



Published in final edited form as:

Vision Res. 2011 April 13; 51(7): 771–781. doi:10.1016/j.visres.2010.09.027.

Contributions of Ideal Observer Theory to Vision Research

Wilson S. Geisler

Center for Perceptual Systems and Department of Psychology, University of Texas at Austin

Abstract

An ideal observer is a hypothetical device that performs optimally in a perceptual task given the available information. The theory of ideal observers has proven to be a powerful and useful tool in vision research, which has been applied to a wide range of problems. Here I first summarize the basic concepts and logic of ideal observer analysis and then briefly describe applications in a number of different areas, including pattern detection, discrimination and estimation, perceptual grouping, shape, depth and motion perception and visual attention, with an emphasis on recent applications. Given recent advances in mathematical statistics, in computational power, and in techniques for measuring behavioral performance, neural activity and natural scene statistics, it seems certain that ideal observer theory will play an ever increasing role in basic and applied areas of vision science.

Keywords

Ideal observer theory; natural scene statistics; natural tasks; psychophysics

Introduction

The major goal of basic vision research is to understand and predict visual performance. Empirical progress toward this goal has come from measurements of natural stimuli, physiological optics, anatomy and neurophysiology of visual pathways, and behavioral performance in adult and developing organisms. Empirical findings in vision research have been interpreted and driven by a wide array of qualitative and quantitative theories and models. Of the quantitative theories, the theory of ideal observers has played a unique and fundamental role, especially during the last 25 years.

There are many different visual tasks human and non-human primates perform under natural conditions and can perform under laboratory conditions. What is ultimately desired is a general theory that parsimoniously explains and quantitatively predicts visual performance in arbitrary natural and laboratory visual tasks. The field is a very long way from such a theory. Instead, vision researchers have been forced to identify specific well-defined tasks, or families of tasks, and then attempt to develop informal or formal models that can explain and predict performance in those specific tasks. For each task or family of tasks the field typically attempts to address a number of fundamental questions, which include: What are the properties of the stimuli in a given task that contribute to the measured performance? How and where are those properties encoded into neural activity along the visual pathway? How are the different sources of task-relevant sensory information combined by the visual

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

system? What are the relative contributions of peripheral and central mechanisms in the task? What are the contributions of ‘bottom-up’ and ‘top-down’ mechanisms in the task? How is the task-relevant information in the neural activity along the visual pathways decoded into behavior?

Twenty-five years ago ideal-observer theory had only been worked out and applied to a very narrow range of simple tasks. In the intervening years it has been applied to much wider range of tasks. This article attempts to summarize some of the different kinds of tasks where ideal observer theory has played a major role in developing models of visual performance and in answering one or more of the questions listed above. Due to space limitations, the primary focus is on behavioral performance, even though ideal observer theory has also played an important role in studies of the underlying neurophysiology. Before getting down to specific tasks, there are some general points to make about the theory of ideal observers.

Ideal observers

An ideal observer is a hypothetical device that performs a given task at the optimal level possible, given the available information and any specified constraints. If the ideal observer can be derived for a given task, then it can serve vision research in several important ways:

1. *Identifying task-relevant stimulus properties.* The ideal observer performs its task optimally; thus, in deriving the ideal observer one is forced to identify, at least implicitly, all the task-relevant properties of the stimuli. This makes it possible to rigorously evaluate and test which relevant stimulus properties real observers exploit when they perform the task.
2. *Describing how to use those properties to perform the task.* The ideal observer explicitly specifies one set of computations that is sufficient to achieve optimal performance in the task. Although there may be other sets of computations that are sufficient to achieve optimal or near-optimal performance, an ideal observer often provides deep insight into the computational requirements of the task.
3. *Providing a benchmark against which to compare the performance of real or model vision systems.* The performance of the ideal observer is a precise ‘information measure’ that describes how the task-relevant information varies across stimulus conditions. In general, real and model (heuristic) observers do not efficiently use all the task-relevant information and hence do not reach the performance levels of the ideal observer. However, if a real or model observer is exploiting the same stimulus properties as the ideal observer, then its performance should parallel that of the ideal observer (e.g., stimulus conditions that are harder for the ideal observer should be harder for the real or model observer). When human performance approaches ideal performance, then the implications for neural processing can become particularly powerful; specifically, all hypotheses (model observers) that cannot approach ideal performance can be rejected. When human performance is far below ideal, there are generally a greater number of models than could explain human performance.
4. *Suggesting principled hypotheses and models for real performance.* Natural selection and learning during the lifespan necessarily drive perceptual systems in the direction of optimal performance in the tasks the organism normally performs in its natural environment. Although perceptual systems may not reach optimum, it is a good bet that they are closer to ideal than to the simple models one might generate from intuition or to explain some experimental result. Thus, a powerful research strategy is to use the ideal observer to guide the generation of hypotheses and models of real performance. This is often done by degrading the ideal observer

with hypothesized neural noise or with hypothesized heuristic computations that approximate ideal computations. Models generated this way are principled and often have very few free parameters.

In all visual tasks, performance is limited at least in part by various sources of random variability. These include variability in the stimuli (e.g., photon noise, heterogeneity of the objects defining a category, variability in scene illumination, variability due to the projection from a 3D environment to the 2D retinal images), variability in the sensory neural representation (e.g., sensory neural noise), and variability in the decoding circuits (e.g., decision and motor neural noise). Thus, ideal observers are properly defined in probabilistic terms, using statistical decision theory and information theory. Most of the ideal observers described here fall within the framework of Bayesian statistical decision theory.

The logic and structure of a Bayesian ideal observer is relatively straight forward. In most visual tasks, there is some actual unknown state of the world ω (e.g., a particular class of physical object) that gives rise to a particular (random) received stimulus \mathbf{S} reaching the eyes. The observer's goal is to make the response \mathbf{r}_{opt} that maximizes the utility (or equivalently minimizes loss) averaged over all possible states of the world (in that task), given the stimulus \mathbf{S} . If some biological constraints are included, then the goal becomes maximizing utility given a neural representation of the stimulus $\mathbf{Z} = \mathbf{g}(\mathbf{S};\theta)$, where $\mathbf{g}(\mathbf{S};\theta)$ is the constraint function that specifies the mapping of the stimulus into a neural representation. For example, \mathbf{Z} might represent the number of photons absorbed in each photoreceptor, and $\mathbf{g}(\mathbf{S};\theta)$ the mapping from the stimulus at the eyes to photons absorbed in each photoreceptor. (The symbol θ is included because in some applications of ideal observer theory it is useful to allow unknown parameters in the mapping from stimulus to neural representation; see later.) Formally, the ideal observer's response is given by

$$\mathbf{r}_{opt}(\mathbf{Z}) = \arg \max_{\mathbf{r}} \left(\sum_{\omega} \gamma(\mathbf{r}, \omega) p(\omega|\mathbf{Z}) \right) \quad (1)$$

where $p(\omega|\mathbf{Z})$ is the posterior probability of each state of the world given the received signal \mathbf{Z} , and $\gamma(\mathbf{r}, \omega)$ is the utility (gain or loss) of making response \mathbf{r} when the true state of the world is ω . If there is no constraint function, then \mathbf{Z} in equation (1) is replaced by \mathbf{S} . The performance of the ideal observer (e.g., accuracy and/or reaction time) can sometimes be determined by direct calculation, but often can be determined only by Monte Carlo simulation (i.e., applying equation (1) to random samples of the signal \mathbf{Z}).

Equation (1) is fairly general; in fact, all of the examples of ideal observers described here are special cases. However, as a concrete example, consider a task where there are just two categories of object and the observer's task is to be as accurate as possible in identifying which object was presented. In this case, the state of the world can take on only two values ($\omega = 1$ and $\omega = 2$) and observer's responses can take on only two values ($\mathbf{r} = 1$ and $\mathbf{r} = 2$). Because the goal is to be as accurate as possible, the proper utility function rewards correct responses ($\gamma(\mathbf{r}, \omega) = 1$ if $\mathbf{r} = \omega$) and does not reward (or punishes) incorrect responses ($\gamma(\mathbf{r}, \omega) = 0$ if $\mathbf{r} \neq \omega$). Substituting into equation (1) shows that the ideal decision rule is simply to make response $\mathbf{r} = 1$ if $p(\omega = 1|\mathbf{Z}) > p(\omega = 2|\mathbf{Z})$ and otherwise make response $\mathbf{r} = 2$. In other words, the rule is simply to pick the object with the highest posterior probability.

