Preprint of:

Kersten, D & Mamassian, P (2009) Ideal Observer Theory. In: Squire LR (ed.) Encyclopedia of Neuroscience, volume 5, pp. 89-95. Oxford: Academic Press.

Ideal Observer Theory

Daniel Kersten
Psychology Department
University of Minnesota

75 East River Road
Minneapolis, MN 55455
USA
kersten@umn.edu

Pascal Mamassian
Laboratoire Psychologie de la Perception
CNRS & Université Paris 5

45 rue des Saints-Pères
75006 Paris
FRANCE
pascal.mamassian@univ-paris5.fr

March, 2007

Abstract

Ideal observer models are applications of Bayesian Statistical Decision Theory to problems of neural information transduction, transmission and utilization. A basic motivation is that because sensory inputs provide noisy or ambiguous information about states of the world, probabilistic methods are required to understand how reliable decisions can be made. Thus the focus is first on modeling the information for a task, independent of the observer under study, and second on comparisons of that model with a test observer, such as a human or neuron. A key rationale for such comparisons is that the ideal observer can be used to normalize performance for task difficulty. An ideal observer can also provide a starting point for modeling perceptual performance.

Keywords: sensation, perception, psychophysics, Bayesian theory, signal detection theory, ideal observer

Ideal observer theory (in-press), *The New Encyclopedia of Neuroscience*, edited by Larry Squire et al., 2008.

Introduction

An ideal observer is ideal in the sense that it achieves statistically optimal performance for a specified task. It makes the best decisions or estimates given uncertain sensory information and prior knowledge. What is "best" is specified in terms of costs and benefits. Ideal observer models can be used to study different types of real "test observers", such as in the psychophysical analysis of the input/output behavior of humans, to spike-train analysis of information carried by single neurons. If human perceptual performance approaches that of an ideal observer, this can in principle rule out a large class of candidate neural mechanisms that are suboptimal relative to the ideal. Although the ideal observer concept can be traced to earlier work, the modern notion of an ideal observer was developed in the 1950s by Peterson, Birdsall, Fox, Tanner, Swets, and others in the context of a general theory of signal detectability. The theory and its initial applications to human auditory and then visual sensitivity are described in a classic 1966 book by Green and Swets. In the late 1970s, Horace Barlow applied ideal observer analysis to higher-level perceptual tasks, such as the perception of symmetric patterns. By the 1990s, comparisons of human and ideal performance were extended to an increasingly wider set of problems, including object and motion perception, perceptual organization, reading, and motor control. Ideal observer analysis has shown, for example, that the way in which human observers integrate depth or motion information takes into account uncertainty in the input image cues, as one might expect from an ideal observer doing the same task.

Defining the ideal observer

The key ingredients in defining an ideal observer are: 1) the generative model; 2) the task requirements, including a measure of performance; and 3) an observer model that specifies the optimal action rule for the given task requirement. The ideal observer can be defined in terms of four classes of random variables on a directed graph (i.e. one with arrows) that represents how the variables influence each other, see Figure (1). These variables represent states of the world s, observations (data) x that result, actions a (e.g. decision or estimate) on the data, and the losses (costs) $L(a, s)$ of action a, given the true state s. The spaces can be discrete or continuous. For example, a state of the world could mean one of two positions of a light switch, or the distance of an object. An action may be only indirectly related to the state, e.g. when throwing a ball, the speed of release of the ball aimed at a distant target vs. an estimate of the state itself (distance). We now go over the elements in greater detail.

The generative model specifies how the states of the world, s, determine the observed data x, i.e. sensory input (e.g. pattern of image intensities). The state space can be discrete or continuous. In general, states and data are multidimensional, $s = (s_1, s_2, s_3, \dots)$ and $x = (x_1, x_2, x_3, \dots)$ and the relationship between states and data can have complex dependencies (a simplified example is shown in Figure 1B). The generative model consists of the prior probability of the state s, $p(S = s)$, and the probability of the observation x given s, $p(X = x|s)$. The prior probability can be simple, representing the probability

