observer analysis, a single efficiency parameter η accounts for the human data over a wide range (Equation 6) (Burgess et al. 1981, Legge et al. 1987, Pelli 1990).

This result is important. It demonstrates that ideal observer analysis can make principled predictions that account for human performance with very few free parameters (e.g., one). Furthermore, it suggests that the general approach described in the Introduction, of first determining the receptive fields that select for the most useful stimulus features and then characterizing the probability distributions that relate the latent variable to the receptive field responses, may prove useful for developing ideal observers for tasks with more natural stimuli.

Target Detection in Correlated Noise Images

White uncorrelated noise has convenient mathematical properties, but it shares few properties with natural images; television snow, a nice example of white noise, looks nothing like images of the real world. This dissimilarity is partly due to the fact that natural images are spatially correlated: Pixels in similar spatial locations tend to have similar values. Natural images are well-characterized by 1/f amplitude spectra (Field 1987). This property of natural images means that, if one frequency is 10 times higher than another, then it will tend to have 10 times less contrast in the image. In other words, more stimulus energy tends to be associated with coarse image detail than with fine image detail.

In a stimulus set defined by correlated Gaussian noise, the shape of the optimal encoding receptive field can again be derived analytically. In correlated Gaussian noise, the optimal receptive field $\mathbf{f}_{opt} \propto \mathbf{C}^{-1}\mathbf{t}$ is shaped like the target modified (i.e., whitened) by the noise covariance. The noise covariance of the stimuli can be set such that the noise amplitude spectrum matches the

1/f amplitude spectra of natural images. Just as with target detection in white noise, in correlated noise a single receptive field extracts all of the useful information from each stimulus about the presence of the target. Human target detection performance in correlated noise again adheres to the pattern predicted by the ideal observer (Abbey & Eckstein 2007, Bradley et al. 2014, Najemnik & Geisler 2005).

Nuisance stimulus variability and the stimulus encoding that is performed by the receptive field jointly impact the variance of the receptive field responses and subsequent performance in the task. Systematic misuse of the available stimulus information deteriorates performance relative to the ideal. If the receptive field is optimal, then the response distributions have the minimum possible variance. If the receptive field is suboptimal, then it encodes the wrong stimulus feature, response variance and stimulus-driven variability in the decision variable increases, and performance decreases. Receptive field correlation,[2]

$$\rho_{\mathbf{f}} = \mathbf{f}_{opt}^{T}\mathbf{f}_{subopt}, \qquad 8.$$

quantifies the (cosine) similarity of the optimal and suboptimal receptive fields. Interestingly, receptive field correlation directly predicts the efficiency

$$\eta = \rho_{\mathbf{f}}^{2} \qquad 9.$$

of an observer using a suboptimal receptive field under certain conditions (Chin & Burge 2020). Receptive field correlation can therefore provide a precise measure of the performance loss caused by a suboptimal receptive field that encodes the wrong feature. For example, an observer that uses the receptive field that is optimal in white noise to detect a 1.5 octave bandwidth Gabor target in 1/f noise has a receptive field correlation of $\rho_{\mathbf{f}} = 0.95$ and an efficiency of $\eta = 0.90$ relative to the ideal observer for 1/f noise. The fact that efficiency and receptive field correlation are related can be useful for diagnosing the cause(s) of human inefficiency (see below).

Target Detection in Natural Images

This section addresses the concern that the general approach that works with noise stimuli cannot be adapted to work with more natural stimuli. Recent work suggests that there is cause for optimism. Consider the task of detecting a target in a natural image (Bex & Makous 2002, Bex et al. 2009, Bradley et al. 2014, Hansen & Essock 2004, Schütt & Wichmann 2017). Natural images vary in far more complex ways than do images of white and 1/f noise. Can a well-chosen receptive field extract all of the information that is useful for detecting a target in natural images, and are its responses Gaussian distributed to natural stimuli? Without some modification of the stimulus-encoding model, the answer is no. If the receptive fields that are optimal for white and 1/f noise images are applied directly to natural images, then the response distributions are heavy tailed and highly non-Gaussian (Olshausen & Field 1997). Poor target detection performance results. [The observation of heavy-tailed response distributions undergirds a great deal of work on the efficient coding hypothesis (for a review, see Simoncelli & Olshausen 2001).] Good detection performance requires a more sophisticated encoding model (**Figure 3*a***). How can this more sophisticated encoding model be determined? Without a mathematical description of natural images, it seems that the development of an image-computable ideal observer for target detection in natural images cannot be successfully formulated. A recent advance is relevant to this issue.

[2]The expression in Equation 8 assumes that the receptive fields are unit vectors. The general expression for receptive field correlation is $\rho_{\mathbf{f}} = \mathbf{f}_{opt}^{T}\mathbf{f}_{subopt}/||\mathbf{f}_{opt}^{T}||||\mathbf{f}_{subopt}||$.

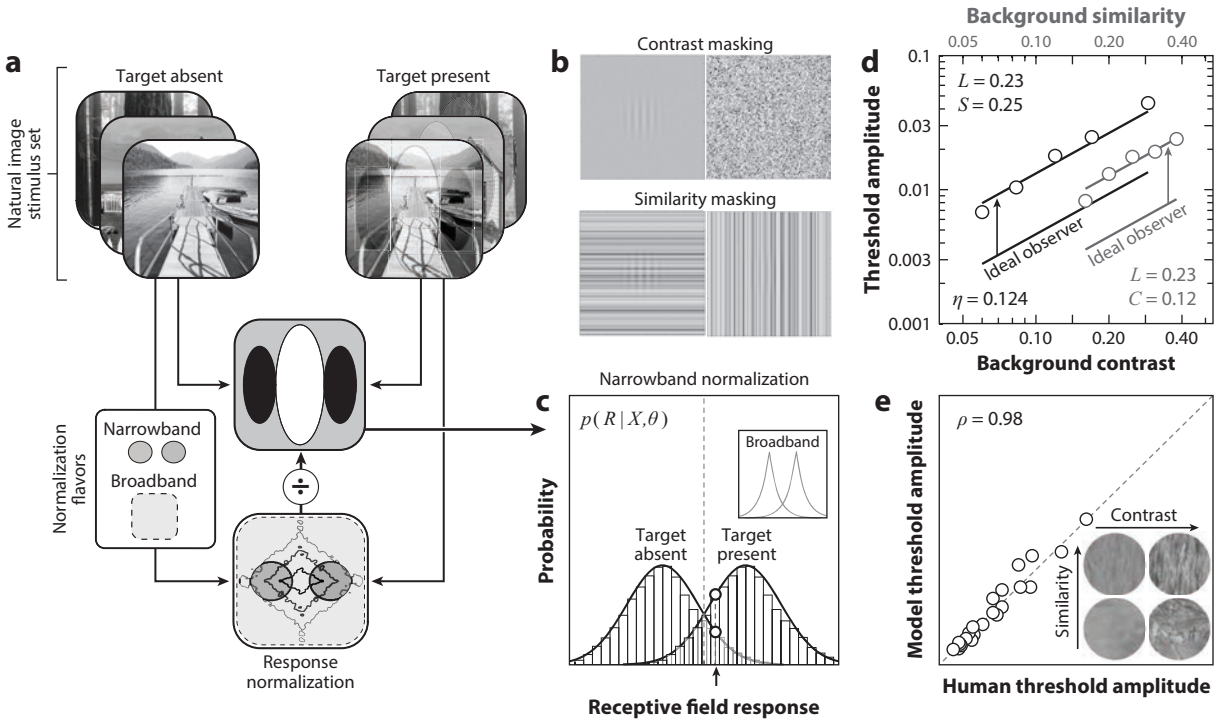

Figure 3

Image-computable ideal observer for target detection in natural images. (*a*) The stimulus set is comprised of natural images with and without an added grating target. Normalizing the response of a linear receptive field by the local luminance, contrast, and spatial similarity to the target maximizes target detection performance. (*Bottom*) The spatial frequency amplitude spectrum of the natural image (*diamond contours*) and the regions of the spectrum that contribute to the narrowband and broadband normalization terms (*shaded regions*). With narrowband normalization, the normalization term is determined by the stimulus energy falling within the spatial frequency passband of the receptive field (*circular shaded regions*). With broadband normalization, the normalization term is determined by all of the energy in the stimulus (*rounded dashed square region*). (*b*) The detectability of a given target decreases (i.e., detection thresholds increase) with increases in image contrast (*top*) and image similarity to the target (*bottom*). Target detectability also decreases with increased luminance (not shown). (*c*) The normalized responses to natural images are Gaussian distributed if the receptive field responses are narrowband normalized. The likelihood that a particular indicated response (*arrow*) was elicited by a target-absent stimulus or a target-present stimulus is represented by the bottom and top circles, respectively. Inset shows the heavy-tailed response distributions that result if broadband normalization, or if no normalization at all, is used. (*d*) Human (*symbols*) and ideal (*lines*) observer detection thresholds for a grating target as a function of contrast (*black*) at fixed luminance and similarity ($L = 0.23$, $S = 0.25$) or as a function of similarity (*gray*) at fixed luminance and contrast ($L = 0.23$, $C = 0.12$). Ideal observer thresholds are derived from natural scene statistics. Scaling the ideal thresholds by efficiency (*arrow*) nicely accounts for the human data. (*e*) Correlation between model and human target detection thresholds (no free parameters). The inset shows example natural image patches with different contrasts and similarities to the target. Thresholds increase (i.e., detectability decreases) as luminance (not shown), contrast, and similarity increase. Panels *b*, *d*, and *e* adapted with permission from Sebastian et al. (2017). To maintain visual similarity to other figures, the target is shown in these panels as vertically oriented; it is horizontally oriented in the original manuscript.

Sebastian et al. (2017) analyzed the natural statistics of image properties that are known to impact human target detection performance. Target detectability decreases as the image that the target is embedded in increases in luminance (Mueller 1951), contrast (Burgess et al. 1981, Legge & Foley 1980), and spatial similarity to the target (Campbell & Kulikowski 1966, Stromeyer & Julesz 1972, Watson & Solomon 1997) (**Figure 3*b***). Sebastian et al. examined the joint impact of these three stimulus factors in natural images on the responses of a target-shaped receptive

field. They also conducted a large psychophysical experiment to measure human target detection performance in natural images.