Although the ideal observer framework as described above is sufficient for present purposes, there are a number of useful elaborations of the framework that should be mentioned here. One conceptual elaboration is the *influence graph* (or *Bayesian network*), which describes

the qualitative mapping between states or properties of the world ω and properties of the stimulus \mathbf{S} (e.g., see Kersten, Mamassian & Yuille 2004; Jacobs & Kruschke 2010). Influence graphs specify the task relevant properties of the world (local environment) and stimulus, and their causal relationships, and they imply how those properties should be treated in computing posterior probabilities for the task. A second elaboration of the framework is to incorporate mechanisms (including ideal Bayesian mechanisms) for learning posterior probability distributions, utility functions, or simple decision rules equivalent to equation (1), either on short (Jacobs & Kruschke 2010) or evolutionary (Geisler & Diehl 2003) time scales. A third elaboration is to take into account biophysical costs (e.g., energy) of neural computations (Laughlin & Sejnowski 2003; Koch et al. 2004; Manning & Brainard 2009) and motor responses (Körding & Wolpert 2006), or more generally fitness (Geisler & Diehl 2003).

Pattern detection, discrimination and identification

The earliest applications of ideal observer theory in vision were concerned with understanding how detection is limited by photon noise and how the performance of real observers compares that of an ideal observer that is limited only by photon noise (e.g., Rose 1948; DeVries 1943; Barlow 1957; Cohn & Lashley 1974). For this ideal observer, the threshold for detecting an increment (or decrement) in intensity increases in proportion to the square root of the background (baseline) intensity. Early studies showed that there are some conditions in which human increment detection performance parallels that of the photon noise limited ideal observer, but, on an absolute scale, humans are substantially less efficient than the ideal observer.

Shortly after the 25th anniversary of Vision Research, photon-noise-limited ideal observers were derived and applied to a wider range of tasks, including various acuity tasks (Geisler 1984; 1989), contrast sensitivity and contrast discrimination tasks in adults (Banks, et al. 1987; Geisler 1989; Banks et al. 1991; Sekiguchi et al. 1993; Arnow & Geisler 1996) and in infants (Banks & Bennett 1988), color discrimination (Geisler, 1989), and letter identification (Beckman & Legge, 2002). These studies also evaluated the additional effects on ideal observer performance of biological constraints such as the optics of the eye, the spatial and chromatic sampling by the photoreceptors, photoreceptor noise, and ganglion cell spatial summation. This work provides insight into how optics, photoreceptors, photon noise, and retinal spatial summation contribute to human performance. The general finding is that human performance is suboptimal, but often parallels ideal observer performance qualitatively (and sometimes quantitatively) for a surprising number of detection and discrimination tasks. In other words, for these tasks the variation in human performance across conditions is often predicted by the information available in the retinal responses (see Geisler 2003 for a review). Nonetheless, the suboptimal performance of human observers implies substantial contributions of central factors.

Barlow (1978) reasoned that it may be possible in psychophysical experiments to largely bypass the effects of photon noise and retinal factors, and hence isolate the effects of some of the central factors, by adding high levels of external noise. This proved to be a powerful insight that spawned a number of studies measuring target detection and identification in Gaussian or Poisson pixel noise. Importantly, using statistically independent Gaussian or Poisson pixel noise makes it relatively easy to derive and determine ideal observer performance. For example in simple detection (where the goal is to maximize accuracy) the ideal observer applies a template matching the shape of the target and then compares the template response to a criterion. Adding external noise raises detection and identification thresholds; however, as expected from bypassing low-level factors, performance generally moves closer to that of the ideal observer (i.e., efficiency increases).

For an ideal observer limited by external noise, the square of contrast detection (or identification) threshold increases linearly with the square of the root-mean-squared (rms) contrast of the external noise. Human thresholds match this prediction approximately both in the fovea (Burgess et al. 1981; Legge et al. 1987; Pelli 1990) and in the near periphery (Najemnik & Geisler 2005), once the external noise contrast exceeds a certain level (see Figures 1a and 1b). Measuring contrast thresholds as a function of external noise contrast allows one to estimate an equivalent internal noise, which can be interpreted as the combined effect of those low-level factors that are swamped (dominated) by the external noise as external noise contrast increases (for review see Pelli & Farell 1999).

Although efficiency is higher with moderate to high levels of external Gaussian or Poisson noise, performance is still generally well below ideal. Several factors probably contribute to this suboptimal performance. One factor is internal uncertainty (Tanner 1961; Nachmias & Kocher 1970; Cohn & Lashley 1974; Pelli 1985), which may include uncertainty about the spatial location of the target (spatial uncertainty) or uncertainty about certain target feature properties such as orientation or shape (channel uncertainty). These are forms of internal noise that necessarily limit performance. Another factor is contrast nonlinearities (e.g., contrast gain control), which may produce masking effects above and beyond those due to the similarity of the target and external noise (Foley & Legge 1981; Foley 1994; Geisler & Albrecht 1997). A third factor is inefficient pooling of target feature information. If the features that the real observer uses to detect the target do not correspond to the template that matches the shape of the target, then performance will be suboptimal. The image features that an observer uses in performing a detection or identification task can be estimated using the *classification image* technique, which is based on ideal observer theory and measures the trial-by-trial correlation between the image noise pixels and the observer's behavioral responses (Ahumada 1996). Measurements of classification images for various kinds of target reveal non-optimal pooling of feature information (Ahumada 1996; 2002; Eckstein et al. 2002; Gold et al. 2000; Murray et al. 2005). Some of this non-optimal pooling is due to uncertainty and contrast nonlinearities. However, these factors can only blur (or sharpen) the classification image; whereas measured classification images frequently reveal missing target features and sometimes added illusory features (Figure 1c; Murray et al. 2005).

Although ideal observer theory has played a fundamental role in much that has been learned about the mechanisms of detection, discrimination and identification over the last 25 years, much remains to be done. One critical direction for future research is to get a better understanding of the differences between foveal and peripheral vision. Detection, discrimination and identification performance vary dramatically across the visual field and understanding the nature of this variation is essential for understanding and predicting performance in most natural tasks. At this point, relatively little is known about how uncertainty, contrast nonlinearities and feature pooling vary with retinal eccentricity. A second critical direction is to move beyond laboratory stimuli to more naturalistic stimuli. Most of what is known about detection, discrimination and identification performance is for simple targets in Gaussian or Poisson noise. Ideal observer theory is likely to play a central role in the push forward on both these fronts.

Pattern estimation

Conceptually, an estimation task can be viewed as a special case of an identification task where the number of categories becomes arbitrarily large. In practice, the two kinds of tasks are typically analyzed differently because often the appropriate utility function in an identification task will penalize all errors equally, whereas often the appropriate utility function in an estimation task will penalize large errors more than small errors.

A fundamental issue in visual neuroscience is how the brain estimates spatial and chromatic image properties from the discrete spatial and chromatic samples provided by the output of the retina. Recent studies suggest that Bayesian ideal observers can be helpful in addressing this issue. For example, Brainard et al. (2008) consider the problem of how the visual system estimates chromatic image properties from the array of different cone types in the human retina. Figure 2a shows the cone mosaics of three of five subjects measured with adaptive optics by Hofer et al. (2005). Not only are the mosaics irregular, but there are large individual differences. Using these mosaics and the measured point spread functions of each subject's eye, Brainard et al. (2008) determined what would be the Bayesian optimal estimate of the 'color' (one of nine color labels) for 0.6 min wide spots of 500, 550, and 600 nm light, presented randomly within the mosaic. They found that the optimal estimates are strongly dependent not only on which cone is stimulated most, but also by the specific cone types within the neighboring region. Further, they found that there was a strong correlation between the distributions of optimal estimates and estimates by the subjects. For example, the percentage of "white" responses is expected to depend strongly on the asymmetry in the proportion of L and M cones within the mosaics (open circles in Figure 2b) and indeed this is what was observed in the subjects' responses (solid circles in Figure 2b). Similarly, the ideal-observer predictions for all color names correlated strongly with the subjects' responses ($r = 0.83$), for all three test spot wavelengths (Figure 2c).

In a related recent example, Geisler & Perry (under review) consider the task of estimating retinal image luminance patterns from the coarse spatial sampling provided by the peripheral ganglion cell mosaic. It is well known that the spatial resolution (inverse kernel size) and sampling rate of the midget ganglion cells falls rapidly with retinal eccentricity. For example, at just 2.5 deg eccentricity the sampling rate in both directions has fallen to about half that in the fovea—a factor of four fewer samples per unit area (see Figure 3a). Geisler and Perry find that if the visual system were decoding the responses at 2.5 deg eccentricity using something simple like linear (or cubic) interpolation, then the representation of a ground truth image (Figure 3b) would be equivalent to the image in Figure 3c. On the other hand, if the visual system were decoding the responses using an ideal Bayesian estimator that takes into account the average local statistics of natural images, then the representation would be equivalent to the image in Figure 3d, which is considerably more accurate (40% smaller mean squared error).