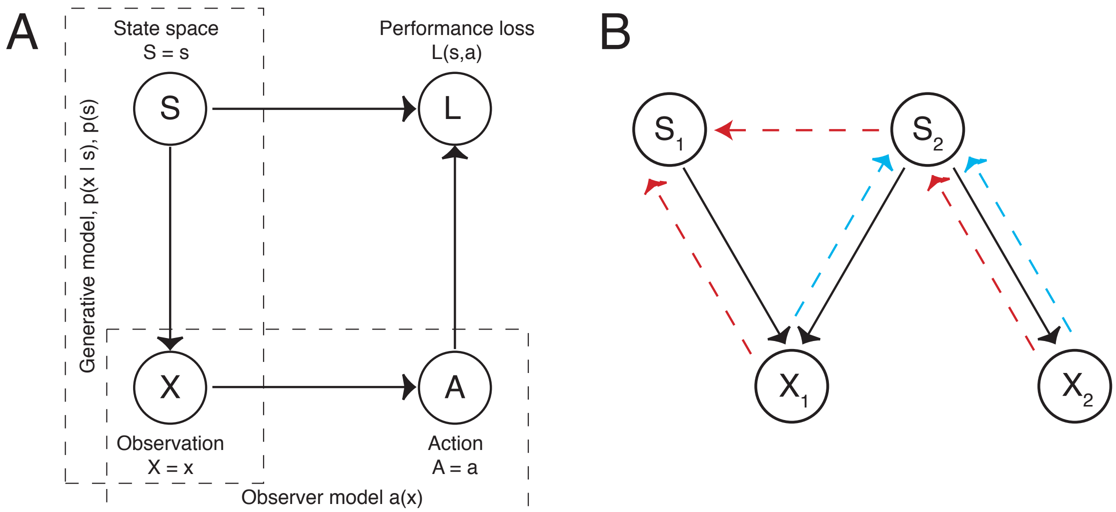

Figure 1: **A.** Components of an Ideal Observer. The nodes represent random variables, and the arrows how they influence each other. The arrows in the graph (formally, a directed acyclic graph) can be interpreted as describing the dependencies between variables. The graph structure determines how to factor the joint distribution for the task, $p(s, x, l, a) = p(l|a, s)p(a|x)p(x|s)p(s)$. From this, one can formalize the problem of estimating components, such as the probability of a loss value $p(L = l)$ (e.g. probability of error $L = 1$), or other performance measures, such as the hit rate $p(L = 0|s_H)$, or false positive rate $p(L = 0|s_L)$ for "signal" state s_H and "noise" state s_L. **B.** The state and observation space can have complex dependencies. This panel shows an example in which state variables S_1 and S_2 both influence observation X_1, but only S_2 directly influences X_2. Solid lines indicate the generative model, and the dashed lines the directions of (inverse) inference for two tasks. Suppose the goal is to estimate S_2, then both observations determine the optimal estimate (blue dashed lines). This is an example of cue integration, where the two observations are conditionally independent of S_2. If the goal is to estimate S_1 , both observations again determine the estimate, but X_2 indirectly affects beliefs about S_1 through S_2 (red dashed lines). This is an example of "explaining away". The graph structure is equivalent to the factorization $p(s_1, s_2, x_1, x_2) = p(x_1|s_1, s_2)p(x_2|s_2)p(s_2)p(s_1)$.

of an hypothesis with just two values, such as a light switch being set to high $S = s_H$ rather than low $S = s_L$. But the prior probability on s can also be used to model complex pattern regularities that take into account covariation between elements of s, such as the probability of a surface tending to have smoothly varying, rather than rapidly changing depths. In this latter case, validating the model for the prior from real world data can be a significant problem by itself; however, in the lab, an experimenter can specify the prior. The probability of an observation x given s describes how the signal s is "encrypted" in the data, and is the likelihood of the hypothesis (the state s) given the data. The description could be as simple as $x = s+n$, where x is light intensity, s is the average level of one of two fixed light intensities, and n is a sample of gaussian noise. Then $p(X = x|s) = p(x - s)$. Or $p(X = x|s)$ could be derived from a more complicated function $x = \phi(s, n)$, where x is a vector representing an image, and ϕ is a function that describes how a shape s together with a confounding variable n, such as viewpoint (or slant angle in the example below), determine the image observation. In the context of a psychophysical experiment, the generative model can be thought of as a probabilistic description of how the stimuli are generated.