First, a stimulus set was obtained by randomly sampling millions of image patches from photographs of natural scenes. Each patch was stored in one of 1,000 bins; all stimuli within a given bin shared a unique combination of luminance, contrast, and target similarity. Next, the responses of a target-shaped receptive field were computed to all stimuli within a bin. Within a bin, the responses across stimuli were Gaussian distributed, $p(R|X, \theta) \sim gauss(\cdot)$, where $\theta = [L, C, S]$ is a vector representing the luminance L, contrast C, and similarity S of the stimuli in the bin (**Figure 3*c***). Across bins, the response standard deviation $\sigma \propto L \times C \times S$ increases in proportion to the product of the stimulus attributes. The image statistics thus specify a multidimensional Weber's law for target detection in natural scenes.

Human performance is beautifully predicted by this multidimensional Weber's law (**Figure 3*d***); a single free parameter—efficiency—nicely predicts human thresholds for all tested conditions, similar to the results with white noise and 1/f noise (see **Figure 2*d***). The model accounts for 96% of the variance in the human thresholds across all tested combinations of luminance, contrast, and similarity (**Figure 3*e***). The stimulus factors that have long been known to limit target detection performance (i.e., luminance, contrast, and similarity), and Weber's law itself, have their roots in the statistics of natural images.

The dependence of human performance on these three stimulus factors is naturally mediated via response normalization (gain control), a widely observed neural computation (Albrecht & Geisler 1991, Carandini & Heeger 2012, Coen-Cagli et al. 2015, Heeger 1992). Normalization updates the expression for receptive field response to $R = \mathbf{f}^T\mathbf{S}/\sigma$ and bakes the relevant image statistics into the encoding model. Normalizing by the product of luminance, contrast, and similarity is equivalent to normalizing by the square root of the stimulus energy in the passband of the receptive field (Iyer & Burge 2019). This type of normalization is known as narrowband normalization (see **Figure 3*a***); including it in the encoding model optimizes target detection performance in natural images. Normalizing by luminance and contrast, but not by similarity, is equivalent to normalizing by all of the stimulus energy and is known as broadband normalization. Using broadband normalization, or failing to normalize altogether, yields heavy-tailed response distributions (**Figure 3*c***, *inset*) and harms target detection performance (Sebastian et al. 2017).

Sebastian et al. (2017) make a strong case that normalization, a ubiquitous property of neural response, is an evolutionary adaptation that functions to optimize target detection performance in natural scenes. Their work suggests that narrowband response normalization is a normative (i.e., ideal) computation for the task of target detection. For other tasks, it may be more appropriate to consider normalization as a biological constraint that is built into the stimulus-encoding model.

Does narrowband normalization occur in cortex? The current consensus, which is based almost entirely on experiments with simple laboratory stimuli, is that broadband normalization (i.e., normalization by luminance and contrast, but not by similarity) provides a better description of neural responses than does narrowband normalization (Busse et al. 2009, Carandini et al. 1997, Heeger 1992), although some results suggest that the story is more nuanced (Cavanaugh et al. 2002, Ruff et al. 2016). Recent results with natural stimuli suggest that narrowband (i.e., feature-specific) normalization may provide a better description than broadband normalization (Burg et al. 2019). It will be interesting to see whether the consensus evolves as additional neurophysiological experiments are performed with natural images.

For all the important insights provided by Sebastian et al. (2017), one shortcoming of this work is that it assumed a receptive field $\mathbf{f} \propto \mathbf{t}$ that is shaped like the target. A target-shaped receptive field, while optimal for white Gaussian noise, is almost certainly suboptimal for detecting targets

in natural images, just as it is in 1/f noise (see above). To exactly quantify the degree of its suboptimality, methods are needed for determining optimal receptive fields for tasks with natural stimuli.

LEARNING OPTIMAL RECEPTIVE FIELDS WITH NATURAL IMAGES

A recently developed Bayesian statistical learning tool called accuracy maximization analysis (AMA) is designed to find the optimal receptive fields for a task via numerical methods, even if an analytical derivation proves elusive (Burge & Jaini 2017, Geisler et al. 2009, Jaini & Burge 2017). To find the optimal receptive fields, the method searches the space of possible receptive fields with a closed-form approximation of the Bayes optimal decoder, given a labeled training stimulus set, a receptive field response (i.e., encoding) model, and a cost function. The encoding model should incorporate narrowband normalization to improve the generality of AMA across both noise and natural images (Iyer & Burge 2019, Sebastian et al. 2017). The optimal receptive fields are those that minimize the cost function for stimuli in the training stimulus set. Receptive fields are learned one at a time or in groups until performance improvements are negligible. If the training stimulus set is sufficiently large and representative, then the receptive fields generalize well to test stimulus sets (i.e., essentially identical receptive fields would be learned with other representative stimulus sets). The receptive fields are learned with a closed-form expression for the optimal decoder, but this decoder is restricted for use with the training stimuli—a somewhat unusual feature of the method—and can generally not be used with test stimuli (but see Jaini & Burge 2017). Thus, a general decoder must be determined that can be used with arbitrary test stimuli. Just as with the models described above, the key step in completing the specification of the ideal observer (i.e., determining the general decoding rules) is to characterize the probability distributions of receptive field response for each value of the latent variable. Once the response distributions are characterized, the likelihoods can be computed, and the posterior probability distribution over the latent variable can be determined for the receptive field response to any arbitrary stimulus.

To develop confidence that AMA is well suited to the target detection tasks that are considered above, it is useful to examine whether the receptive fields $\hat{\mathbf{f}}_{opt}$ learned by AMA match the optimal receptive fields in cases where the optimal receptive fields can be derived analytically. AMA correctly learns the optimal receptive field for detecting targets in white and correlated Gaussian noise (**Figure 4*a*,*b***). This technique is therefore promising for learning the receptive fields that optimize target detection performance in natural images. From 5,000 natural images (half of which had an embedded Gabor target), the optimal receptive field was estimated using AMA with an encoding model that includes narrowband normalization (**Figure 4*c***). If this learned receptive field is indeed optimal, then its receptive field correlation ($\rho_f = 0.92$) with the receptive field used by Sebastian et al. (2017) implies that the human efficiency relative to the model observer reported in their paper ($\eta = 0.124$) was approximately 18% higher than human efficiency relative to the true ideal (see Equation 9, **Figure 3*d***).

With AMA, one receptive field, or a small set of receptive fields, encodes and transmits information about useful stimulus features for further processing. Thus, ideal observers that are developed with the help of AMA are most conservatively described as ideal given the hard constraint that the receptive fields encode the stimulus with a particular or a parameterized set of encoding models. If the assumed encoding models are well matched to properties of real sensory-perceptual systems, then ideal observer models developed with the help of AMA have the potential to be quite useful for the scientific study of those systems.

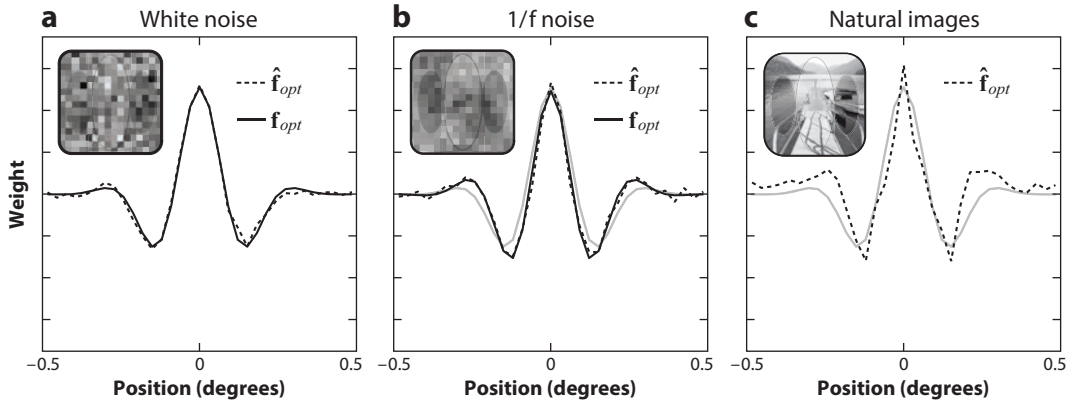

Figure 4

Evaluating the optimality of accuracy maximization analysis (AMA) receptive fields. Optimal receptive fields for target detection in (*a*) white noise, (*b*) 1/f noise, and (*c*) natural images are shown. The solid curve is the optimal receptive $\mathbf{f}_{opt}$ that was derived analytically; the dashed curve is the receptive field $\hat{\mathbf{f}}_{opt}$ learned numerically with AMA. When the optimal receptive fields can be derived analytically (panels *a* and *b*), the similarity of the receptive fields obtained analytically and numerically is excellent ($\rho_{\mathbf{f}} > 0.99$). The optimal receptive field for white noise is replotted in panels *b* and *c* as a gray curve for comparative purposes.

IDEAL OBSERVERS FOR ESTIMATION TASKS WITH NATURAL STIMULI

Ideal observers have recently been developed for the tasks of estimating binocular disparity, motion speed, and focus error (i.e., defocus blur) from natural images. In the target-detection task described above, only a single receptive field is required to extract the latent variable of interest. But these estimation tasks are more complex than target detection. More than one stimulus feature, and thus more than one receptive field, is useful for these tasks. In this case, the expression for the posterior probability of the latent variable is

$$p(X_i|\mathbf{S}) \cong p(X_i|\mathbf{R}) = \frac{p(\mathbf{R}|X_i)p(X_i)}{p(\mathbf{R})}, \qquad 10.$$

where $\mathbf{R} = [R_1, R_2, \ldots, R_N] = \hat{\mathbf{f}}_{opt}^T \mathbf{S}/\sigma$ are the responses of a set of N optimal task-specific receptive fields that incorporate narrowband response normalization into the stimulus-encoding model.