This increase in accuracy over linear interpolation is typical of the 1000 natural images tested (42% decrease in mean squared error on average). Given these potential gains in accuracy, it is likely that the visual system would exploit the average natural image statistics and apply more sophisticated estimation heuristics than simple linear interpolation. The Bayesian ideal observer provides principled hypotheses for how the visual system might do this.

Another nice example of pattern estimation is the study of color constancy by Brainard & Freeman (1997). Specifically, they considered the task of estimating the (natural) illuminant from the responses of three sensors to an image produced by illuminating a random sample of natural reflectance patches. They find that the maximum *a posteriori* (MAP) estimate, which weights all errors equally, performs poorly compared to a maximum local mass (MLM) estimate, where the utility decreases as a Gaussian function of the error size. Further, their results show that a Bayesian ideal observer can, under at least some realistic circumstances, achieve a good estimate of the illuminant and hence good color constancy. More recently, Brainard et al. (2006) find that a similar ideal observer with a single prior over natural illuminants and surface reflectance functions tracks the successes and failures of human color constancy.

These examples illustrate the potential value of ideal observer theory for gaining insight into the computational issues associated with decoding the responses of retinal neurons and for developing principled hypotheses for behavioral and neural performance.

Contour and region grouping

To interpret natural images the visual system makes use of various perceptual grouping mechanisms, which work to organize local image features into clusters (groups) that are likely to derive from the same physical source (e.g., the same physical object). Some of the presumably more primitive perceptual grouping mechanisms are those that combine local contour elements into extended contours and those that combine local texture elements into extended regions. Ideal observer theory has been useful for gaining insight into how these mechanisms should work and has provided a useful benchmark for comparison with human performance in perceptual grouping tasks.

Applications of ideal observer approaches to contour grouping began about the time that measurements of the statistical properties of contours in natural images began to appear (Kruger 1998; Geisler, et al. 2001; Sigman et al. 2001; Elder & Goldberg 2002). These studies clearly demonstrated the Gestalt principle of good continuation has a solid physical basis in the structure of natural images, and hence that some form of this principle would be useful for linking contour elements into groups that correspond to the same physical source (contour). Feldman (2001) proposed a Bayesian model for subjective contour integration and Geisler et al. (2001) proposed a Bayesian model for performance in contour detection (integration) tasks like those developed by Field, Hayes and Hess (1993). However, neither of these models is a true ideal observer. To date, deriving the ideal observer for standard contour integration tasks has proven to be intractable. Yuille, Fang, Schrater & Kersten (2003) devised a more restrictive generative model for contour shape and background noise where it is possible to derive the ideal observer for a naturalistic contour detection (integration) task. They find that human observers approximate ideal performance, but are most efficient in detecting approximately straight contours and contours that obey “ecological” statistics.

More recently, Geisler & Perry (2009) derived an ideal observer for a contour occlusion task, where the observer is presented with two contour elements (taken directly from natural images) and separated by a circular occluder. The observer's task is to decide whether the two elements belong to the same or different physical contours (sources). The ideal decision rule was derived directly from the pair-wise statistics of contours in natural images, and is essentially a big table (see Figure 4a). For any pair of contour elements across the occluder, either one of them can be regarded as the reference element. This element is represented by the horizontal black line segment at the center of Figure 4a. Every other line segment drawn in the figure represents a particular geometrical and contrast polarity relationship to the reference element where the ideal decision rule is to respond that the elements belong to the same physical contour. Every possible line segment not drawn in the diagram represents a relationship where the ideal decision rule is to respond that the elements belong to different physical contours. Comparison of ideal observer performance with human performance in the contour occlusion task (without feedback) shows that there is essentially no part of the stimulus space where humans systematically deviate from the ideal decision rule in Figure 4a. The overall percentage of variance in the human responses (7 subjects) predicted by the ideal observer was approximately 99%; the percentage of variance predicted for hits, correct rejections, false alarms and misses is shown in Figure 4b (note that the scatter at low frequencies is expected given the logarithmic plots). These results demonstrate that the human visual system has incorporated, to close approximation, the decision rule in Figure 4a, which represents the average statistics of contours in natural scenes. Geisler and Perry

argue that the results also reject other proposed models of (and algorithms for) contour completion and contour integration.

Shape, depth and cue integration

Understanding how the visual system recovers the three dimensional structure of the environment from a pair of two dimensional retinal images is one of the most central and complex issues in vision science. To gain deep understanding will require a rigorous description of the relationship between 3D structure and the properties of retinal images, and a rigorous description of how those image properties are encoded and decoded by the eye and brain. The Bayesian ideal observer framework is well suited for addressing both parts, although most studies have focused on the second. Specifically, it has long been known that there are various retinal image properties that alone can provide some information about depth and shape. In recent years, the Bayesian framework has been elegantly applied to the question of how these various image properties (cues) are encoded and combined by the visual system. The development of Bayesian approaches to this and other complex perceptual problems began approximately 20 years ago (e.g., Clarke & Yuille 1990; Blake, Bühlhoff & Sheinberg 1993; Kersten, Mamassian & Yuille 2004, and the collection of papers in Knill & Richards 1996).

Although relatively little is known about the detailed statistical relationship between natural images and the 3D structure of natural environments, it is well known that there are a vast number of scene properties that are extracted and combined by the visual system when estimating depth and 3D shape. The Bayesian ideal observer framework is well suited to addressing the issue of how the visual system combines different scene properties to determine 3D structure (Blake, Bühlhoff & Sheinberg 1993; Landy et al. 1995; Yuille & Bühlhoff 1996).

In simple tasks where the scene properties are statistically independent and Gaussian, the optimal estimate for the combined cues is the weighted sum of the estimates computed for each cue separately, where the weights depend on how reliable are the individual estimates. For example in the case of two cues

$$\widehat{\omega} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \widehat{\omega}_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \widehat{\omega}_2 \quad (2)$$

where $\widehat{\omega}_1, \widehat{\omega}_2$ are the two estimates and σ_1^2, σ_2^2 the variances of the estimates.

Experimental evidence demonstrates that under a number of circumstances human observers combine cues in approximately this fashion (Knill & Saunders 2003; Mamassian & Landy 2001; Ernst & Banks 2002; Jacobs 2002). Two examples are illustrated in Figure 5. Knill & Saunders (2003) examined how texture and stereo cues (Figure 5a) are combined in a surface-slant discrimination task. The dashed and dotted curves in Figure 5b show measured slant difference thresholds for texture and stereo cues alone, as a function of the baseline slant. The solid black curve shows the measured thresholds for both cues, and the other solid curve shows the parameter-free prediction of equation (2) obtained by estimating the relative variances from the thresholds for the single cue conditions. Ernst & Banks (2002) examined how visual stereo and haptic cues (Figure 5c) are combined in a height discrimination task, where the variability of the stereo information was directly manipulated. According to equation (2) the relative weight placed on the visual information should decline as the level of noise added to the stereo signal increases, as shown by the solid curve and gray area in

Figure 5d. The solid symbols in the figure show that in this task human observers adjust their weights in an approximately optimal fashion. Similar findings have been reported for visual and auditory spatial localization (Alais & Burr 2004).

These elegant studies were designed so that the multiple cues provide approximately statistically independent sources of information, making the simple linear rule of equation (2) the ideal rule. However, in many situations such statistical independence does not hold (Clark & Yuille 1990; Landy et al. 1995; Yuille & Bülhoff 1996), and hence the ideal rules are nonlinear and can be more difficult to derive. The simple cue integration studies were also designed to consider tasks where there is only one *a priori* “model” of the world (e.g., the world contains a single object whose properties are sampled from fixed probability distributions). However, real-world tasks are often best described as estimation given a mixture of *a priori* models (Yuille & Bülhoff 1996; Knill 2003). For example, in a more natural localization task there might be two *a priori* models of stimulus generation: one model might be of a single animal that makes a sound and a visible movement, and the other model might be of two animals, one that makes a sound and one that makes a visible movement (Körding et al. 2007). In the former case, the two cues should be combined to get a single estimate of location, whereas in the latter case there should be two separate estimates of location. In other words, the task involves two concurrent sub-tasks: discrete model selection and continuous location estimation. Finally, the simple cue integration studies were designed so that ideal observer predictions for the combined cues could be generated from the thresholds in the single cue conditions, without explicitly processing the input stimuli.