The task requirements are two-fold: 1) what is to be done with the data, e.g. detect the presence or absence of the signal (e.g. "high vs. low contrast?", or "animal present or not?"), identify the signal (e.g. " which letter "A" , "B", or "C"?), estimate a continuous value (e.g. what is the depth of the object?); 2) the value of achieving or not achieving the goal. Not all elements of the state variable, $s = (s_1, s_2, s_3, \dots)$, are equally important to determine accurately. In classical signal detection theory, state variables are divided into two types: the relevant variables to be inferred accurately ("signals"), and those to be discounted ("noise") with associated losses for right and wrong answers. Generally, the value of accuracy can be expressed in terms of a "loss function" (alternatively a "gain" function, equal to one minus the loss function.) For example, different degrees of relevance can be placed on the state variables by specifying a loss function, $L(a_1, a_2; s_1, s_2)$ which puts a cost on choosing actions, a_1, a_2, when the true states are s_1 and s_2. If the action is an estimate of s, $a = \hat{s}$, loss may be expressed in terms of differences, $L(\hat{s} - s)$ (see loss function panel in Figure 3).

Optimal action or decision rule. Of the many possible observer models, the ideal observer is defined as one that uses an action rule, $a(x)$, that minimizes the risk for the task. The expected risk is defined as the loss averaged over both state and observation variables: $R(a) = \sum_{s,x} L(s, a(x))p(s, x)$. For a given observation, x, the decision rule is to choose action a so as to minimize $R(a|x) = \sum_{s,x} L(s, a(x))p(s|x)$. Often the action space has a simple relationship with the signal state space. For example, in a detection or discrimination task the state space is discrete and binary, and the action is to decide whether the signal was sent or not (answer yes or no to questions such as: "is Thomas there?" or "is Thomas bigger than Simon?"). For identification, the state space is discrete with multiple values ($S = \{Thomas, Simon, Hermann\ or Jacob\}$), and the actions are decisions based on x ("Is it Thomas? Simon? Hermann?, or Jacob?"). For estimation, the action is an estimate of the state variable, $a = \hat{s}$ from observation x (e.g. "how tall is she?"). If the loss is uniformly high for all wrong decisions, and low for the correct decision, then the optimal action is to pick the most probable state s given the data x. This is the *maximum a posteriori* (MAP) observer: $a(x) = \text{argmax}_s\, p(s|x)$. This rule yields the smallest average error. If the loss is the squared error between state and estimated state variables, then the optimal action is to choose the mean of the posterior. Next, we illustrate task requirements with examples for detection and estimation tasks.

Applications of Ideal Observer Theory

One can distinguish two types of applications of ideal observer theory. For the purposes of this article, we refer to the first as *ideal observer analysis*, where the test observer (humans or neural system) competes against the ideal observer in the same, well-defined, laboratory task. Because the experimenter controls the conditions for the stimuli and task, no observer can do better (on average) for the specified task. In this case, the performance of an ideal observer can be thought of as a benchmark against which to measure human performance. The emphasis is on departures from ideal performance which provide clues as to the underlying, generally sub-ideal, mechanisms. Ideal observer analysis is a useful addition to the experimenter's toolbox alongside linear and non-linear systems analysis. Ideal observer analysis has been used to show that humans perform some tasks with strikingly high efficiency relative to the ideal observer, but quite poorly for others. One of the historic success stories was to show that human ability to discriminate light intensity under night time viewing conditions could be extremely high. Following on the work of Hecht, Schlaer and Pirenne in the 1940s, Horace Barlow showed in the early 1960s that human quantum efficiencies were sufficiently high as to rule out any explanation that required the retinal transduction of more than a few photons.