To develop the ideal observer for each task, the authors of the studies described below obtained a labeled training set of natural images, determined the optimal receptive fields using AMA, and then determined the optimal processing and decoding rules from the response distributions associated with each value of the latent variable (see **Figure 1*a***). Ideal observer performance was then compared to human performance. Additionally, the receptive fields and the computations associated with decoding the responses into estimates were compared to the properties of neurons that may underlie performance in these tasks. Image-computable ideal observers for estimation tasks with natural images provide a rich explanatory framework for understanding observed properties of sensory-perceptual performance and the underlying neurophysiology. These developments suggest that a bright future is in store for ideal observer analysis of natural tasks with natural stimuli.

Ideal Observer for Disparity Estimation

The ability of animals to estimate the three-dimensional structure of the environment is critical for many aspects of survival. Depth differences in the scene cause image differences in the left and right eyes due to their different vantage points on the scene. These image differences—the binocular disparities—are powerful cues to depth. Mammals (Blakemore 1970, Scholl et al. 2013), birds (Fox et al. 1977, van der Willigen 2011), cephalopods (Feord et al. 2020), and insects (Nityananda et al. 2016, Rossel 1983) all use disparity to estimate depth. However, before disparities can be used to estimate depth, the disparities themselves must be estimated. The manner in which visual systems estimate disparity is a long-studied topic in vision science (Banks et al. 2004, Burge & Geisler 2014, Chauhan et al. 2018, Cormack et al. 1991, Cumming & DeAngelis 2001, DeAngelis et al. 1991, Goncalves & Welchman 2017, Hibbard 2008, Julesz 1964, Ogle 1952, Ohzawa et al. 1990, Parker 2007, Qian 1997, Read & Cumming 2007, Tyler & Julesz 1978, Welchman 2016, Wheatstone 1838). The fundamental challenge in disparity estimation is to determine what local region of the left-eye image goes with what local region of the right-eye image; the problem of estimating binocular disparity is thus known as the correspondence problem.

The current consensus is that the human visual system estimates binocular disparity (i.e., solves the correspondence problem) via local cross-correlation (Banks et al. 2004, Cormack et al. 1991, Tyler & Julesz 1978). The response properties of binocular neurons in cortex are commonly described by the disparity energy model, which proposes that classic disparity-selective neurons are obtained via quadratic combination of Gabor receptive field responses (Cumming & DeAngelis 2001, DeAngelis et al. 1991, Ohzawa et al. 1990). The disparity energy model has been shown to be the computational equivalent of local cross-correlation (Anzai et al. 1999b). However, the disparity energy model does not indicate how information in different spatial frequency channels should be integrated (Read & Cumming 2007), nor does it provide an account of why disparity discrimination thresholds rise exponentially as disparity increases, two longstanding questions in the field. Image-computable ideal observers for disparity estimation provide principled answers to both of these questions.

To develop an image-computable ideal observer for disparity estimation, a large stereo-image database of natural scenes was collected; each image had laser-based distance measurements coregistered to each image pixel (Burge et al. 2016). A novel routine that makes use of this distance data to establish corresponding points was used to sample stereo-image patches with known amounts of disparity at the center pixel of each patch (Iyer & Burge 2018a). These images were then passed through the optics of the eye and sampled by the photoreceptors, both of which shape and constrain the information available for further processing. Optimal receptive fields were learned on the training stimulus set with AMA using an encoding (i.e., receptive field response) model that incorporates narrowband response normalization (Burge & Geisler 2014, Iyer & Burge 2018b). The optimal receptive fields share many properties with the receptive fields of binocular simple cells in cortex (**Figure 5*a*,*b***). Optimal receptive fields select for a range of spatial frequencies, have approximately 1.5 octave bandwidths, and exhibit both phase and position coding like that observed in cortex (Burge & Geisler 2014). These receptive field properties are well-matched to what is observed in cat and macaque cortex (Anzai et al. 1999a, Burge & Geisler 2014, Cumming & DeAngelis 2001, De Valois et al. 1982, DeAngelis et al. 1991, Prince et al. 2002, Ringach 2002).

To determine how to optimally decode the receptive field responses into estimates, the response distributions must be characterized. Across large numbers of stimuli sharing the same disparity, the responses $p(\mathbf{R}|X_i)$ are mean zero and Gaussian distributed with response covariance matrices Σ_i that change systematically with disparity (Burge & Geisler 2014, Burge & Jaini 2017) (**Figure 5*c*,*d***). Just as with target detection, the probability distributions that characterize the

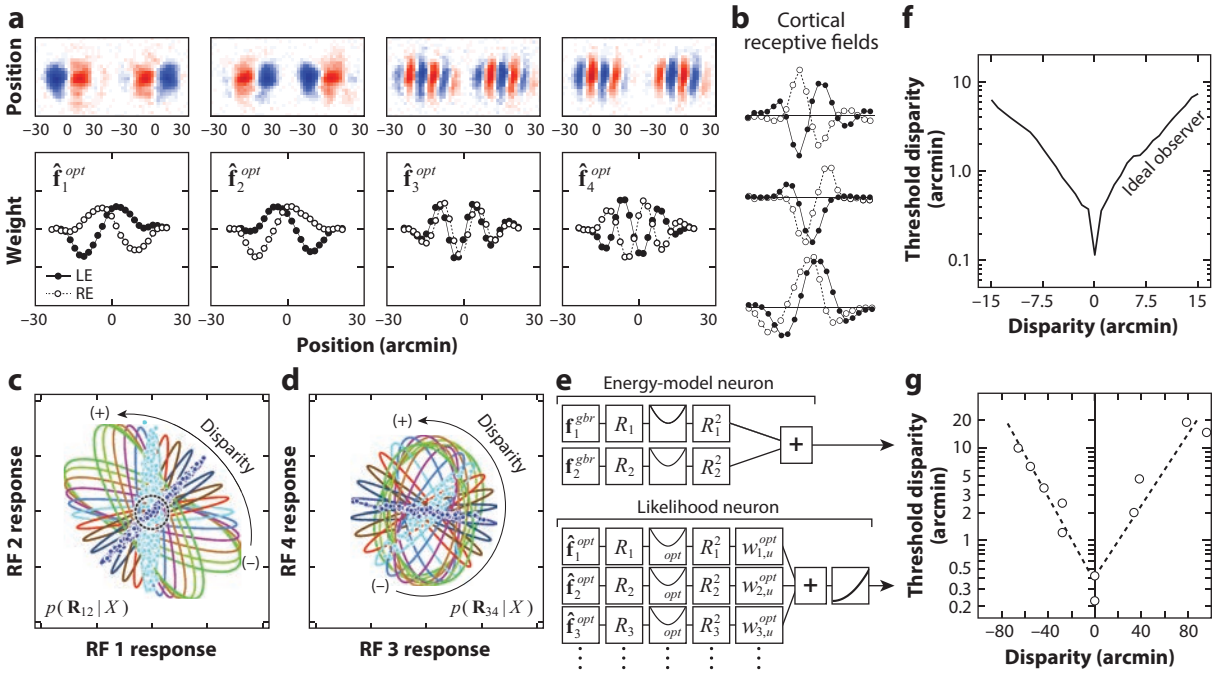

Figure 5

Ideal observer for binocular disparity estimation. (*a*) Optimal binocular receptive fields (RFs), learned with accuracy maximization analysis (AMA), for disparity estimation in natural stereo-images. Left-eye (LE) and right-eye (RE) RF components (*top*) and horizontal slices through the RF (*bottom*) are shown. The RFs select for a range of spatial frequencies and phase differences similar to those observed in disparity-sensitive neurons in cortex. Panel adapted from Iyer & Burge (2018b). (*b*) Binocular RFs in the cat visual cortex. Panel adapted with permission from DeAngelis et al. (1991). (*c*) Response distributions of RFs 1 and 2 to stimuli with each of multiple disparities (*colors*). Symbols represent responses to individual stimuli; ellipses represent the best-fit Gaussian to each response distribution. Stimuli that cause responses near zero (*dashed black circle*) cannot be accurately estimated with RFs 1 and 2. Panel adapted from Burge & Geisler (2014). (*d*) Response distributions of RFs 3 and 4. Symbols represent the responses of RFs 3 and 4 to stimuli that elicited responses within the dashed black circle in panel *c*. Thus, RFs 3 and 4 can be used to discriminate the disparity of the many stimuli that RFs 1 and 2 cannot. However, RFs 3 and 4 in isolation are substantially worse than RFs 1 and 2 at discriminating large disparities, as indicated by the heavy overlap of the response distributions corresponding to those disparities. Fortunately, the ideal observer optimally uses all receptive fields in tandem. (*e*) Computations underlying a standard disparity energy-model neuron and a likelihood neuron. In the energy model, the binocular RFs $\mathbf{f}_{gbr}$ are assumed to be shaped like Gabors. In the likelihood neuron, the RFs $\hat{\mathbf{f}}_{opt}$ are the optimal binocular RFs for the task as estimated by AMA. (*f*) Ideal observer-predicted thresholds in natural scenes. Panel adapted with permission from Burge & Geisler (2014). (*g*) Human disparity discrimination thresholds for two-point stimuli. Panel adapted with permission from Blakemore (1970).

receptive field responses link the natural image statistics to the computations that support optimal performance in the task. As discussed below, the response distributions specify how to construct neurons that are both maximally selective for disparity and maximally invariant to nuisance stimulus variability in natural stereo-images.

Computing the likelihood that a particular value of the latent variable elicited the observed response is the key computation required for converting the receptive field responses into optimal estimates. Each response distribution dictates the details of the computations. The likelihood of a particular disparity $L(X_u; \mathbf{R}) = gauss(\mathbf{R}; \mathbf{0}, \Sigma_u)$ is obtained by evaluating the Gaussian corresponding to that disparity given the observed receptive field response. Thus, computing the likelihood $L(X_u; \mathbf{R}) \propto \exp[-0.5\mathbf{R}^T \Sigma_u \mathbf{R}] = \exp[Q_{\mathbf{w}_u}(\mathbf{R})]$ requires combining the receptive field

responses in a weighted quadratic sum and then passing the sum through a static exponential output nonlinearity, where the weights $\mathbf{w}_u = f(\Sigma_u)$ are simple functions of the response covariance associated with a particular disparity (Burge & Geisler 2014, Jaini & Burge 2017).

