

A Bayesian approach to the evolution of perceptual and cognitive systems

Wilson S. Geisler*, Randy L. Diehl

*Department of Psychology, Center for Perceptual Systems,
University of Texas at Austin, Austin, TX 78712, USA*

Received 27 May 2002; received in revised form 13 December 2002; accepted 19 December 2002

Abstract

We describe a formal framework for analyzing how statistical properties of natural environments and the process of natural selection interact to determine the design of perceptual and cognitive systems. The framework consists of two parts: a Bayesian ideal observer with a utility function appropriate for natural selection, and a Bayesian formulation of Darwin's theory of natural selection. Simulations of Bayesian natural selection were found to yield new insights, for example, into the co-evolution of camouflage, color vision, and decision criteria. The Bayesian framework captures and generalizes, in a formal way, many of the important ideas of other approaches to perception and cognition.

© 2003 Cognitive Science Society, Inc. All rights reserved.

Keywords: Natural selection; Ideal observer; Scene statistics; Color perception; Camouflage evolution

Perceptual and cognitive systems, including the developmental and learning mechanisms that shape them during the lifespan, are the result of evolution by natural selection. Yet historically most approaches to the study of perception and cognition acknowledge only implicitly the role of natural selection. Here we propose a Bayesian theoretical framework that makes explicit the relationship between the statistical properties of the environment, the evolving genome, and the design of perceptual and cognitive systems. The proposed framework grew out of recent applications of Bayesian statistical decision theory in perception and cognition and recent efforts to measure the statistical properties of natural environments; however, as we will see, the Bayesian framework encompasses many important insights of previous theoretical approaches to perception and cognition. In what follows, we briefly summarize the

* Corresponding author. Tel.: +1-512-471-5380; fax: +1-512-471-7356.

E-mail address: geisler@psy.utexas.edu (W.S. Geisler).

formal Bayesian framework and show how it can be used to test specific hypotheses about the design of perceptual and cognitive systems. We then describe some connections between the Bayesian framework and other theoretical approaches to perception and cognition.

1. Bayesian framework

In recent years, the concepts of Bayesian statistical decision theory have transformed research in perception by providing a rigorous mathematical framework for representing physical and statistical properties of the environment, describing the tasks that perceptual systems perform, and deriving appropriate computational theories of how to perform those tasks optimally, given the properties of the environment and the costs and benefits associated with different perceptual outcomes. These computational theories of optimal performance are referred to as “ideal observers.” Derivation of ideal observers can be a very useful exercise because it forces one to consider carefully the various constraints that apply to the perceptual task, and because the ideal observer provides an appropriate benchmark against which to compare human performance. Furthermore, an ideal observer can be modified in various ways, making it an excellent starting point for developing testable models (Barlow, 1977; Green & Swets, 1966; Geisler, 1989). Although ideal observer theory has been applied primarily to perceptual tasks, it has also been applied to various cognitive tasks.

Specification of an ideal observer requires descriptions of: the task, the probability of each possible state of the environment before the stimulus is received (prior probability distribution), the probability of each possible stimulus given each possible state of the environment (likelihood distribution), and the costs and benefits associated with each possible combination of response and state of the environment (utility function). Traditionally, ideal observers were considered only for laboratory tasks with relatively simple stimuli. For example, a task might involve detecting a spot of light against a white noise background in a two-alternative forced choice paradigm, with the subject’s goal being to maximize detection accuracy (i.e., maximize average utility, where the utility is positive for correct responses and negative for errors). In these cases, the prior probability distributions, likelihood distributions, and utility functions are directly given by the experimental design.

Recently, investigators have begun considering ideal observers for complex natural tasks and stimuli. This is an important development because it speaks directly to the design of perceptual systems through evolution and learning. In these cases, the prior probability and likelihood distributions are based on measurements of physical and statistical properties of natural environments. For example, if the task in a given environment is to detect edible fruit in background foliage, then the prior probability and likelihood distributions are estimated by measuring representative spectral illumination functions for the environment and spectral reflectance functions for the fruits and foliage. In a similar vein, utility functions should be based on measurements of the costs and benefits associated with different stimulus–response outcomes in natural environments. At the most fundamental level, the utility function must be the “fitness” function—the birth and death rates associated with different stimulus–response outcomes. Geisler and Diehl (2002) referred to ideal observers with such utility functions as “maximum fitness” ideal observers.

While ideal observer theory provides an appropriate benchmark for evaluating perceptual and cognitive systems, it will not, in general, accurately predict the design and performance of real systems, which are limited by a number of factors that will be described later. Here, we mention only one of these factors: Evolution through natural selection is an incremental process where each change must produce an increase in fitness; thus the real observer may correspond to a local maximum in the space of possible solutions, whereas the ideal observer corresponds to the global maximum in the space of possible solutions.