A number of ideal observer analyses of shape and depth estimation move beyond the simple cue integration paradigm. Hogervorst & Eagle (1998; 2000) measured velocity and acceleration discrimination of local patches of dots undergoing uniform motion and then used those measurements to generate ideal observer predictions for structure-from-motion tasks. They found that the ideal observer predicts many aspects of human discrimination thresholds and bias in the structure-from-motion tasks. Knill (2003) and Körding et al. (2007) derive ideal observers for tasks that involve model selection and find that the ideal observers predict important aspects of human performance in slant estimation of textured surfaces and in spatial auditory-visual localization, respectively. Liu & Kersten (1995; 1998) derived various ideal observers for recognizing, from static monocular images, 3D jointed stick-shaped objects under conditions where the 3D objects can be randomly transformed (e.g., rotated) relative to observer's viewpoint. Using these ideal observers they were able to reject certain classes of existing 2D template-based models for 3D object recognition.

In addition to formal ideal observer analyses, the principles of Bayesian ideal observers can be used to make qualitative predictions. This approach has produced a number of novel insights into shape and depth perception, particularly concerning the interactions between shape, lighting, and surface properties (for reviews see Mamassian et al. 1998; Kersten et al. 2004).

The visual system has evolved and learned to perform certain tasks in natural environments and hence Bayesian ideal observer models are most likely to provide principled quantitative hypotheses for brain mechanisms in the cases where the stimuli match those occurring in natural tasks. Thus, derivation of the most relevant ideal observers for shape and depth perception must await measurement of the relevant natural scene statistics (Geisler 2008). Nonetheless, even without such measurements the studies described above show that principled hypotheses can be obtained by deriving Bayesian observers based on plausible generative models of the mapping between the environment and the retinal images. These ideal observers and the experiments motivated by them have produced major advances in

our understanding of both the computational principles of shape and depth estimation and how humans do it.

Motion

Motion is a ubiquitous and fundamental property of retinal stimulation under natural viewing conditions. However, there have been relatively few attempts to study motion perception from the perspective of ideal observer theory, in part because of the complexity of the stimuli and the related difficulty in measuring the relevant natural scene statistics (Geisler 2008). Early studies compared human and ideal observer performance for 2D motion direction discrimination (Watamaniuk 1993) and 3D heading discrimination (Crowell & Banks 1996) in random dot displays. The ideal observers (or modest modifications of them) predicted many aspects of the human performance in these tasks. More recently, Weiss et al. (2002) considered the influence of the prior distribution of motion velocity in natural scenes on the estimation of 2D motion direction. They note that when retinal image information is poor (e.g., contrast is low) an ideal Bayesian observer will put greater reliance on the prior probability distributions and bias its estimates accordingly. Under the assumption that the prior probability for local speed decreases monotonically as local speed increases, Weiss et al. are able to qualitatively explain a number of motion illusions. The ideal Bayesian observer framework has also been used to motivate and explain powerful demonstrations of the role of shadow motion on the perception of motion in depth (Kersten et al. 1997).

Optimal features

As described above, ideal observer theory is typically used to determine optimal performance given precisely specified tasks, stimuli, and biological constraints. However, ideal observer theory can be used in a rather different way; namely, to determine the statistical properties (features) of natural images that are optimal for performance in specific natural visual tasks. This application has its origin in information theory where a classic goal is to find the encoding of input signals and decoding of output signals that maximizes the transmission rate through a noisy communication channel; i.e., find the encoding and decoding that maximizes the *mutual information* between the input and output of the channel (Cover & Thomas 1991). Applications of information theory in neuroscience have been primarily concerned with characterizing the information transmitted by a given neural circuit (see Reinagle 2001 for a review). In computer vision applications, Ullman et al. (2002; 2007) determined which image patches (object parts), when used as feature templates, maximize the mutual information between the distribution of input stimuli and distribution of output responses in a natural object identification task. They obtained excellent performance relative to other methods that had been applied to similar tasks.

Similarly, Geisler et al. (2009) propose and demonstrate a general method for determining optimal features for performing natural tasks, which is based on direct application of Bayesian ideal observer theory. Their method is illustrated in Figure 6 for the task of identifying which side of a surface boundary in foliage images corresponds to the foreground. A large set of ground-truth training patches, rotated to a canonical vertical orientation, was used as input. The goal of the method is to learn the set of linear weighting functions (RFs) that provide the best identification accuracy when the responses of units having those weighting functions are processed with the Bayesian optimal decoder. In terms of equation (1), these RFs correspond to the parameters θ and the vector of their responses corresponds to $\mathbf{g}(\mathbf{S};\theta)$. The method produces sensible RFs (see Figure 6) that perform better on test samples than RFs determined by other more generic procedures (e.g., principle components analysis). Thus, ideal observers (and similar concepts in information theory)

have great potential for providing deep insight into the statistical properties of natural scenes that are relevant for performance in natural tasks.

Overt and covert attention

Efficient performance in most natural tasks requires attention mechanisms that are able to dynamically select image locations or image properties for certain kinds of specialized processing. For example, in multiple fixation visual search tasks, overt attention mechanisms must select locations to direct the specialized high-resolution processing available in the fovea. Poor fixation selection can greatly increase search time. Furthermore, within each fixation, or in a single fixation search task, covert attention mechanisms must select retinal image locations where image features are allowed to contribute to the decision about target location. If features from irrelevant locations are not suppressed at some processing stage prior to behavior, then performance will necessarily be degraded (e.g., see Doshier et al. 2004; Palmer et al. 2000). In recent years, ideal observer theory has been used to determine what would be optimal overt and covert attention mechanisms for different tasks. These ideal attention mechanisms have served both as a baseline for comparison with human performance and as a starting point for proposing principled models of attention.

Two examples from the visual search literature serve to illustrate these applications of ideal observer theory. Eckstein (1998) measured target detection accuracy in a single fixation search task where the number of potential target locations (set size) was varied from 2 to 12 (see Figure 7a). On each trial, the subjects remained fixated on the central cross, the potential locations of the target were cued with rectangular boxes, and the subjects judged (in a 2AFC task) whether the display contained the target (open ellipse or tilted ellipse in the three example displays). The symbols in Figure 7b show the performance accuracy of one subject for the three different types of search display. The solid curves labeled “parallel” show the predictions of an ideal attention mechanism that can select just the cued locations under the assumption that the (effective) internal noise limiting detection performance is statistically independent at each potential target location. Ideal performance necessarily declines with set size because of the greater chance that some noise features will be mistaken for target features. The dashed curve labeled “serial” shows the prediction of a suboptimal attention mechanism that can only select one location independent of set size. Interestingly, under these conditions exactly the same optimal attention mechanism and parallel subsequent processing predicts performance with both the “feature” and “conjunction” type displays, which have traditionally been viewed as involving quite different processing (Triesman & Gelade 1980; Wolfe 2000). This example demonstrates that the ideal observer approach can be useful for identifying those conditions under which human attention mechanisms are near optimal and, of course, when they are not.

Najemnik & Geisler (2005; 2008) measured the speed, eye movements and accuracy for localizing a small target in a multiple fixation search task, where the contrasts of the target and background noise texture were varied (see Figure 7c). Search always began at the center of the display; the white dots and lines show a hypothetical fixation sequence. Najemnik and Geisler derived an ideal searcher for this task that is limited only by the variable sensitivity of the human visual system with retinal location. To characterize this sensitivity map, they directly measured target detectability (d') at many different retinal locations in 2AFC detection task, with the target location cued on each trial (see Figure 1b). The ideal searcher is limited by this d' map, but otherwise, in the search task, it processes the entire display in parallel, optimally updates after each fixation the posterior probability of the target being at each possible location, and then uses an optimal overt attention mechanism to select the next fixation. This optimal attention mechanism considers each possible fixation location and chooses the one that on average will produce the highest probability of correctly locating the

target after the eye movement is made. Najemnik and Geisler also considered suboptimal attention mechanisms including random and maximum *a posteriori* (MAP) selection. MAP selection is an important alternative because it is equivalent to the common sense strategy of directing fixations to the peripheral display locations with features most similar to the target. They found that human search time (median number fixations) and accuracy were similar to that of the ideal searcher and MAP searcher, and much better than that of the random searcher (ruling out all possible random search models). As shown in Figure 7d, they also found that the distribution of human fixation locations in the search display was more similar to the ideal than to the MAP searcher or random searcher (not shown). The curious asymmetric fixation distributions predicted by the ideal and MAP searchers are due to the fact that the human retinal d' map is elongated in the horizontal direction (i.e., humans can detect the target further into the horizontal periphery than the vertical periphery). Because of this asymmetry the ideal searcher fixates more in the top and bottom of the display and the MAP searcher fixates more in the sides of the display. Thus, humans are qualitatively more like the ideal searcher, which chooses fixations to gain the most information about where the target is located, and less like the MAP searcher, which chooses locations that “look” most like the target. This example demonstrates how ideal observer theory can provide deep insight into complex tasks and how it can generate novel and sometimes quite unanticipated predictions.