The response of a hypothetical neuron—a likelihood neuron—that performs these quadratic computations represents the likelihood $R_u^L \propto \exp[Q_{\mathbf{w}_u}(\mathbf{R})]$ that a particular disparity elicited the observed receptive field response. Appropriately changing the weights constructs a likelihood neuron with a different preferred disparity. The likelihood neurons are unimodal, are approximately log-Gaussian shaped, and are largely invariant to nuisance stimulus variability (see below). Computational models of estimation from neural populations often incorporate the assumption that each neuron is invariant and unimodally tuned to the stimulus property of interest (Girshick et al. 2011, Jazayeri & Movshon 2006, Ma et al. 2006). But the computations that lead to invariant unimodal tuning are often not specified. The ideal observer computations described in this review constitute a recipe for how to construct neurons that are selective for disparity and maximally invariant to nuisance stimulus variability.

Each likelihood neuron effectively carries out an augmented disparity-energy-like computation (Burge & Geisler 2014, Burge & Jaini 2017) (**Figure 5*e***). Because the optimal receptive fields select for different spatial frequencies, the quadratic combination rules provide a principled specification for how to weight (i.e., combine) image information across spatial frequency. In accordance with intuition, these rules indicate that receptive fields that select for high-frequency image features are less useful for coding large disparities and thus should receive less weight for these disparities.

The ideal observer computations also provide a normative justification for discrepancies between neurons in the cortex and the classic disparity energy model. For example, most disparity-selective complex cells in macaque V1—possible neurophysiological analogs of likelihood neurons—are driven by more than two stimulus features (Tanabe et al. 2011). If the analogy holds approximately, then one can conclude that more than two stimulus features (i.e., receptive fields) drive the responses of most disparity-selective complex cells because more than two stimulus features provide useful information about disparity (Burge & Geisler 2014, Tanabe et al. 2011).

Finally, the ideal observer predicts disparity discrimination thresholds that rise exponentially as disparity increases (**Figure 5*f***). This pattern of discrimination thresholds was first reported in humans with two-point stimuli (Blakemore 1970) (**Figure 5*g***) and has been widely replicated with bar stimuli (McKee et al. 1990), grating stimuli (Badcock & Schor 1985), random-dot stereograms (Schumer & Julesz 1984, Stevenson et al. 1992), and natural stimuli (White & Burge 2018). This is the first principled account of why disparity discrimination thresholds follow this exponential law. Thus, the ideal observer for disparity estimation provides a principled account of how natural image statistics shape the pattern of human disparity (i.e., stereo-depth) discrimination thresholds and numerous properties of disparity-sensitive neurons in cortex. The qualitative similarities between the ideal and human discrimination performance could be probed more rigorously with an experiment in which human and ideal observers view matched stimuli. Such an analysis has been performed in the domain of motion estimation.

Ideal Observer for Speed Estimation

Accurate estimation of the speed of retinal image motion is critical for estimating the motion of objects and the self through the environment. A great deal of work has been devoted to understanding how retinal image motion is estimated by the visual system (Adelson & Bergen 1985, Britten et al. 1992, Gekas et al. 2017, Heeger 1987, Jogan & Stocker 2015, Kane et al. 2011, Nishimoto & Gallant 2011, Simoncelli & Heeger 1998, Sinha et al. 2018, Stocker & Simoncelli 2006, Weiss et al. 2002). But as with the estimation of binocular disparity, comparatively little

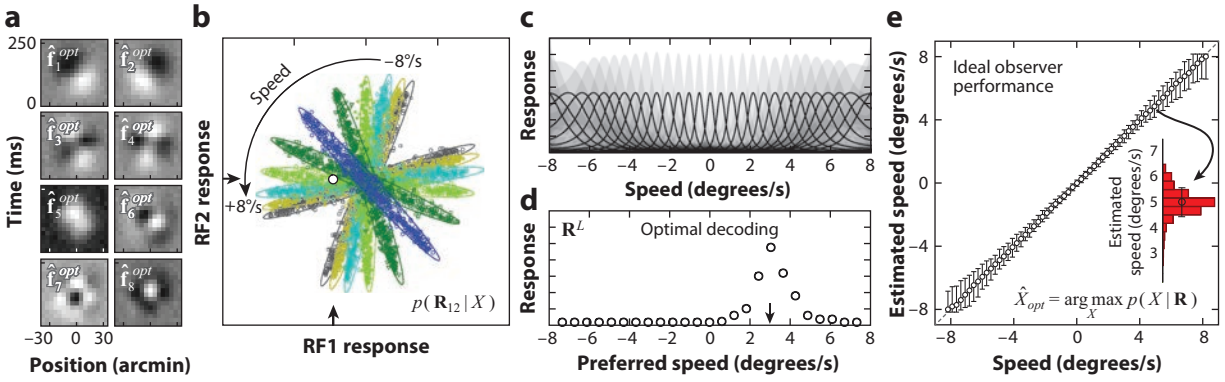

Figure 6

Ideal observer for retinal speed estimation. (*a*) Space-time receptive fields (RFs) using accuracy maximization analysis (AMA). (*b*) RF response distributions for the first two RFs across a range of speeds (*colors*). (*c*) Speed tuning curves of a population of likelihood neurons that are constructed from all eight RFs in panel *a* and that have speed preferences across the same range of speeds represented in panel *b*. Shaded regions represent ±1 standard error on the expected response due to natural stimulus variability. (*d*) Likelihood neuron population activity $\mathbf{R}^L$ in response to a particular natural image movie drifting at a particular speed. The optimal estimate can be decoded from the population response (*arrow*). The joint response of RF 1 and RF 2 to the movie associated with this population response is shown in panel *b* (*arrows* and *white circle*). (*e*) Ideal observer speed estimation performance. The red histogram indicates, for one speed, the stimulus-driven variation σ^2_{ideal} in the ideal observer decision variable (Equation 5). Figure adapted from Chin & Burge (2020).

effort has been spent on developing ideal observers for motion estimation with natural stimuli. This section briefly discusses the development of an ideal observer for speed estimation with natural image movies (Burge & Geisler 2015), but the focus of the section is to show how the ideal observer can be used to determine the cause of human inefficiency in a task (Chin & Burge 2020).

First, an image-computable ideal observer for speed estimation with local patches of naturalistic image movies was developed with the methods described above (Burge & Geisler 2015, Chin & Burge 2020). Results are analogous to those with disparity. Optimal space-time receptive fields are similar to those observed in cortex (**Figure 6*a***). The receptive field responses to naturalistic movies having the same speed are Gaussian distributed with covariance matrices that change systematically with speed (**Figure 6*b***). Hypothetical likelihood neurons for speed, which carry out augmented motion-energy-like computations (Adelson & Bergen 1985), have unimodal speed tuning curves that are approximately log-Gaussian in shape (**Figure 6*c***). These tuning curves are shaped similarly to the speed tuning curves of neurons in macaque area MT (Liu & Newsome 2006, Nover et al. 2005). In addition, just as with neurons exhibiting good disparity selectivity, motion-selective neurons in cortex tend to be driven by more than two stimulus features (Rust et al. 2005). For any given stimulus (see **Figure 6*b***), the evoked response across a population of likelihood neurons represents the likelihood function (**Figure 6*d***). Assuming a uniform prior and the cost function in Equation 3, the peak of the population response represents the optimal estimate (see **Figure 6*d***); assuming a nonuniform prior [e.g., a speed-zero prior (Weiss et al. 2002)] has little effect on these results (Burge & Jaini 2017). The performance of the ideal observer across many thousands of stimuli is shown in **Figure 6*e***. The variance of ideal observer speed estimates σ^2_{ideal} reflects the irreducible impact of nuisance stimulus variability on speed estimation performance (**Figure 6*e***, red histogram; see also **Figure 1*a***).

To determine whether the ideal observer predicts human performance, ideal and human speed discrimination were measured with matched natural stimuli in a two-interval forced choice experiment (Burge & Geisler 2015, Chin & Burge 2020). The ideal estimates for multiple stimuli