Geisler and Diehl (2002) proposed an extended Bayesian framework to represent these additional factors that may limit performance. The framework consists of two complementary components: a maximum-fitness ideal observer theory that allows inclusion of specific anatomical and physiological constraints, and a Bayesian formulation of the theory of natural selection referred to as “Bayesian natural selection.”¹ In the Bayesian formulation of natural selection, each allele vector² (i.e., each instance of a polymorphism) in each species under consideration is represented by a fundamental equation, which describes how the number of organisms carrying that allele vector at time $t + 1$ is related to (1) the number of organisms carrying that allele vector at time t , (2) the prior probability of a state of the environment at time t , (3) the likelihood of a stimulus given the state of the environment, (4) the likelihood of a response given the stimulus, and (5) the fitness given the response and the state of the environment. The process of natural selection is represented by iterating these fundamental equations in parallel over time, while updating the allele vectors using appropriate probability distributions for mutation and sexual recombination. A great strength of this framework is that maximum-fitness ideal observer theory and Bayesian natural selection are specified in terms of exactly the same prior probability distributions, likelihood distributions, and utility functions. Recognizing that mechanisms of learning and plasticity are also a result of natural selection, we believe that this extended Bayesian framework is appropriate for analyzing real perceptual and cognitive systems and for understanding their relationship to optimal design.

Next, we describe in more detail the individual components of the Bayesian approach.

1.1. Properties of the natural environment

The importance of characterizing natural environments for understanding biological design has been widely recognized ever since Darwin. However, until recently, there have been few attempts to describe quantitatively the physical properties and regularities of natural environments, in part because of the focus on laboratory tasks and in part because of difficulties in making the necessary measurements. In the domain of perception, many relevant properties of the environment can be characterized by measuring natural scene statistics. Some natural scene statistics that have been measured include energy spectra of natural sound sources (Attias & Schreiner, 1997), illumination spectra of natural light sources (Dixon, 1978; Judd, MacAdam, & Wyszecki, 1964; Regan et al., 2001), reflectance spectra of natural surfaces (Krinov, 1947; Lythgoe & Partridge, 1989; Regan et al., 2001; Osorio & Vorobyev, 1996), distributions of edge orientation in natural images (Coppola, Purves, McCoy, & Purves, 1998; Switkes, Mayer, & Sloan, 1978), spatial frequency spectra of natural images (Burton & Moorehead, 1987; Field, 1987), distributions of distance along the visual axis in natural scenes (Huang, Lee, & Mumford, 2000), co-occurrence statistics of geometrical relationships between edges in natural images

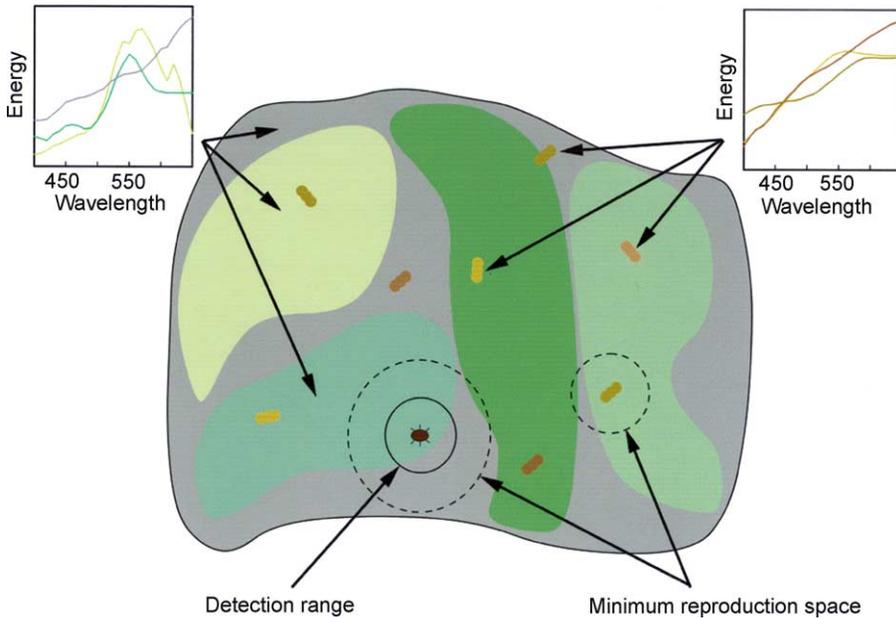


Fig. 1. A hypothetical world consisting of a mobile predator species (the six-legged organism) and a static prey species (the rod-shaped organisms) in a simple environment. The insets illustrate representative chromatic spectra arising from the background and from the surface of the prey. The dashed circles indicate minimum reproduction spaces, which set the maximum density of predators and prey that the environment could support under the best of circumstances (for each species considered separately). The solid circle represents the effective detection range of the predator's sensory system.