The second application of ideal observer theory is as an approximate theory or model of performance. The primary distinction between the two applications is that the second typically makes simplifying assumptions about the generative model (see Estimation example below), including the nature of the input representation (e.g. the ideal observer may have geometrical features as input, whereas the human observer receives from image intensities as input, from which geometrical information is derived), or how the data are caused by the states of the world (e.g. gaussian noise is added, when the actual noise may be non-gaussian). In other words, the ideal observer is optimal with respect to a generative model that may differ in detail from the true generative model, either as defined in a particular laboratory study, or real world task. The primary reason for these simplifying assumptions is practical–the true ideal observer is too hard to calculate. Even though humans are in general suboptimal, an ideal observer model may nonetheless go a long way to explaining the observed behavior. In this second type of application, Ideal observer theories are equivalent to *Bayesian theories* of perceptual performance. As a modeling tool, a general advantage of the Bayesian ideal observer approach is that it avoids commitment to untestable mechanistic details, such as a particular neural architecture, while still providing quantitative predictions of behavior.

Terms	*Equations*
Conditioning	$p(x\|y) = p(x,y)/p(y)$
Marginalization	$p(x) = \int p(x,y)dy$
Bayes rule	$p(s\|x) = p(x\|s)p(s)/p(x)$
Signal-to-noise ratio for equal variance gaussian case	$d'_{snr} = \frac{\mu_s - \mu_n}{\sigma}$
Relation of z & P	$z(P): \ P = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-x^2/2} dx$
Sensitivity from false alarm and hit rates	$d'_{perf} = z(P_{FA}) - z(P_H)$
ROC in terms of z values for unequal variance case	$z(P_H) = \frac{\sigma_n}{\sigma_s} z(P_{FA}) - \frac{\mu_s - \mu_n}{\sigma_s}$
d' for 2AFC from proportion correct	$d'_{2afc} = -\sqrt{2} z(P_c)$
Efficiency	$E = (d'_{snr_I}/d'_{snr_T})^2$

Examples of Ideal observer analysis

Signal detection

Figure (2) illustrates a classic task in signal detection theory in which a switch is set to "high" or "low" and the observer has to guess the setting based on an observation, in this case a measurement of light intensity.

The generative model. The switch setting is the state variable. If the switch settings are equally likely, independent of the data, then the prior is constant and equal to a half, $p(S = s) = 1/2$. The second part of the generative model, the likelihood, specifies how the observations depend on the state. Suppose that when the switch is set to high, the average light intensity is higher than when the switch is low. However, because of trial-to-trial fluctuations (noise), the measured or observed light intensity varies. So sometimes the measured light intensity is higher for the "low" than the "high" switch setting. Note that in ideal observer analysis, one may be able to test the validity of the generative model, apart from questions of optimal inference. Thus, for example, the light switch model assumes additive gaussian noise (although at low intensities where fluctuations in photon emission and absorption may dominate, a Poisson model is a better choice.) The top panel