Future

Normative models, especially Bayesian ideal observers, have been fruitfully applied to a rapidly expanding range of problems in vision science over the last 25 years. There is every reason to think that this trend will continue, especially given the pace of advances in statistical modeling/mathematics and computational power.

An ultimate goal for basic vision science is to understand and predict visual performance in natural tasks, and thus it seems likely that there will be a growing trend to develop ideal observers for increasingly naturalistic tasks (Geisler & Ringach 2009). Ideal observer analysis is particularly important in the study of natural tasks, because as tasks and stimuli become more complex, it becomes harder to intuit what kinds of neural computations would be sensible or adequate (let alone optimal) and hence harder to generate plausible hypotheses, design experiments, or interpret results. As the examples presented here illustrate, even in relatively simple tasks the predictions of an ideal observer can be quite unexpected (except in retrospect), and hence can provide novel and deep insight into the problem under investigation.

Developing ideal observers for naturalistic tasks presents a number of challenges. The most fundamental challenge is to characterize natural tasks and natural stimuli. Some of the tools for quantitatively measuring and characterizing the statistical properties of natural signals relevant for specific (usually simple) tasks are in place and others are being developed. On the other hand, the science of identifying and characterizing natural tasks is not as well developed. It is obvious that most natural tasks involve some mixture of perceptual and motor subtasks, and there have been some attempts to empirically identify sequences of subtasks at a coarse level (e.g., see Land & Hayhoe 2000). However, it is likely that even brief perceptual subtasks are most properly characterized as being composed of subtasks. For example, as mentioned earlier, many natural perceptual tasks might best be characterized as a concurrent combination of two subtasks: discrete model selection and continuous estimation (Knill 2003; Kording et al. 2007).

If a natural task and the relevant natural scene statistics can be characterized, then the next challenge is to derive the appropriate ideal observer. There has been progress in finding

methods for deriving ideal observers for perceptual tasks with naturalistic stimuli (see references in this paper) and methods for deriving ideal controllers for sensorimotor systems with realistic mechanical properties (e.g., see Kording & Wolpert 2006). The next step of deriving ideal agents that optimally process naturalistic stimuli and optimally control naturalistic motor systems in natural tasks may soon be taken.

Conclusion

The aim of this article was to summarize some of the contributions made by ideal observer theory to vision research in the last 25 years. There are many other interesting examples that could have been selected both in psychophysics and in neurophysiology. Nonetheless, these examples amply illustrate the power of the ideal observer approach both in providing theoretical insight and in guiding experimental design. It seems certain that ideal observer theory (and information theory) will play an even larger role in the future of basic and applied vision science.

Acknowledgments

David Brainard and David Knill provided helpful comments and suggestions. Supported by NIH grants EY02688 and EY11747.

References

- Ahumada AJ. Perceptual classification images from vernier acuity masked by noise. *Perception*. 1996; 25:18.
- Ahumada A Jr. Classification image weights and internal noise level estimation. *Journal of Vision*. 2002; 2(1):121–131. [PubMed: 12678600]
- Arnou TL, Geisler WS. Visual detection following retinal damage: Predictions of an inhomogeneous retino-cortical model. *SPIE Proceedings: Human Vision and Electronic Imaging*. 1996; 2674:119–130.
- Banks MS, Bennett PJ. Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates. *Journal of the Optical Society of America A*. 1988; 5:2059–2079.
- Banks MS, Geisler WS, Bennett PJ. The physical limits of grating visibility. *Vision Research*. 1987; 27(11):1915–1924. [PubMed: 3447346]
- Banks MS, Sekuler AB, Anderson SJ. Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America A*. 1991; 8(11): 1775–1787.
- Barlow HB. Increment thresholds at low intensities considered as signal/noise discriminations. *Journal of Physiology, London*. 1957; 136:469–488. [PubMed: 13429514]
- Barlow HB. The efficiency of detecting changes of density in random dot patterns. *Vision Research*. 1978; 18:637–650. [PubMed: 664351]
- Beckmann PJ, Legge GE. Preneural limitations on letter identification in central and peripheral vision. *Journal of the Optical Society of America A*. 2002; 19(12):2349–2362.
- Blake A, Bülthoff HH, Sheinberg D. Shape from texture: ideal observers and human psychophysics. *Vision Research*. 1993; 33:1723–1737. [PubMed: 8236859]
- Bloj MG, Kersten D, Hurlbert AC. Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*. 1999; 402:877–79. [PubMed: 10622251]
- Boyaci H, Doerchner K, Snyder JL, Maloney LT. Surface color perception in three dimensional scenes. *Vis Neurosci*. 2006; 23:311–21. [PubMed: 16961962]
- Brainard DH, Freeman WT. Bayesian color constancy. *Journal of the Optical Society of America A*. 1997; 14(July 1997):1393–1411.
- Brainard DH, Longere P, Delahunt PB, Freeman WT, Kraft JM, Xiao B. Bayesian model of human color constancy. *Journal of Vision*. 2006; 6:1267–1281. [PubMed: 17209734]

- Brainard DH, Williams DR, Hofer H. Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots. *Journal of Vision*. 2008; 8(5):15, 1–23. [PubMed: 18842086]
- Burgess AE, Wagner RF, Jennings RJ, Barlow HB. Efficiency of human visual signal discrimination. *Science*. 1981; 214:93–94. [PubMed: 7280685]
- Clark, JJ.; Yuille, AL. *Data Fusion for Sensory Information Processing*. Boston: Kluwer; 1990.
- Cohn TE, Lasley DJ. Detectability of a luminance increment: Effect of spatial uncertainty. *Journal of the Optical Society of America*. 1974; 64(12):1715–1719. [PubMed: 4443844]
- Cover, TM.; Thomas, JA. *Elements of Information Theory*. New York: John Wiley & Sons, Inc.; 1991.
- Crowell JA, Banks MS. Ideal observer for heading judgments. *Vision Research*. 1996; 36(3):471–490. [PubMed: 8746236]
- De Vries HL. The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica*. 1943; X(7):553–564.
- Doshier BA, Liu SH, Blair N, Lu Z. The spatial window of the perceptual template and endogenous attention. *Vision Research*. 2004; 44:1257–1271. [PubMed: 15066390]
- Eckstein MP. The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*. 1998; 9:111–118.
- Elder JH, Goldberg RM. Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*. 2002; 2:324–353. [PubMed: 12678582]
- Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 2002; 415(6870):429–433. [PubMed: 11807554]
- Feldman J. Bayesian contour integration. *Perception & Psychophysics*. 2001; 63(7):1171–1182. [PubMed: 11766942]
- Field DJ, Hayes A, Hess RF. Contour integration by the human visual system: Evidence for a local ‘association field’. *Vision Research*. 1993; 33(2):173–193. [PubMed: 8447091]
- Foley JM, Legge GE. Contrast detection and near-threshold discrimination in human vision. *Vision Research*. 1981; 21:1041–1053.
- Foley JM. Human luminance pattern vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*. 1994; 11:1710–1719.
- Geisler WS. The physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*. 1984; 1:775–782.
- Geisler WS. Sequential ideal-observer analysis of visual discriminations. *Psychological Review*. 1989; 96:267–314. [PubMed: 2652171]
- Geisler, WS. Ideal observer analysis. In: Chalupa, L.; Werner, J., editors. *The Visual Neurosciences*. Boston: MIT Press; 2003. p. 825–837.
- Geisler WS. Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*. 2008; 59:167–192.
- Geisler WS, Albrecht DG. Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience*. 1997; 14:897–919. [PubMed: 9364727]
- Geisler WS, Diehl RL. A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*. 2003; 27:379–402.
- Geisler WS, Najemnik J, Ing AD. Optimal stimulus encoders for natural tasks. *Journal of Vision*. 2009; 9(13):17, 1–16. [PubMed: 20055550]
- Geisler WS, Perry JS. Optimal point prediction statistics for natural images. 2010
- Geisler WS, Perry JS, Super BJ, Gallogly DP. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*. 2001; 41:711–724. [PubMed: 11248261]
- Geisler WS, Perry JS. Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience*. 2009; 26:109–121. [PubMed: 19216819]
- Geisler WS, Ringach D. Natural systems analysis. *Visual Neuroscience*. 2009; 26:1–3. [PubMed: 19347975]
- Gold JM, Murray RF, Bennett PJ, Sekuler AB. Deriving behavioural receptive fields for visually completed contours. *Current Biology*. 2000; 10:663–666. [PubMed: 10837252]