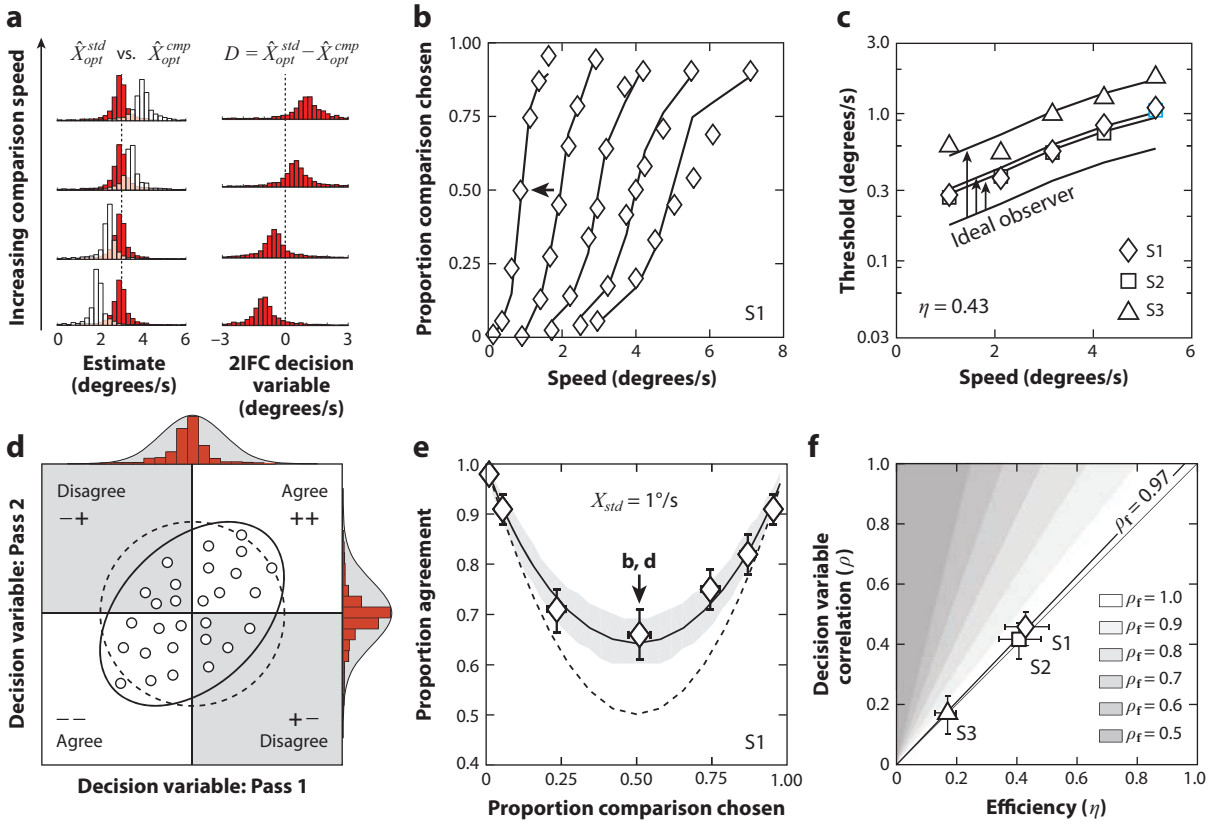

Figure 7

Using the ideal observer for speed estimation to predict and understand human speed discrimination performance. (*a*) Ideal observer estimates (*left*) and corresponding decision variable (*right*) distributions in a two-interval forced choice (2IFC) speed discrimination experiment. (*b*) Human psychometric data (*symbols*) with ideal observer predictions scaled by human efficiency (*curves*). A single efficiency parameter matches the slopes of all five functions to the human data. (*c*) Human and ideal speed discrimination thresholds with naturalistic image movies. (*d*) Decision variable correlations that equal efficiency (*solid ellipse*) and zero (*dashed circle*). Response agreement increases as decision variable correlation increases. (*e*) Proportion response agreement as a function of proportion comparison chosen for the left-most psychometric function in panel *b*. Similar results are found across all tested speeds and all three observers. The solid curve represents the predicted response agreement if efficiency equals decision variable correlation. The dashed curve represents the binomial prediction if stimulus-driven variability is negligible, and decision variable correlation equals zero. (*f*) For all human observers, decision variable correlation almost exactly equals efficiency. This result implies that humans use (encode and process) the available stimulus information near optimally (i.e., receptive field correlation ρ_f equals 1.0). The result also implies that human inefficiency is due near exclusively to noise or to other factors that are uncorrelated with the stimulus. Figure adapted from Chin & Burge (2020).

at the standard and comparison speeds ($\hat{X}_{opt}^{std}$ and $\hat{X}_{opt}^{cmp}$, respectively) in each of four conditions help illustrate why discrimination improves as speed differences increase (**Figure 7*a***). In the context of this task, the fundamental performance limits imposed by nuisance stimulus variability are reflected in the variance of the ideal decision variable $D = \hat{X}_{opt}^{cmp} - \hat{X}_{opt}^{std}$ in each condition (e.g., **Figure 7*a***). Representative psychometric data from one human subject is shown in **Figure 7*b***. The variance of the human speed estimates σ_{human}^2 can be derived from this data using standard tools from signal detection theory (Green & Swets 1966). Across all conditions, a single free parameter—efficiency—accounts for 96% of the variance in the raw human psychometric data

($\eta = 0.43$; see **Figure 7*b***, Equation 6). The ideal observer also predicts the pattern of thresholds in each human observer, just as it does for target detection in noise and natural images (**Figure 7*c***). These results demonstrate that the ideal observer provides a quantitative account of the pattern of human performance, but they do not indicate why humans underperform the ideal.

To determine the cause(s) of human inefficiency, each observer repeated the original experiment such that two responses were collected for each unique trial, a so-called double pass experiment (Burgess & Colborne 1988). Different performance-limiting factors have signature impacts on the correlation of the human decision variable—and thus the proportion of times that responses agree—across passes in a given condition (**Figure 7*d***). Stimulus-driven variability in the human decision variable is correlated across passes and increases decision variable correlation and response agreement. Internal noise is uncorrelated across passes and decreases decision variable correlation and response agreement. Decision variable correlation, which is estimated from the proportion of response agreements and disagreements, thus indicates the relative influence of internal and external sources of variability on the human decision variable (Chin & Burge 2020, Sebastian & Geisler 2018).

Under the hypothesis that the deterministic computations performed by the human observer are ideal (i.e., receptive field correlation ρ_f equals 1.0, and subsequent computations for processing and decoding are all optimal), the stimulus-driven variability in the human decision variable will equal the stimulus-driven variability in the ideal decision variable. If this is the case, then efficiency should predict human decision variable correlation:

$$\rho = \frac{\sigma^2_{ideal}}{\sigma^2_{ideal} + \sigma^2_{noise}} = \eta. \qquad 11.$$

If the human observer uses suboptimal receptive fields, then the relationship between decision variable correlation and efficiency $\rho = \eta/\rho_f^2$ is scaled by the inverse square of receptive field correlation (see Equation 9).

Efficiency nicely predicts the pattern of response agreement in the first human observer without additional free parameters (**Figure 7*e***). Similar results are obtained for all human observers. Hence, estimated decision variable correlation almost exactly equals efficiency for all human observers (**Figure 7*f***). Thus, an image-computable ideal observer grounded in the statistics of natural scenes predicts both the pattern of human thresholds and the repeatability of human responses. The results suggest that the only substantial suboptimality in the visual system is the presence of internal noise or some other factor that is uncorrelated with the stimulus (Equation 11). The results also indicate that the deterministic computations performed by the human visual system are nearly optimal. These results place strong constraints on the neural machinery that carries out the computations supporting performance in this task. Furthermore, these results demonstrate that ideal observer analysis, in conjunction with well-designed experiments, can be used not just to predict overall human performance with natural stimuli, but also to determine the sources of human performance limits.

Despite these encouraging results, there is still much work to be done. The ideal observer for speed estimation is limited in that the labeled training set for which the ideal observer was determined consisted of rigidly drifting movies that moved leftward and rightward at one isolated spatial location. Retinal images of real-world scenes move in all directions, and local motions include not just drifting, but also looming, transparency, and motion discontinuities caused by occlusions (Gekas et al. 2017, Nitzany & Victor 2014, Schrater et al. 2001). Constructing stimulus sets and developing ideal observers that include these complexities will be a challenge. However, the results described in this section constitute a solid foundation upon which to build.

Ideal Observer for Focus Error Estimation

Defocus blur (i.e., blur due to focus error) is nearly always present in the retinal image because the human lens can focus only at one distance at a time, and three-dimensional scenes have objects at many distances. Defocus blur impacts many biological and sensory-perceptual processes. It can stimulate eye growth (Wallman & Winawer 2004, Wildsoet & Wong 1999), drive accommodation (Campbell et al. 1958; Cholewiak et al. 2018; Fincham 1951; Flitcroft 1990; Kotulak & Schor 1986; Kruger et al. 1993, 1997), support predatory behavior across the animal kingdom (Harkness 1977, Nagata et al. 2012, Schaeffel et al. 1999), and influence the perception of depth (Burge et al. 2019, Held et al. 2010, Watt et al. 2005, Zannoli et al. 2016). The human ability to detect and discriminate optical blur has been measured in a variety of different contexts (Artal et al. 2004, Held et al. 2012, Sebastian et al. 2015, Wang & Ciuffreda 2005). However, until recently, it was not known whether focus error could be estimated from individual natural images. It may seem that a solution should not be possible. First, under certain conditions, the point spread function is identical for focus errors of the same magnitude but opposite signs; thus, focus error estimation can suffer from a sign ambiguity. Second, it is often unclear whether poor image quality is due to an error in focus or to some property of the scene (e.g., fog). However, regularities in the statistical properties of natural images and the optical properties of the human eye make a solution possible.

An ideal observer analysis demonstrated that the sign and magnitude of focus error can be accurately estimated from individual natural images formed in the human eye (Burge & Geisler 2011). The same can be done in smartphone (Burge 2017) and digital single-lens reflex (DSLR) cameras (Burge & Geisler 2012). The ideal observer capitalizes on two optical effects. First, the human eye focuses different wavelengths from the same target differently, a property called chromatic aberration (Thibos et al. 1992). If the eye is focused too close for a target, then the short-wavelength image is blurrier, and the long-wavelength image is sharper; if the eye is focused too far, then the long-wavelength image is blurrier (**Figure 8*a***). The difference in blur between the images in each color channel creates a cue to the sign of focus error (Burge & Geisler 2011, Fincham 1951, Flitcroft 1990). [Note that astigmatism similarly causes differential blur in different orientation channels and also creates a cue to the sign of focus error (Burge & Geisler 2011).] Second, when the eye is focused at the wrong distance, defocus blur more aggressively decreases the contrast of high compared to low frequency image features (i.e., fine versus coarse image detail), an effect that increases with the magnitude of focus error. Focus error (i.e., blur) thus causes a deterministic change in the shape of the amplitude spectrum (**Figure 8*b***). These deterministic shape changes would be useless if natural images had unpredictable amplitude spectra. Fortunately, as discussed above, individual natural images tend to have spectra that are characterized by a 1/f falloff. Although natural stimulus variability causes the shape of the amplitude spectrum to vary randomly from image patch to image patch, these random shape changes tend to be small relative to the deterministic shape changes caused by focus error. Hence, because the measurable signal changes more due to the latent variable (i.e., focus error) than to nuisance stimulus variability, the ideal observer can estimate the sign and magnitude of focus error with accuracy and precision (Burge 2017; Burge & Geisler 2011, 2012).