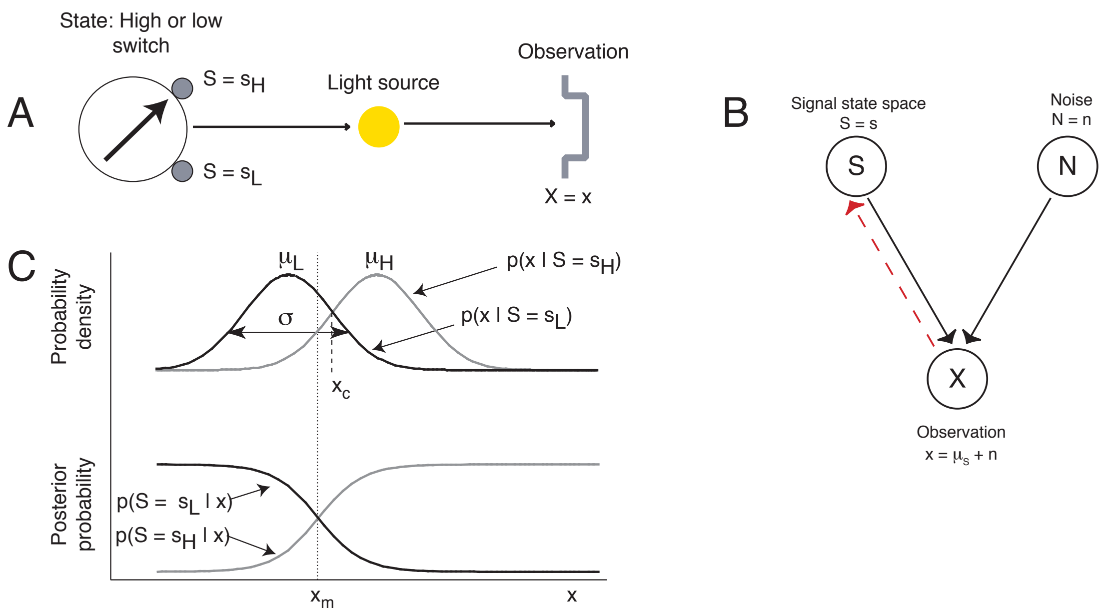

Figure 2: Example of a simple ideal observer for a binary task, signal detection. **A**. Schematic of the physical model in which a light source produces an observation x, of light intensity. **B**. Graph showing causal dependencies corresponding to factorization in terms of the components of the generative model: $p(x, s, n) = p(x|s, n)p(s)p(n)$. The direction of the inference for the detection task is shown by the dashed arrow. **C**. Top panel shows the gaussian probability densities of the observation under the two hypotheses (switch states), also known as the likelihood functions corresponding to different means, μ_L, μ_H. The standard deviations are the same, σ. The bottom panel shows graphs of the posterior probabilities. When the switch probabilities are not equal, choosing the highest likelihood no longer produces the lower error rate. The ideal observer guesses the state with highest posterior probability.

of Figure (2C) shows the gaussian distributions of observations under the two possible state conditions. The noise is additive because the high switch setting produces the same distribution as that of the lower except that the mean is shifted by an added amount. The standard deviation remains the same.

The task. Suppose that given an observation, the observer has to decide whether it was due to a high or a low switch setting. Further suppose that the task requires that the probability of error $p(L = 1)$ be as small as possible. Loss is represented by $1 - \delta_{s,\hat{s}}$ (the Kronecker delta function, $\delta_{s,\hat{s}}$, is one if $s = \hat{s}$ and zero otherwise). Another common task in psychophysics is the two-alternative forced-choice task (2AFC). The test (e.g. a human subject) and ideal observers share the same state, observation, and action space, as well as loss function ("scoring system"). In general, they differ in their mappings of observations to actions, e.g. their decision rules.

The optimal decision rule. Intuitively, one would guess that the best strategy might be to choose s_H if $p(x|s_H) > p(x|s_L)$, and in fact this is the optimal rule, for the constant prior case, if one wants to have the smallest error rate. In our example, this rule is equivalent to choosing s_H if $x > x_c$, where x_c is the point where the two curves in the upper panel of Figure (2C) cross each other. However, the general rule for minimizing error is the maximum a posteriori (MAP) rule which takes into account the prior probability, and picks the state with the higher posterior probability, i.e. choose s_H if $p(s_H|x) > p(s_L|x)$. The posterior $p(s|x)$ is determined by the prior *and* likelihood through Bayes rule: $p(s|x) = p(x|s)p(s)/p(x)$. If, for example, the prior probability of the switch setting "high" is bigger than the probability of "low", this is equivalent to moving the criterion to x_m, i.e.: choose s_H if $x > x_m$. (Compare the upper and lower panels of Figure 2C.) But there are other rules depending on the loss function one chooses. The loss function is a 2×2 "payoff" matrix with loss values for: "hit" (true positive), "false alarm" (false positive), "correct rejection" (true negative), and "miss" (false negative). One can assign distinct losses to each of these outcomes, in which for example, false alarms are more or less costly. In general, this results in a different rule r based on minimizing the risk: choose s_H if $r(s_H|x) > r(s_L|x)$. Standard results in signal detection theory show that for this problem, the optimal rule is equivalent to: choose s_H if $x > x_m$, where the value of x_m (the *criterion*) depends on the loss matrix.

How to compare test and ideal observers? The test observer can perform no better than the ideal, and in general will do worse, because of suboptimal action rules, such as systematically failing to use all of the relevant information in the input stimulus. In this specific example, an observer can be suboptimal by putting the criterion x_m at the wrong place–a "response bias". As the criterion shifts, the values of the hit and false alarm rates also shift. A graph of hit vs. false alarm rates is known as the Receiver Operating Characteristic (ROC). A key contribution of signal detection theory was to show that underlying human sensitivity to a sensory signal (d') could be teased apart from the value of the criterion.