(Elder & Goldberg, 1998; Geisler, Perry, Super, & Gallogly, 2001; Sigman, Cecchi, Gilbert, & Magnasco, 2001), and co-occurrence statistics of spatial frequency components in natural images (Simoncelli & Schwartz, 1999). For some other examples of measurements of natural scene statistics, see Purves and Lotto (2003).

The importance of such measurements for understanding the design of perceptual systems is illustrated in Fig. 1, which represents a simple hypothetical world that will be used here for illustrating the concepts of Bayesian ideal observers and Bayesian natural selection. In this world, there exist two competing species, a mobile predator species (e.g., an insect) with a visual system, and a static prey species (e.g., a plant) with no perceptual system. It is obvious that success of the predator will depend critically on its ability to detect prey and that success of the prey will depend on its ability to avoid detection. Detection performance may be determined, for example, by the chromatic properties of the prey and of the background. These properties are usefully described by probability distributions for the wavelength spectra reflected from the prey and from the background. Several samples of such spectra are shown in the insets of Fig. 1. An efficient perceptual system in the predator is one that responds as differently as possible to the prey and background spectra, and an efficient camouflage system is one in which the prey and background spectra are as similar as possible. Thus, one would expect both the perceptual system of the predator and the surface characteristic of the prey to reflect the chromatic scene statistics of their environment.

Two questions that come to mind when considering natural scene statistics are whether it is practical to measure them, given the variety and complexity of natural environments, and whether they are regular and structured enough to be usefully exploited by the brain. In many cases, the answer to both questions is yes. For example, the sample spectra shown in Fig. 1 are actual spectra measured in a particular natural environment. As can be seen, they are relatively broad and smooth. Further, there is considerable similarity in shape among different natural spectra. In fact, using principal components analysis it can be shown that natural reflectance spectra and natural illumination spectra are each adequately described (>99% of the variance accounted for) by simple models with just a few parameters (Maloney, 1986). Thus, not only are natural spectra highly regular and structured, but the number of dimensions needed to describe them is small enough that the probability distributions over these spaces can be estimated from a moderate number of samples. Analyses of samples of natural spectra have shown that locations of the peak spectral sensitivities of cone photopigments in dichromatic mammals are often nearly optimal for the discrimination of foliage (Lythgoe & Partridge, 1989; Chiao, Vorobyev, Cronin, & Osorio, 2000). Similarly, locations of the peak spectral sensitivities of the cone photopigments in trichromatic primates are often nearly optimal for discrimination of fruit from foliage (Osorio & Vorobyev, 1996; Regan et al., 2001). Furthermore, the statistics of natural reflectance and illumination spectra appear to be exploited by the brain to solve many complex tasks, including perceptual grouping, visual search, and identification of surface reflectance independent of illumination (Maloney & Wandell, 1986; Brainard & Freeman, 1997; Regan et al., 2001).

Another example concerns the geometry of edge elements extracted from natural images. All possible geometrical relationships between pairs of such elements can be described by three parameters: distance, direction, and orientation difference. Measurements of probability distributions over this three-dimensional space reveal two important regularities that appear to hold in every natural image: (a) nearby edge elements tend to be parallel, (b) edge elements that belong to the same physical contour tend to be nearby and co-circular (tangent to the same imaginary circle). Models of perceptual grouping based directly on these probability distributions are able to predict human contour detection performance in complex backgrounds under a wide range of conditions (Geisler et al., 2001). These results show that the natural scene statistics of edge geometry are very regular and structured, and suggest that they are incorporated rather precisely into perceptual grouping mechanisms.

In the Bayesian framework, natural scene statistics as well as known physical properties of the environment (e.g., the principle of inertia, the Poisson character of light, the Doppler shift in sound transmission) are represented by prior probability and likelihood distributions.³ The prior probability distribution describes the probability of each possible state of the environment prior to the stimulus arriving at the organism. The possible states of the environment can be represented by a vector, $\omega = (\omega_1, \omega_2, \dots)$, whose components are the various environmental properties (including the properties of other species) relevant to the perceptual or cognitive task under consideration. In the example world of Fig. 1, these components might include the reflectance and illumination spectra for the background and prey, and the number of prey within detection range. It is important to emphasize that the components that need to be included in the environment vector depend upon the task. If the task is to detect potential mates rather than prey, then the relevant components might be the reflectance and illumination spectra for

the background and conspecifics, and the number of other conspecifics within detection range. Here, we represent the prior probability distribution with the notation $p(\omega; t)$.