The ideal observer for focus error estimation uses receptive fields that extract information about both the overall shape of the image's amplitude spectrum and the difference in spectral shape between two (or all three) color channels (**Figure 8*c***). The optimal receptive fields select for spatial frequencies similar to those known to drive human accommodation (MacKenzie et al. 2010, Mathews & Kruger 1994, Owens 1980, Walsh & Charman 1988). The optimal receptive fields also have properties that are strikingly similar to the properties of double-opponent chromatic cells in the early visual cortex. These cells have been studied almost exclusively in the context of color

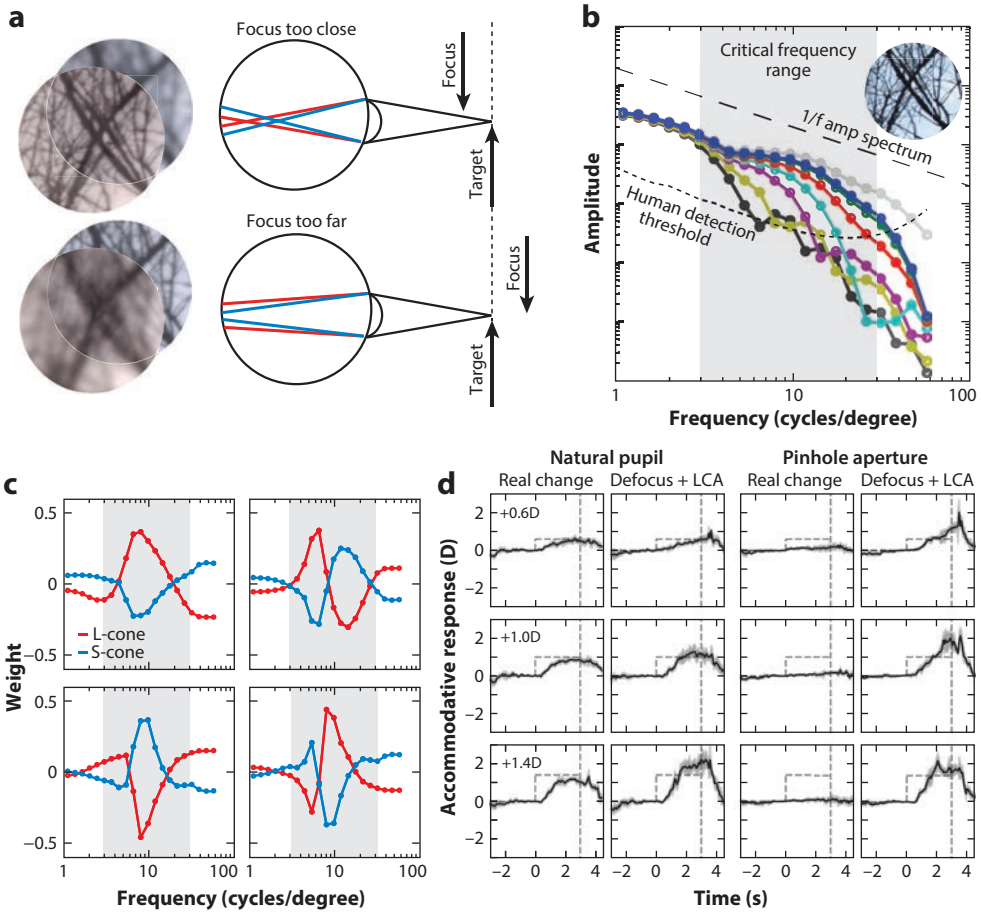

Figure 8

Estimating focus error from defocus blur. (*a*) Chromatic aberration in the human eye. Short (*blue*) wavelengths are focused more strongly than long (*red*) wavelengths by the human eye. When the eye is focused too close for the target (*top*), short-wavelength light will create a blurrier image. When the eye is focused too far (*bottom*), long-wavelength light will create a blurrier image. Chromatic aberration thus creates a cue to the sign of focus error. (*b*) Amplitude spectra of the image in panel *a* for focus errors ranging from 0.0 to 2.0 diopters (*colors*). The perfectly focused image has a 1/f amplitude spectrum (*gray*). Different focus errors are associated with differently shaped spectra (*colors*). The shaded region indicates the range of spatial frequencies that is most useful for estimating focus error; this range shifts toward lower spatial frequencies if the range of focus error is increased. (*c*) Optimal receptive fields for estimating focus error. The receptive fields select for spatial frequencies similar to those known to drive human accommodation (*shaded regions*) and have properties similar to double-opponent chromatic receptive fields in the cortex. Panels *b* and *c* adapted with permission from Burge & Geisler (2011). (*d*) With a natural pupil, graphically rendered defocus with chromatic aberration drives accommodation like a real change in focus error. With a pinhole aperture, real changes in focus error cause negligible changes in image blur, thereby removing the possibility that feedback could be provided by accommodative fluctuations. Nevertheless, with rendered defocus and chromatic aberration, the accommodative response scales approximately with the magnitude of the simulated focus error (*dashed lines*), similar to how it responds to real changes in focus error with a natural pupil. This result suggests that the accommodative system estimates the sign (i.e., direction) and magnitude of focus error from individual images. Panel adapted with permission from Cholewiak et al. (2018).

vision (Conway 2001, Johnson et al. 2008, Shapley & Hawken 2011, Shapley et al. 2019). The results from the ideal observer suggest that double-opponent chromatic cells are well suited to support focus error estimation. The hypothesis that double-opponent chromatic cells may mediate the processing of focus error was first made 30 years ago (see Burge & Geisler 2011, Flitcroft 1990). Unfortunately, this hypothesis has received little attention. Hopefully, it will be investigated in the coming years.

The ideal observer results also prompt one to ask whether the human visual system can estimate the sign and magnitude of focus error from individual images. This question has received considerable attention over the years (Fincham 1951, Flitcroft 1990). Several experiments have obtained evidence indicating that the accommodative system makes use of chromatic aberration (Kruger et al. 1993, 1997; Smithline 1974). However, a definitive test must rule out the possibility that online feedback could have influenced the results. Ruling this possibility out is difficult because the power of the human lens fluctuates continuously (Charman & Heron 1988); these microfluctuations could provide online feedback, even for briefly presented targets (Charman & Tucker 1978, Kotulak & Schor 1986, Walsh & Charman 1988).

A recent report provides the best evidence to date that the accommodative system estimates the sign and magnitude of focus errors (Cholewiak et al. 2018). Step changes in the accommodative distance of a target were introduced with real and rendered focus errors (i.e., defocus blur and chromatic aberration), and the accommodative response was measured. With a natural pupil, accommodative responses were always in the correct direction and increased systematically with the accommodative demand for both real and rendered blur (**Figure 8*d***). With a pinhole pupil and real focus error, the impact of chromatic aberration and defocus on image quality are eliminated, the image does not change with accommodation or focus error, and the accommodative responses are eliminated. However, with a pinhole pupil and rendered focus error (i.e., defocus blur and chromatic aberration), the accommodative response increases systematically with the accommodative demand, just as it does in both conditions with a natural pupil (**Figure 8*d***). This empirical result implies that the accommodative system can indeed estimate the sign and magnitude of focus error with reasonable accuracy from an individual image.

FUTURE DIRECTIONS

The psychophysics and neurophysiology of motion-in-depth perception are newly invigorated areas of study (Cormack et al. 2017). Image-computable ideal observers for motion-in-depth estimation with natural stimuli should contribute to progress in this exciting area of research. The task of estimating self-motion and object motion with components toward or away from the head almost certainly plays an outsized role in the reproductive fitness of predator and prey animals. Successes and failures are likely to influence whether predators will eat or starve, and whether prey will be eaten or survive. Psychophysical work has shown that signals relevant for motion-in-depth are processed via two distinct computations: the rate of change of binocular disparity over time and interocular differences in motion speed (i.e., interocular differences in the rate of position changes over time) (Czuba et al. 2010, Harris & Watamaniuk 1995, Rokers et al. 2008). The underlying neural mechanisms are just beginning to be probed with modern techniques (Bonnen et al. 2020, Czuba et al. 2014, Rokers et al. 2009, Sanada & DeAngelis 2014). Recently, it has also been shown that blur differences between the eyes can cause dramatic misperceptions of the distance and three-dimensional direction of moving objects (Burge et al. 2019). This surprising result indicates that binocular disparity, image motion, and defocus blur—the three cues that this review is focused on—all contribute to the accuracy and inaccuracy of motion-in-depth estimation. Despite this progress, it is unknown what the optimal receptive fields are for motion-in-depth

estimation in natural images, how the responses of those receptive fields should be processed, what the fundamental limits of performance are, and how closely human performance approaches these limits. An ideal observer analysis could provide answers to these questions. Similar questions could be tackled in other domains of study.

Another interesting direction for future work is the development of ideal observers that make optimal use of spatial context. All of the ideal observers described in this review were constrained to operate only on very local regions of a stimulus (see **Figure 1*b–e***). However, natural scenes and images tend to be correlated across space; the properties of nearby image and scene locations tend to be similar (Field 1987). For many estimation tasks, the accuracy of the estimate at one location can be enhanced by combining (i.e., pooling) nearby estimates consistent with the natural scene statistics (Kim & Burge 2018, 2020). Image-computable ideal observers that take maximum advantage of contextual information in natural scenes may represent the next big step forward in ideal observer analyses of natural visual tasks.

CONCLUSIONS

The ultimate purpose of vision science is to understand how vision works in the real world with natural stimuli. Image-computable ideal observers are powerful tools for obtaining that understanding. Newly collected stimulus databases and recent advances in receptive field learning (i.e., task-specific dimensionality reduction) have enabled the development of image-computable ideal observers for a broad new range of sensory-perceptual tasks with natural stimuli. These ideal observers have shown that natural stimulus variability (*a*) dictates the optimal computations for each task, (*b*) predicts the response properties of neurons in cortex, and (*c*) shapes the pattern of human performance. The methods and results discussed in this review should motivate future efforts to develop image-computable ideal observers for progressively more sophisticated tasks. Such efforts promise to further enrich our understanding of the links between natural image and scene statistics and the properties of sensory-perceptual systems.

DISCLOSURE STATEMENT

The author holds a United States patent on focus error estimation in individual images: US Patent Application No. 13/965,758. Reference No.: 5934 US. File No.: 93331–001910US-882167. Filing date: August 13, 2013.

ACKNOWLEDGMENTS

The author thanks David Brainard, Martin Banks, and Wilson Geisler for helpful discussion. This work was supported by National Institutes of Health grant R01-EY028571 from the National Eye Institute and the Office of Social and Behavioral Science Research and startup funds from the University of Pennsylvania.