In theory, the ideal and test observer can be compared in terms of their "scores" (average loss, e.g. proportion of errors) or d'_{perf} given the same states and resulting observations. However, it is often not practical to directly compare error rates (when the ideal is making a modest number of mistakes, the test observer may be near chance and would thus require too many measurements to estimate d'_{perf_T}); alternatively one can compare state parameters in the generative model that produce identical performances (e.g. sensitivities d'_{perf_T}, d'_{perf_I} for test and ideal in an equal-variance gaussian noise detection task are determined by the signal-to-noise ratios d'_{snr_T} and d'_{snr_I}). Suppose, for example, that a human and ideal observer both have the same proportion correct in a 2AFC task, then the ideal observer can be used as a benchmark, where a standard measure of comparison is statistical efficiency $E = (\#\ samples\ required\ by\ ideal / \#\ samples\ required\ by\ test\ observer)$. In the case of additive gaussian noise, this is equivalent to $E = (d'_{snr_I}/d'_{snr_T})^2$. If the noise is the same, efficiency takes on the simple form of the ratio of the squared threshold values of the ideal to the human. In the case of detecting an image pattern in additive gaussian noise, an observer can be suboptimal by failing to use all of the image samples (pixels), or by performing as if there was additional noise. These two sources of suboptimality have been teased apart in psychophysical experiments, measured in terms of calculation (or central) or transduction efficiencies. Efficiency provides a unit-free measure of performance that allows for the stimulus complexity and task constraints.

Estimation

Perception is critical to a wide range of tasks, of which signal detection is just one. Consider the problem of determining the dimensions of a three-dimensional object from a two-dimensional projected image. In this case, a major source of uncertainty results from loss of information due to projection. Figure 3 illustrates a Bayesian ideal observer model for a simple task to estimate the height and slant for a class of elliptical disks. Given an elliptical shape on the retina, what object could have produced it? It could be an elliptical disk, but a flat circular disk could also produce an elliptical image.

The generative model. For simplicity, assume that the "world" consists of elliptical disks of width equal to one, but with various heights h. These disks can have various slants, α, with respect to a viewer. The states are continuous, and can be represented by, $s = (s_1, s_2) = (h, \alpha)$. One can hypothesize a prior, $p(s) = p(h, \alpha)$ on these values. Figure 3 shows the graph of a bivariate gaussian prior that assumes the average slant is $\pi/4$ and the average height is 1. If viewed from a far distance, the width of the projection of any disk from this state space is a constant value (say 1); but the height of the projection depends on both the true height of the disk and the slant. An observation of the height, x of the retinal image provides data that constrains but, without additional information, does not uniquely determine the possible states. In the previous example, uncertainty resulted from added gaussian noise. Here, although we assume some additional measurement noise, the critical uncertainty results from the property that the surface could be slanted away from the viewer by various unknown degrees. For a distant view, the image observation x can be approximated as $x \sim h\cos(\alpha) + n$. Assuming gaussian noise, this determines the likelihood function, $p(x - h\cos(\alpha))$, where $p(n) = e^{-n^2/2\sigma^2}/(\sigma\sqrt{2\pi})$ is the standard formula for a gaussian distribution with zero mean.

The task. Given an observation, say $x = 1/2$, the ideal observer needs to determine the height and slant with the least risk. Figure 3 illustrates the steps. It first computes the likelihood of the observation for all possible values of the state variables. The likelihood has a high ridge of constant height, meaning that there are infinitely many pairs of values of height and slant that are equally likely. Thus, for example, a height of 4 inclined at 83° is no less likely than a head-on slant of zero together with a height of 1/2. This ambiguity can be resolved by multiplying the likelihood by the prior distribution to obtain the posterior distribution, $p(h, \alpha|x)$. At this point, one could stop, and pick off the peak values of the posterior as the most probable values of height and slant. This would correspond to the decision rule for the MAP observer, described above. However, suppose the task is more complex, requiring greater precision, let's say, in estimating the slant than the height.