- Hogervorst MA, Eagle RA. Bias in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society of London B, Biological Sciences*. 1998; 265:1587–1593.
- Hogervorst MA, Eagle RA. The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *Journal of Experimental Psychology, Human Perception and Performance*. 2000; 26:934–955. [PubMed: 10884003]
- Jacobs, RA. Visual cue integration for depth perception. In: Rao, R.; Olshausen, BA.; Lewicki, MS., editors. *Probabilistic Models of the Brain: Perception and Neural Function*. Cambridge, MA: MIT Press; 2002.
- Jacobs RA, Kruschke JK. Bayesian learning theory applied to human cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2010 n/a. 10.1002/wcs.80
- Kersten D, Mamassian P, Knill DC. Moving cast shadows induce apparent motion in depth. *Perception*. 1997; 26:171–92. [PubMed: 9274752]
- Kersten D, Mamassian P, Yuille AL. Object perception as Bayesian inference. *Annual Review of Psychology*. 2004; 55:271–304.
- Knill DC. Mixture models and the probabilistic structure of depth cues. *Vision Research*. 2003; 43(7): 831–854. [PubMed: 12639607]
- Knill, DC.; Richards, W., editors. *Perception as Bayesian Inference*. Cambridge: Cambridge University Press; 1996.
- Knill DC, Saunders J. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*. 2003; 43(24):2539–58. [PubMed: 13129541]
- Koch K, McLean J, Segev R, Freed MA, Berry MJ, Balasubramanian V, et al. How much the eye tells the brain. *Current Biology*. 2004; 16:1428–1434. [PubMed: 16860742]
- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS ONE*. 2007; 2(9):e943. doi:10.1371/journal.pone.0000943. [PubMed: 17895984]
- Körding KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*. 2006; 10:319–326. [PubMed: 16807063]
- Krüger N. Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*. 1998; 8:117–129.
- Land M, Hayhoe M. In what ways do eye movements contribute to everyday activities? *Vision Research*. 2001; 41(special issue on Eye Movements and Vision in the Natural World):3559–3566. [PubMed: 11718795]
- Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*. 1995; 35(3):389–412. [PubMed: 7892735]
- Laughlin SB, Sejnowski TJ. Communication in neuronal networks. *Science*. 2003; 301:1879–1874. [PubMed: 14512620]
- Legge GE, Kersten D, Burgess AE. Contrast discrimination in noise. *Journal of the Optical Society of America A*. 1987; 4(2):391–404.
- Liu Z, Knill DC, Kersten D. Object classification for human and ideal observers. *Vis Res*. 1995; 35:549–68. [PubMed: 7900295]
- Liu Z, Kersten D. Dobservers for human 3D object recognition? *Vis Res*. 1998; 38:2507–19. [PubMed: 9798012]
- Mamassian P, Knill DC, Kersten D. The perception of cast shadows. *Trends Cogn Sci*. 1998; 2:288–95. [PubMed: 21227211]
- Mamassian P, Landy MS. Interaction of visual prior constraints. *Vision Research*. 2001; 41:2653–2668. [PubMed: 11520511]
- Manning JR, Brainard DH. Optimal design of photoreceptor mosaics: Why we do not see color at night. *Visual Neuroscience*. 2009; 26:5–19. [PubMed: 19193250]
- Murray RF, Bennett PJ, Sekuler AB. Classification images predict absolute efficiency. *Journal of Vision*. 2005; 5(2):5, 139–149.

- Nachmias J, Kocher EC. Visual detection and discrimination of luminance increments. *Journal of the Optical Society of America*. 1970; 60(3):382–389. [PubMed: 5437483]
- Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. *Nature*. 2005; 434:387–391. [PubMed: 15772663]
- Najemnik J, Geisler WS. Eye movement statistics in humans are consistent with an optimal strategy. *Journal of Vision*. 2008; 8(3):4, 1–14. [PubMed: 18484810]
- Palmer J, Verghese P, Pavel M. The psychophysics of visual search. *Vision Research*. 2000; 40:1227–1268. [PubMed: 10788638]
- Pelli, DG. The quantum efficiency of vision. In: Blakemore, C., editor. *Vision: Coding and Efficiency*. Cambridge: Cambridge University Press; 1990. p. 3–24.
- Pelli DG. Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*. 1985; 2(9):1508–1532.
- Pelli DG, Farell B. Why use noise? *Journal of the Optical Society of America A*. 1999; 16(3):647–653.
- Reinagel P. How do visual neurons respond in the real world? *Curr Opin Neurobiol*. 2001; 11:437–42. [PubMed: 11502389]
- Rose A. The sensitivity performance of the human eye on an absolute scale. *Journal of the Optical Society of America*. 1948; 38(2):196–208. [PubMed: 18901781]
- Saunders JA, Knill DC. Perception of 3D surface orientation from skew symmetry. *Vision Research*. 2001; 41:3163–3183. [PubMed: 11711141]
- Sekiguchi N, Williams DR, Brainard DH. Efficiency in detection of isoluminant and isochromatic interference fringes. *Journal of the Optical Society of America A*. 1993; 10(10):2118–2133.
- Sigman M, Cecchi GA, Gilbert CD, Magnasco MO. On a common circle: Natural scenes and Gestalt rules. *Proc National Academy of Science*. 2001; 98:1935–1940.
- Tanner WP Jr. Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*. 1961; 89:752–765. [PubMed: 13775211]
- Treisman AM, Gelade G. A feature integration theory of attention. *Cognitive Psychology*. 1980; 12:97–136. [PubMed: 7351125]
- Ullman S. Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Science*. 2007; 11:58–64.
- Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*. 2002; 5:682–687.
- Watamaniuk SNJ. Ideal observer for discrimination of the global direction of dynamic random-dot stimuli. *Journal of Optical Society A*. 1993; 10:16–28.
- Weiss Y, Simoncelli E, Adelson EH. Motion illusions as optimal percepts. *Nat Neurosci*. 2002; 5:598–604. [PubMed: 12021763]
- Wolfe JM. Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*. 1994; 1:202–238.
- Yuille, AL.; Bulthoff, HH. Bayesian decision theory and psychophysics. In: Knill, DC.; Richards, W., editors. *Perception as Bayesian inference*. Cambridge: Cambridge University Press; 1996.
- Yuille, AL.; Fang, F.; Schrater, P.; Kersten, D. Human and ideal observers for detecting image curves. Paper presented at Neural Information Processing Systems (NIPS); 2003.

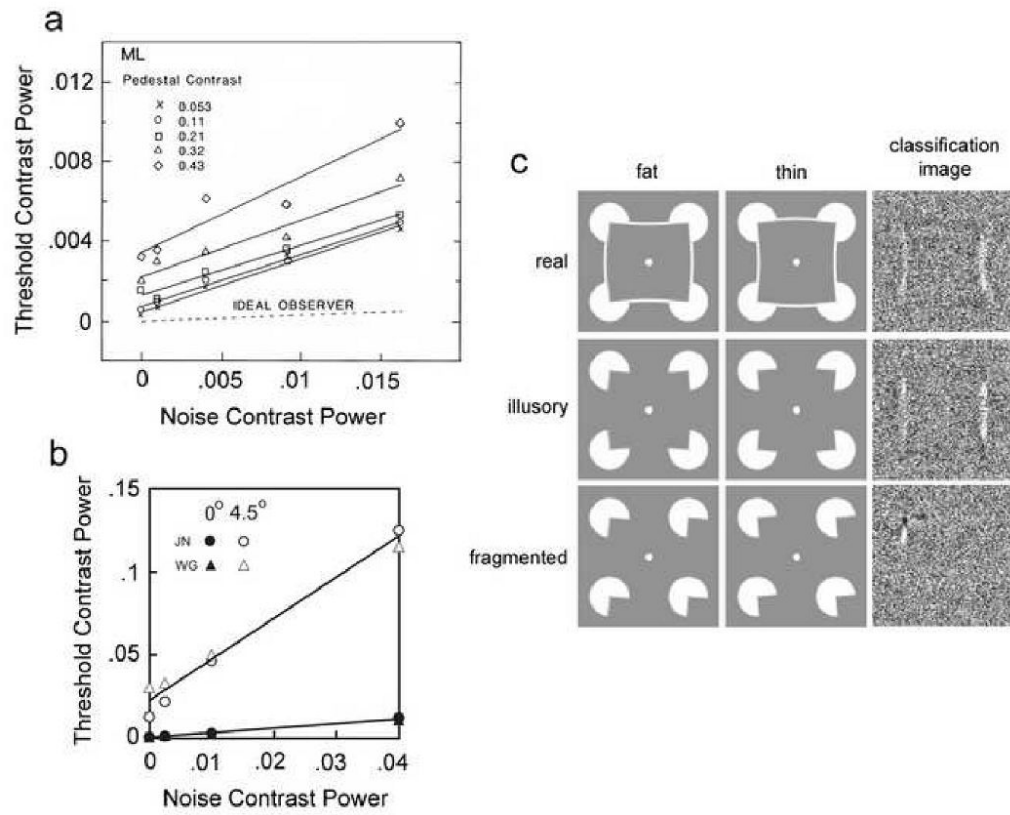


Figure 1. Detection and discrimination in Gaussian noise. **a.** Detection threshold as a function of white noise power for one observer for five pedestal contrasts. (adapted from Legge et al. 1987). **b.** Detection threshold for a 6 cpd target as a function of 1/f noise contrast for two observers at two retinal eccentricities (adapted from Najemnik & Geisler 2005). **c.** Classification images for shape discrimination in white noise (adapted from Gold et al. 2000).