LITERATURE CITED

Abbey CK, Eckstein MP. 2007. Classification images for simple detection and discrimination tasks in correlated noise. *J. Opt. Soc. Am. A* 24(12):B110–24

Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2(2):284–99

Albrecht DG, Geisler WS. 1991. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis. Neurosci.* 7(6):531–46

Anzai A, Ohzawa I, Freeman RD. 1999a. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J. Neurophysiol.* 82(2):874–90

Anzai A, Ohzawa I, Freeman RD. 1999b. Neural mechanisms for processing binocular information: I. Simple cells. *J. Neurophysiol.* 82(2):891–908

Artal P, Chen L, Fernández EJ, Singer B, Manzanera S, Williams DR. 2004. Neural compensation for the eye's optical aberrations. *J. Vis.* 4(4):281–87

Badcock DR, Schor CM. 1985. Depth-increment detection function for individual spatial channels. *J. Opt. Soc. Am. A* 2(7):1211–15

Banks MS, Geisler WS, Bennett PJ. 1987. The physical limits of grating visibility. *Vis. Res.* 27(11):1915–24

Banks MS, Gepshtein S, Landy MS. 2004. Why is spatial stereoresolution so low? *J. Neurosci.* 24(9):2077–89

Bex PJ, Makous W. 2002. Spatial frequency, phase, and the contrast of natural images. *J. Opt. Soc. Am. A* 19(6):1096–106

Bex PJ, Solomon SG, Dakin SC. 2009. Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *J. Vis.* 9(10):1

Bishop CM. 2006. *Pattern Recognition and Machine Learning*. Berlin: Springer

Blakemore C. 1970. The range and scope of binocular depth discrimination in man. *J. Physiol.* 211(3):599–622

Bonnen K, Czuba TB, Whritner JA, Kohn A, Huk AC, Cormack LK. 2020. Binocular viewing geometry shapes the neural representation of the dynamic three-dimensional environment. *Nat. Neurosci.* 23:113–21

Bradley C, Abrams J, Geisler WS. 2014. Retina-V1 model of detectability across the visual field. *J. Vis.* 14(12):22

Britten KH, Shadlen MN, Newsome WT, Movshon JA. 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12(12):4745–65

Burg MF, Cadena SA, Denfield GH, Walker EY, Tolias AS, et al. 2019. Learning divisive normalization in primary visual cortex. bioRxiv 767285. **https://doi.org/10.1101/767285**

Burge J. 2017. Accurate image-based estimates of focus error in the human eye and in a smartphone camera. *J. Soc. Inf. Disp.* 1:18–23

Burge J, Geisler WS. 2011. Optimal defocus estimation in individual natural images. *PNAS* 108(40):16849–54

Burge J, Geisler WS. 2012. Optimal defocus estimates from individual images for autofocusing a digital camera. In *Proc. SPIE 8299*, *Digital Photography VIII*, art. 82990E. Bellingham, WA: SPIE

Burge J, Geisler WS. 2014. Optimal disparity estimation in natural stereo images. *J. Vis.* 14(2):1

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(2):e1005281

Burge J, McCann BC, Geisler WS. 2016. Estimating 3D tilt from local image cues in natural scenes. *J. Vis.* 16(13):2

Burge J, Rodriguez-Lopez V, Dorronsoro C. 2019. Monovision and the misperception of motion. *Curr. Biol.* 29(15):2586–92.e4

Burgess AE, Colborne B. 1988. Visual signal detection. IV. Observer inconsistency. *J. Opt. Soc. Am. A* 5(4):617–27

Burgess AE, Wagner RF, Jennings RJ, Barlow HB. 1981. Efficiency of human visual signal discrimination. *Science* 214(4516):93–94

Busse L, Wade AR, Carandini M. 2009. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* 64(6):931–42

Campbell FW, Kulikowski JJ. 1966. Orientational selectivity of the human visual system. *J. Physiol.* 187(2):437–45

Campbell FW, Westheimer G, Robson JG. 1958. Significance of fluctuations of accommodation. *J. Opt. Soc. Am.* 48(9):669

Carandini M, Heeger DJ. 2012. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13(1):51–62

Carandini M, Heeger DJ, Movshon JA. 1997. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17(21):8621–44

Cavanaugh JR, Bair W, Movshon JA. 2002. Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *J. Neurophysiol.* 88(5):2547–56

Charman WN, Heron G. 1988. Fluctuations in accommodation: a review. *Ophthalmic Physiol. Opt.* 8(2):153–64

Charman WN, Tucker J. 1978. Accommodation and color. *J. Opt. Soc. Am.* 68(4):459–71

Chauhan T, Masquelier T, Montlibert A, Cottereau BR. 2018. Emergence of binocular disparity selectivity through Hebbian learning. *J. Neurosci.* 38(44):9563–78

Chin BM, Burge J. 2020. Predicting the partition of behavioral variability in speed perception with naturalistic stimuli. *J. Neurosci.* 40(4):864–79

Cholewiak SA, Love GD, Banks MS. 2018. Creating correct blur and its effect on accommodation. *J. Vis.* 18(9):1

Coen-Cagli R, Kohn A, Schwartz O. 2015. Flexible gating of contextual influences in natural vision. *Nat. Neurosci.* 18(11):1648–55

Conway BR. 2001. Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *J. Neurosci.* 21(8):2768–83

Cormack LK, Czuba TB, Knöll J, Huk AC. 2017. Binocular mechanisms of 3D motion processing. *Annu. Rev. Vis. Sci.* 3:297–318

Cormack LK, Stevenson SB, Schor CM. 1991. Interocular correlation, luminance contrast and cyclopean processing. *Vis. Res.* 31(12):2195–207

Cumming BG, DeAngelis GC. 2001. The physiology of stereopsis. *Annu. Rev. Neurosci.* 24:203–38

Czuba TB, Huk AC, Cormack LK, Kohn A. 2014. Area MT encodes three-dimensional motion. *J. Neurosci.* 34(47):15522–33

Czuba TB, Rokers B, Huk AC, Cormack LK. 2010. Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. *J. Neurophysiol.* 104(5):2886–99

De Valois RL, Albrecht DG, Thorell LG. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vis. Res.* 22(5):545–59

De Vries HL. 1943. The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica* 10(7):553–64

DeAngelis GC, Ohzawa I, Freeman RD. 1991. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* 352(6331):156–59

Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870):429–33

Feord RC, Sumner ME, Pusdekar S, Kalra L, Gonzalez-Bellido PT, Wardill TJ. 2020. Cuttlefish use stereopsis to strike at prey. *Sci. Adv.* 6(2):eaay6036

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4(12):2379–94

Fincham E. 1951. The accommodation reflex and its stimulus. *Br. J. Ophthalmol.* 35(7):381–93

Flitcroft DI. 1990. A neural and computational model for the chromatic control of accommodation. *Vis. Neurosci.* 5(6):547–55

Fox R, Lehmkuhle SW, Bush RC. 1977. Stereopsis in the falcon. *Science* 197(4298):79–81

Geisler WS. 1984. Physical limits of acuity and hyperacuity. *J. Opt. Soc. Am. A* 1(7):775–82

Geisler WS. 1989. Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96(2):267–314

Geisler WS. 2003. Ideal observer analysis. In *The Visual Neuerosciences*, Vol. 10, ed. L Chalupa, J Werner, pp. 825–37. Cambridge, MA: MIT Press

Geisler WS. 2008. Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59:167–92

Geisler WS. 2011. Contributions of ideal observer theory to vision research. *Vis. Res.* 51(7):771–81

Geisler WS, Davila KD. 1985. Ideal discriminators in spatial vision: two-point stimuli. *J. Opt. Soc. Am. A* 2(9):1483–97

Geisler WS, Najemnik J, Ing AD. 2009. Optimal stimulus encoders for natural tasks. *J. Vis.* 9(13):17

Gekas N, Meso AI, Masson GS, Mamassian P. 2017. A normalization mechanism for estimating visual motion across speeds and scales. *Curr. Biol.* 27(10):1514–20.e3

Girshick AR, Landy MS, Simoncelli EP. 2011. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14(7):926–32

Goncalves NR, Welchman AE. 2017. "What not" detectors help the brain see in depth. *Curr. Biol.* 27(10):1403–8

Green DM, Swets JA. 1966. *Signal Detection Theory and Psychophysics*. Hoboken, NJ: Wiley

Hansen BC, Essock EA. 2004. A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *J. Vis.* 4(12):1044–60

Harkness L. 1977. Chameleons use accommodation cues to judge distance. *Nature* 267:346–49

Harris JM, Watamaniuk SN. 1995. Speed discrimination of motion-in-depth using binocular cues. *Vis. Res.* 35(7):885–96

Heeger DJ. 1987. Model for the extraction of image flow. *J. Opt. Soc. Am. A* 4(8):1455–71

Heeger DJ. 1992. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9(2):181–97

Held RT, Cooper EA, Banks MS. 2012. Blur and disparity are complementary cues to depth. *Curr. Biol.* 22(5):426–31

Held RT, Cooper EA, O'Brien JF, Banks MS. 2010. Using blur to affect perceived distance and size. *ACM Trans. Graph.* 29(2):19

Hibbard PB. 2008. Binocular energy responses to natural images. *Vis. Res.* 48(12):1427–39

Iyer AV, Burge J. 2018a. Depth variation and stereo processing tasks in natural scenes. *J. Vis.* 18(6):4

Iyer AV, Burge J. 2018b. Optimal binocular disparity estimation in the presence of natural depth variation. *J. Vis.* 18(10):627

Iyer A, Burge J. 2019. The statistics of how natural images drive the responses of neurons. *J. Vis.* 19(13):4

Jaini P, Burge J. 2017. Linking normative models of natural tasks to descriptive models of neural response. *J. Vis.* 17(12):16

Jazayeri M, Movshon JA. 2006. Optimal representation of sensory information by neural populations. *Nat. Neurosci.* 9(5):690–96

Jogan M, Stocker AA. 2015. Signal integration in human visual speed perception. *J. Neurosci.* 35(25):9381–90

Johnson EN, Hawken MJ, Shapley R. 2008. The orientation selectivity of color-responsive neurons in macaque V1. *J. Neurosci.* 28(32):8096–106

Julesz B. 1964. Binocular depth perception without familiarity cues. *Science* 145(3630):356–62

Kane D, Bex P, Dakin S. 2011. Quantifying "the aperture problem" for judgments of motion direction in natural scenes. *J. Vis.* 11(3):25