The optimal decision rule. In this more general case, we use a rule that minimizes risk. A simple model of risk is to define loss in terms of the absolute values of the errors, i.e. the difference between the true and estimated state variables. Then assign a uniformly low loss to errors of height that fall within a wide range, but penalize slant errors more narrowly, as shown by the loss function in Figure 3. The risk is then a convolution of the loss function with the posterior: $R(h, \alpha|x) = \int p(h', \alpha'|x)L(h - h', \alpha - \alpha')dh'd\alpha'$, and the decision rule is to pick the pair of state variables corresponding to the lowest values of R. It is straightforward to show that if the tolerance to errors in height and slant are infinitely wide and narrow, respectively, then height effectively drops out of the risk, and minimizing risk corresponds to MAP estimation on the marginal posterior, $p(\alpha|x)$.

How to compare test and ideal observers? Even if human estimates differ in detail from an ideal observer model, one can still test the degree to which information is optimally combined. Consider the posterior term, which is proportional to the product of the likelihood and prior terms. The ideal observer combines the image cues and the prior information in such a way as to take into account the reliability of each of the sources. Thus, smaller variances on the prior would bias the final estimate toward the prior means, and conversely highly reliable (e.g. noise free) image data would move the estimates of slant and/or height toward values consistent with the data. Task differences also affect shifts in the optimal estimates of the state variables. Although this example illustrates the balance between prior and likelihood, there is a similar trade-off in the case of cue integration. For example, if there was additional information for the slant of the surface (e.g. a texture gradient), then the ideal observer would combine and weight the cues according to their reliabilities. For the case of conditional independence and gaussian noise, the optimal cue weighting is given by: $s_{opt} = \frac{s_1 r_1}{r_1 + r_2} + \frac{s_2 r_2}{r_1 + r_2}$, where s_i is the estimate of the state variable from cue i, and r_i is the reciprocal of the variance for that cue.

Related areas

Ideal observer theory has close connections to several other areas of neuroscience and cognitive science. Optimal control theory has a similar structure to that illustrated in Figure (1)A. In applications to motor control, the state variables change continuously in time and correspond to internal physical parameters that influence a movement. The Bayes risk function is replaced by a cost function (e.g. physical energy of a movement), and the action rule is replaced by a control law.

Ideal observer analysis can be naturally extended to optimal learning. Bayes optimal learning corresponds to updating the parameters (e.g. mean and variance) of the posterior distribution as new data comes in. Rather than using the data to infer the causes or states of the generative model, one uses Bayes rule to learn the values of the parameters that determine the posterior.

Computational techniques such as expectation-maximization (EM) and Bayesian belief propagation, have provided the means to deal with more complicated and non-gaussian generative models, with applications to inference and learning. Future work should produce a growing number of applications to the experimental study of cognitive and neural processes.

Comparisons of ideal and human perceptual behavior inevitably lead to the question of the nature of the underlying neural mechanisms that would support ideal-like computation. One of the central issues is how information about uncertainty may be represented and processed in the nervous system.