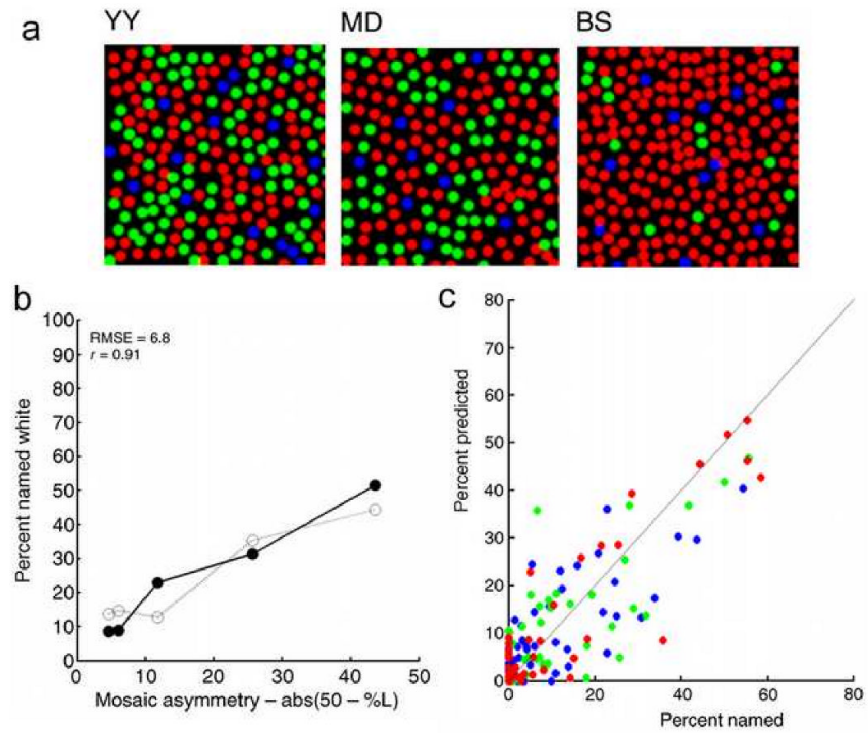


Figure 2. Optimal chromatic decoding given the human cone mosaic. (Adapted from Brainard, Williams & Hofer 2008). **a.** Cone mosaics of three (out of five) human observers measured with adaptive optics. **b.** Solid symbols show the percentage of trials where the subject reported seeing “white” when a small (receptor size) spot was flashed at a random location in the receptor lattice; open symbols show parameter free ideal-observer predictions. **c.** Observed vs. predicted color names for three different spot wavelengths ($r = 0.83$).

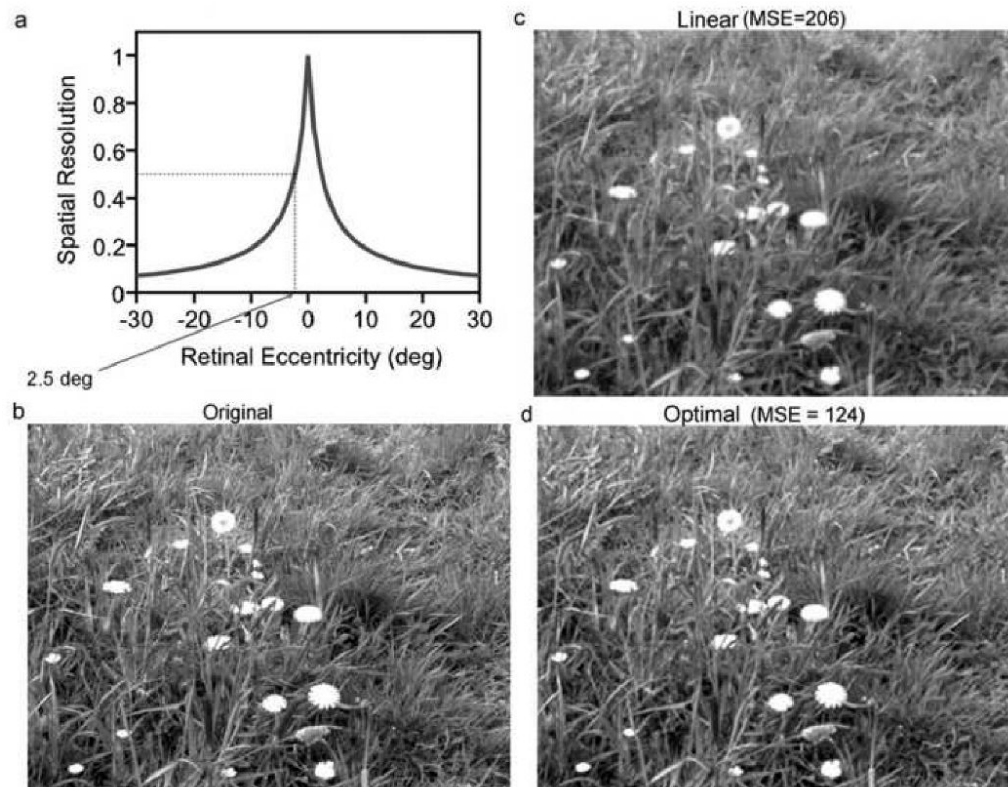


Figure 3.

Optimal decoding given peripheral ganglion cell sampling. **a.** Approximate relative spatial resolution of the human visual system as a function of retinal eccentricity. Resolution is half the foveal resolution at 2.5 deg eccentricity. **b.** Calibrated gray scale image at full (foveal) resolution. **c.** Half-resolution image (2.5 deg eccentric) upsampled using linear interpolation. **d.** Half-resolution image upsampled using ideal decoder based on natural image statistics.

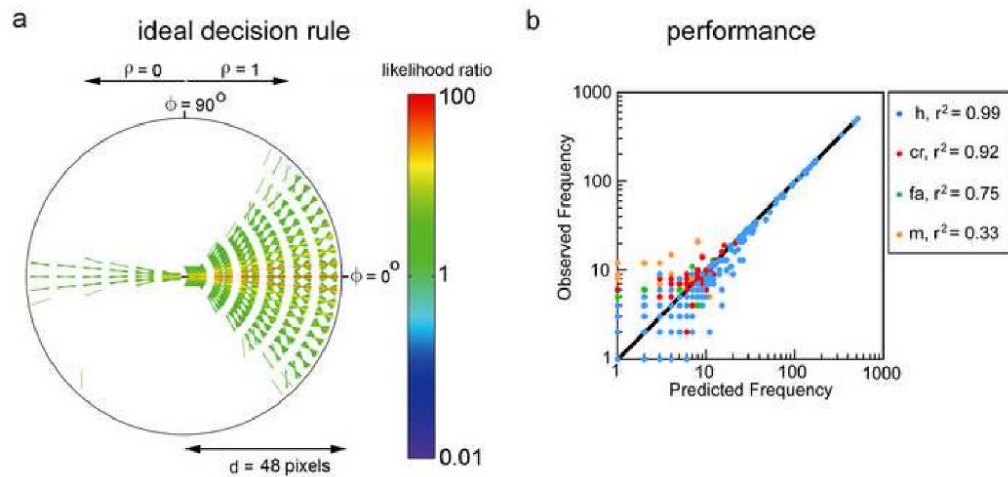


Figure 4. Optimal grouping across an occlusion. **a.** Parameter free ideal decision rule based on natural image statistics. The central black line segment represents one of a pair of contour elements. Each other line segment drawn in the figure represents a particular geometrical and contrast polarity relationship for the case when the optimal decision is that the contour elements belong to the same physical contour (source). **b.** Comparison of ideal (solid line) and human (symbols) performance in a contour occlusion task (Adapted from Geisler & Perry 2009.)