Kim S, Burge J. 2018. The lawful imprecision of human surface tilt estimation in natural scenes. *eLife* 7:31448

Kim S, Burge J. 2020. Natural scene statistics predict how humans pool information across space in surface tilt estimation. *PLOS Comput. Biol.* 16(6):e1007947

Knill DC, Richards W. 1996. *Perception as Bayesian Inference*. Cambridge, UK: Cambridge Univ. Press

Kotulak JC, Schor CM. 1986. A computational model of the error detector of human visual accommodation. *Biol. Cybern.* 54(3):189–94

Kruger PB, Mathews S, Aggarwala KR, Sanchez N. 1993. Chromatic aberration and ocular focus: Fincham revisited. *Vis. Res.* 33(10):1397–411

Kruger PB, Mathews S, Katz M, Aggarwala KR, Nowbotsing S. 1997. Accommodation without feedback suggests directional signals specify ocular focus. *Vis. Res.* 37(18):2511–26

Landy MS, Banks MS, Knill DC. 2011. Ideal-observer models of cue integration. In *Sensory Cue Integration*, ed. J Trommershäuser, K Kording, MS Landy, pp. 5–29. Oxford, UK: Oxford Univ. Press

Landy MS, Maloney LT, Johnston EB, Young M. 1995. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vis. Res.* 35(3):389–412

Legge GE, Foley JM. 1980. Contrast masking in human vision. *J. Opt. Soc. Am.* 70(12):1458–71

Legge GE, Kersten D, Burgess AE. 1987. Contrast discrimination in noise. *J. Opt. Soc. Am. A* 4(2):391–404

Liu J, Newsome WT. 2006. Local field potential in cortical area MT: stimulus tuning and behavioral correlations. *J. Neurosci.* 26(30):7779–90

Ma WJ, Beck JM, Latham PE, Pouget A. 2006. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9(11):1432–38

MacKenzie KJ, Hoffman DM, Watt SJ. 2010. Accommodation to multiple-focal-plane displays: implications for improving stereoscopic displays and for accommodation control. *J. Vis.* 10(8):22

Marr D. 1982. *Vision*. New York: W.H. Freeman & Company
Mathews S, Kruger PB. 1994. Spatiotemporal transfer function of human accommodation. *Vis. Res.* 34(15):1965–80
McKee SP, Levi DM, Bowne SF. 1990. The imprecision of stereopsis. *Vis. Res.* 30(11):1763–79
Mueller CG. 1951. Frequency of seeing functions for intensity discrimination of various levels of adapting intensity. *J. Gen. Physiol.* 34(4):463–74
Nagata T, Koyanagi M, Tsukamoto H, Saeki S, Isono K, et al. 2012. Depth perception from image defocus in a jumping spider. *Science* 335(6067):469–71
Najemnik J, Geisler WS. 2005. Optimal eye movement strategies in visual search. *Nature* 434(7031):387–91
Nishimoto S, Gallant JL. 2011. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J. Neurosci.* 31(41):14551–64
Nityananda V, Tarawneh G, Rosner R, Nicolas J, Crichton S, Read J. 2016. Insect stereopsis demonstrated using a 3D insect cinema. *Sci. Rep.* 6:18718
Nitzany EI, Victor JD. 2014. The statistics of local motion signals in naturalistic movies. *J. Vis.* 14(4):10
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(43):10049–60
Ogle KN. 1952. On the limits of stereoscopic vision. *J. Exp. Psychol.* 44(4):253–59
Ohzawa I, DeAngelis GC, Freeman RD. 1990. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249(4972):1037–41
Olshausen BA, Field DJ. 1997. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* 37(23):3311–25
Owens DA. 1980. A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings. *Vis. Res.* 20(2):159–67
Parker AJ. 2007. Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.* 8(5):379–91
Pelli DG. 1990. The quantum efficiency of vision. In *Vision: Coding and Efficiency*, ed. C Blackmore, pp. 3–24. Cambridge, UK: Cambridge Univ. Press
Portilla J, Simoncelli EP. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* 40(1):49–71
Prince SJD, Cumming BG, Parker AJ. 2002. Range and mechanism of encoding of horizontal disparity in macaque V1. *J. Neurophysiol.* 87(1):209–21
Qian N. 1997. Binocular disparity and the perception of depth. *Neuron* 18(3):359–68
Read JCA, Cumming BG. 2007. Sensors for impossible stimuli may solve the stereo correspondence problem. *Nat. Neurosci.* 10(10):1322–28
Ringach DL. 2002. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.* 88(1):455–63
Rokers B, Cormack LK, Huk AC. 2008. Strong percepts of motion through depth without strong percepts of position in depth. *J. Vis.* 8(4):6
Rokers B, Cormack LK, Huk AC. 2009. Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nat. Neurosci.* 12(8):1050–55
Rose A. 1948. The sensitivity performance of the human eye on an absolute scale. *J. Opt. Soc. Am.* 38(2):196–208
Rossel S. 1983. Binocular stereopsis in an insect. *Nature* 302(5911):821–22
Ruff DA, Alberts JJ, Cohen MR. 2016. Relating normalization to neuronal populations across cortical areas. *J. Neurophysiol.* 116(3):1375–86
Rust NC, Schwartz O, Movshon JA, Simoncelli EP. 2005. Spatiotemporal elements of macaque v1 receptive fields. *Neuron* 46(6):945–56
Sanada TM, DeAngelis GC. 2014. Neural representation of motion-in-depth in area MT. *J. Neurosci.* 34(47):15508–21
Schaeffel F, Murphy CJ, Howland HC. 1999. Accommodation in the cuttlefish (*Sepia officinalis*). *J. Exp. Biol.* 202:3127–34

Scholl B, Burge J, Priebe NJ. 2013. Binocular integration and disparity selectivity in mouse primary visual cortex. *J. Neurophysiol.* 109(12):3013–24

Schrater PR, Knill DC, Simoncelli EP. 2001. Perceiving visual expansion without optic flow. *Nature* 410(6830):816–19

Schumer RA, Julesz B. 1984. Binocular disparity modulation sensitivity to disparities offset from the plane of fixation. *Vis. Res.* 24(6):533–42

Schütt HH, Wichmann FA. 2017. An image-computable psychophysical spatial vision model. *J. Vis.* 17(12):12

Sebastian S, Abrams J, Geisler WS. 2017. Constrained sampling experiments reveal principles of detection in natural scenes. *PNAS* 114(28):E5731–40

Sebastian S, Burge J, Geisler WS. 2015. Defocus blur discrimination in natural images with natural optics. *J. Vis.* 15(5):16

Sebastian S, Geisler WS. 2018. Decision-variable correlation. *J. Vis.* 18(4):3

Shapley R, Hawken MJ. 2011. Color in the cortex: single- and double-opponent cells. *Vis. Res.* 51(7):701–17

Shapley R, Nunez V, Gordon J. 2019. Cortical double-opponent cells and human color perception. *Curr. Opin. Behav. Sci.* 30:1–7

Simoncelli EP, Heeger DJ. 1998. A model of neuronal responses in visual area MT. *Vis. Res.* 38(5):743–61

Simoncelli EP, Olshausen BA. 2001. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24:1193–216

Sinha SR, Bialek W, de Ruyter van Steveninck RR. 2018. Optimal local estimates of visual motion in a natural environment. arXiv:1812.11878 [q-bio.NC]

Smithline LM. 1974. Accommodative response to blur. *J. Opt. Soc. Am.* 64(11):1512–16

Stevenson SB, Cormack LK, Schor CM, Tyler CW. 1992. Disparity tuning in mechanisms of human stereopsis. *Vis. Res.* 32(9):1685–94

Stocker AA, Simoncelli EP. 2006. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9(4):578–85

Stromeyer CF, Julesz B. 1972. Spatial-frequency masking in vision: critical bands and spread of masking. *J. Opt. Soc. Am.* 62(10):1221–32

Tanabe S, Haefner RM, Cumming BG. 2011. Suppressive mechanisms in monkey V1 help to solve the stereo correspondence problem. *J. Neurosci.* 31(22):8295–305

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(19):3594–600

Tyler CW, Julesz B. 1978. Binocular cross-correlation in time and space. *Vis. Res.* 18(1):101–5

van der Willigen RF. 2011. Owls see in stereo much like humans do. *J. Vis.* 11(7):10

Vos JJ, Walraven PL. 1972. An analytical description of the line element in the zone-fluctuation model of color vision: I. Basic concepts. *Vis. Res.* 12(8):1327–44

Wallman J, Winawer J. 2004. Homeostasis of eye growth and the question of myopia. *Neuron* 43(4):447–68

Walsh G, Charman WN. 1988. Visual sensitivity to temporal change in focus and its relevance to the accommodation response. *Vis. Res.* 28(11):1207–21

Wang B, Ciuffreda KJ. 2005. Foveal blur discrimination of the human eye. *Ophthalmic Physiol. Opt.* 25(1):45–51

Watson AB, Solomon JA. 1997. Model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A* 14(9):2379–91

Watt SJ, Akeley K, Ernst MO, Banks MS. 2005. Focus cues affect perceived depth. *J. Vis.* 5(10):834–62

Weiss Y, Simoncelli EP, Adelson EH. 2002. Motion illusions as optimal percepts. *Nat. Neurosci.* 5(6):598–604

Welchman AE. 2016. The human brain in depth: how we see in 3D. *Annu. Rev. Vis. Sci.* 2:345–76

Westheimer G. 1979. The spatial sense of the eye. Proctor lecture. *Investig. Ophthalmol. Vis. Sci.* 18(9):893–912

Westheimer G. 1982. The spatial grain of the perifoveal visual field. *Vis. Res.* 22(1):157–62

Wheatstone C. 1838. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos. Trans. R. Soc. Lond.* 128:371–94

White D, Burge J. 2018. Human binocular disparity estimation with natural stereo-images. *J. Vis.* 18(10):993

Wildsoet CF, Wong R. 1999. A far-sighted view of myopia. *Nat. Med.* 5(8):879–80

Zannoli M, Love GD, Narain R, Banks MS. 2016. Blur and the perception of depth at occlusions. *J. Vis.* 16(6):17