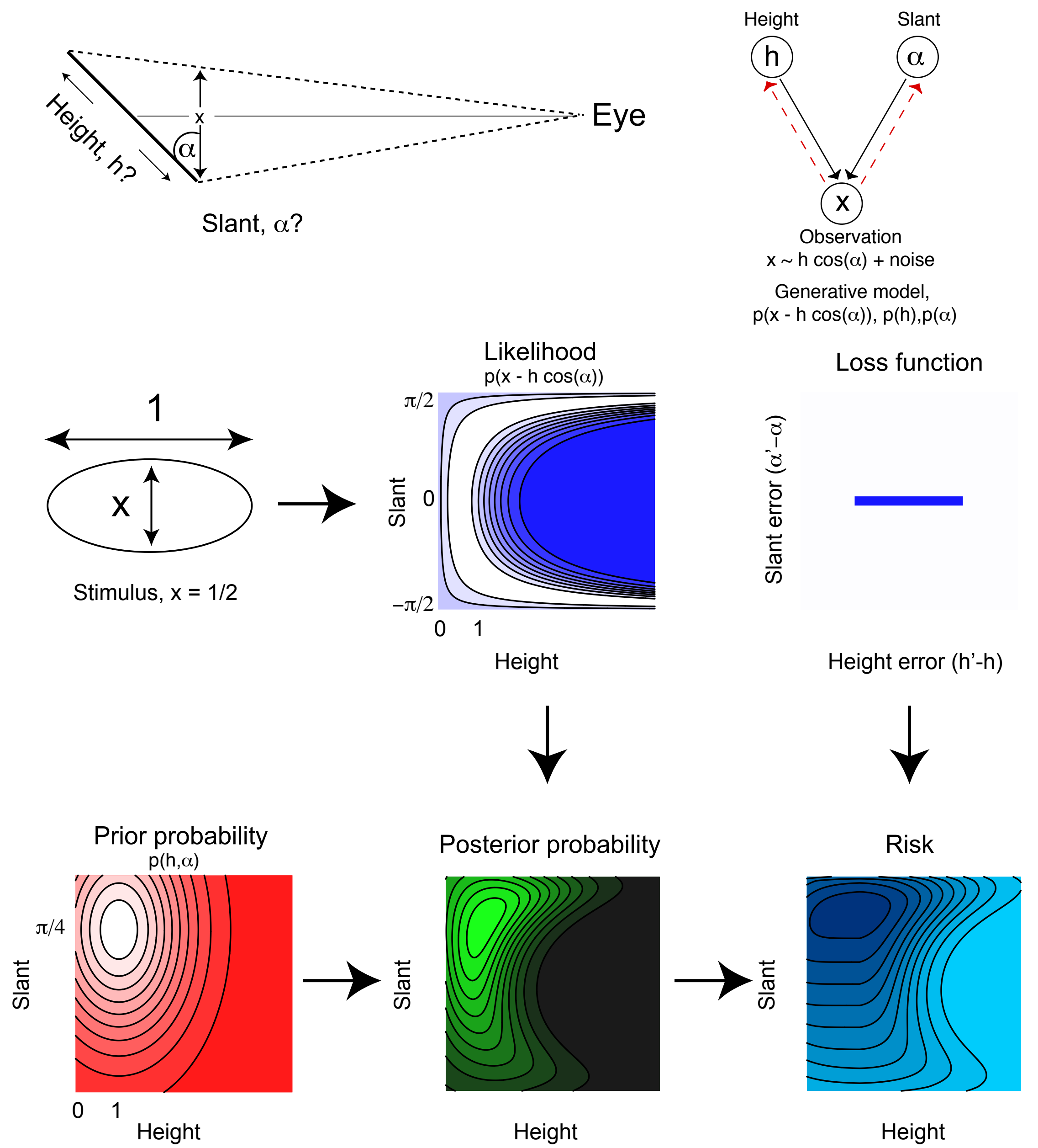

Figure 3: Example of ideal observer theory applied to object perception. See main text for details.

.

Further readings

Barlow, H. B. (1962). A method of determining the overall quantum efficiency of visual discriminations. *J. Physiol. (Lond.)*, **160**, 155-168.

Doya, K., Ishii, S., Pouget, A., & Rao, R.P.N. (Eds.) (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. The MIT Press. Cambridge MA.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: Wiley.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**(6870), 429-433.

Eckstein, M. P. (1998). The lower efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, **9**, 111-118.

Green, D.M., & Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. Wiley. New York.

Griffiths, T.L. & Yuille, A.L. (2006) Technical Introduction: A Primer on Probabilistic Inference. Department of Statistics, UCLA. Department of Statistics Papers. Paper 2006010103. http://repositories.cdlib.org/uclastat/papers/2006010103

Geisler, W. S. (2003) Ideal Observer analysis. In: L. Chalupa and J. Werner (Eds.), The Visual Neurosciences. Boston: MIT press, 825-837.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271-304.

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn Sci*, **10**(7), 319-326.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge University Press. Cambridge.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Res*, **46**(28), 4646-4674.

Tanner, W. P. J., Birdsall, T. G., & Clarke, F. R. (1960). *The Concept of the Ideal Observer in Psychophysics* (Technical Report No. 98). University of Michigan, Communications and Signal Processing Laboratory (CSPL) http://hdl.handle.net/2027.42/7891

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature*, **31,**, 419-422.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci*, **5**(6), 598-604.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford ; New York: Oxford University Press.