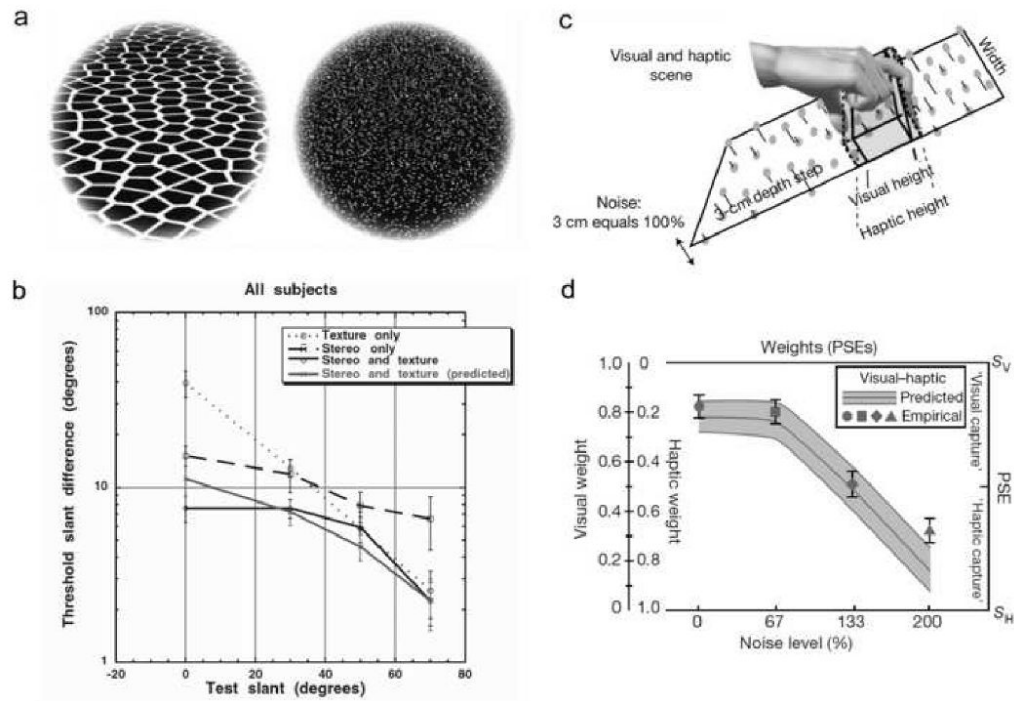


Figure 5. Cue combination in 3D perception. **a.** Stimuli for measuring how texture and stereo cues combine in surface slant discrimination. **b.** Slant threshold as a function baseline slant angle for single and combined cues. Red curve is parameter free prediction of equation (2). **c.** Stimuli for measure how stereo and haptic cues combine in size discrimination. **d.** Relative weights assigned to the two cues as a function of the amount of noise added to the stereo display. Symbols are estimated from psychophysical thresholds; solid curve and gray area are parameter-free predictions based on equation (2).

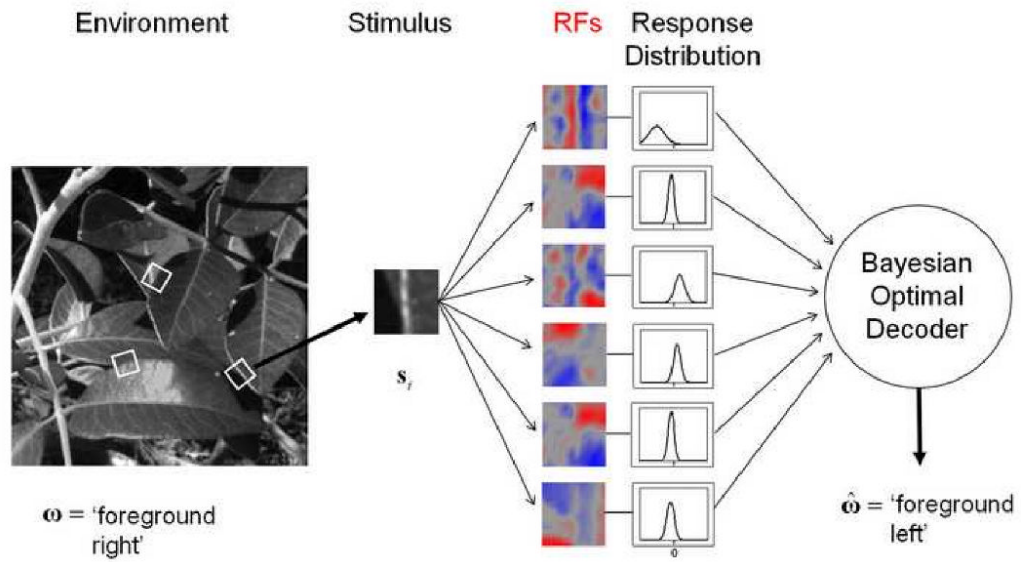


Figure 6.

Accuracy Maximization Analysis (AMA): a method of using ideal observer theory to find optimal features for natural tasks. The Bayesian optimal decoder knows the mean and variance of the responses of each RF to each stimulus in the training set. In this case, the accuracy of the decoder in the task can be approximated with a closed form expression. The receptive fields are found one at a time by modifying their shape until maximum accuracy on the training set is obtained. Final RFs for a foreground identification task are shown.

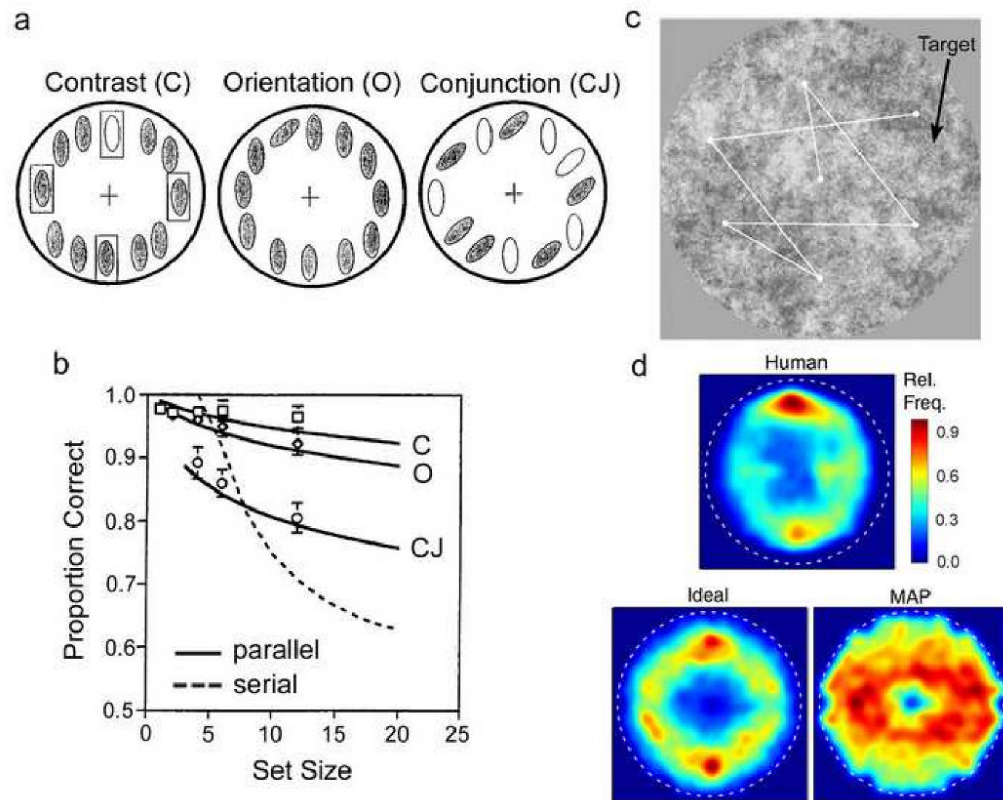


Figure 7.

Ideal observers for overt and covert attention. **a.** Schematic of three types of stimuli in a single fixation search task. The subjects' task was to indicate (following a brief presentation followed by a mask) whether or not the display contained the target object at one of the cued locations (rectangular boxes). **b.** Accuracy in the single fixation search task as a function of the number of cued locations where the target might appear, for a subject and two model searchers. Solid curves show ideal observer predictions. **c.** Stimuli for multiple fixation search task. A small Gabor target was randomly located in background texture of noise having the average power spectrum of natural images. Search began at the center of the display; the white dots and lines show a hypothetical fixation sequence. **d.** These temperature plots show the distribution of fixation locations in the display combined over all trials (excluding the first fixation which was always at the center of the display), for the human and two model searchers.