The parameter t in the prior probability distribution is needed to represent the fact that some environmental properties may change over time. For example, because of predation, the prior probabilities of the number of prey and of the reflectance spectra of the prey may change over time, whereas the prior probability for the reflectance spectra of the background may be relatively stable over time. Obviously, being able to adjust to changing prior probabilities can be valuable for the organism, and indeed many forms of learning and plasticity can be conceptualized as mechanisms for dynamically adjusting to changing prior probabilities. On the other hand, learning and plasticity mechanisms are unnecessary, and often lead to decrements in performance, if applied to environmental properties that are statistically stable over time. (We return to this point later.)

Given a particular state of the environment (a value of ω), there will be a probability distribution (the likelihood distribution) over the set of possible stimuli that could reach the organism. A stimulus can be represented by a vector, $s = (s_1, s_2, \dots)$, whose components are the dimensions relevant to the perceptual or cognitive task under consideration. In the case of light, the dimensions might be intensity, location (direction), wavelength, and time; in the case of sound, the dimensions might be intensity, frequency, and time. We represent the likelihood distribution, given a particular state of the environment, with the notation, $p(s|\omega)$. The likelihood distribution describes the mapping between states of the environment and proximal stimuli. For example, in the case of vision, it might represent the effects of projective geometry on images formed in the eye, and in the case of hearing, the effects of the inverse square law on sound pressure level at the ear drum.

1.2. Tasks

Perceptual and cognitive systems have evolved because they enhance performance in tasks relevant to survival and reproduction, such as detecting and localizing predators or prey, navigating through the environment, identifying materials, estimating the three-dimensional geometry of the environment, recognizing specific objects encountered before, and remembering advantageous behaviors for a previously encountered local environment. In the Bayesian framework, precise definition of perceptual and cognitive tasks is a crucial first step. It is impossible to determine what counts as optimal performance, or even what properties of the environment are relevant, without specifying the task.

1.3. Costs and benefits

Specifying the task requires a description of the costs and benefits associated with each possible combination of state of the environment (ω) and response (r). In the Bayesian framework, these costs and benefits are represented by a utility function. In the typical laboratory study, the utility function is defined by the experimental design, either explicitly or implicitly. For example, typical instructions are to maximize accuracy or to minimize response time without making errors. In a categorization task, maximizing accuracy would correspond to a utility function with a positive value for correct responses and a negative value for incorrect

responses. Minimizing response time without making errors might correspond to a utility function that is a positive-valued decreasing function of response time for correct responses and a negative-valued function for errors.

In natural tasks the utility function is generally more complex. Natural tasks that organisms perform can be fully understood only within the context of Darwinian natural selection, and in natural selection the utility function is a fitness function—the birth and death rates associated with different combinations of environmental states and responses. For example, if the task is navigating to a food source at a given location, then the fitness function may place the greatest value on reaching the food while minimizing expenditure of energy and time and exposure to physical harm, because this maximizes the chances for survival and reproduction. For some natural tasks (e.g., detection of a food source), the fitness function may reduce to a simple utility function such as maximizing accuracy, but generally this will not be the case. The appropriate utility (fitness) function for natural tasks is the growth factor, $\gamma(\omega, r)$, associated with each possible combination of state of the environment and response. This growth factor is equal to $1 + \text{birth rate} - \text{death rate}$. (Thus, if the birth and death rates are equal there will, on average, be no change in population size.)

1.4. Ideal observer analysis

Once the task, the environmental prior probabilities, the stimulus likelihoods, and the utility function are specified, it becomes possible in principle to determine optimal performance in the task by deriving a Bayesian ideal observer. As an illustration of how this is done, consider the task of judging shape and orientation of an object in the environment from its image on the retina. Suppose, in particular, that the task is to estimate the aspect ratio and slant of a thin elliptical disk. The retinal image (which is also an ellipse) is inherently ambiguous because many combinations of aspect ratio and slant of the disk can give rise to exactly the same image. Thus, the stimulus in Fig. 2 can be seen as an elliptical disk in the frontal plane or a circular disk in a slanted plane. To interpret the image, an ideal observer first computes the likelihood function $p(X|\alpha, \sigma)$, that is, the probability of the aspect ratio X of the image, for each possible combination of aspect ratio α and slant σ of the object.⁴ The C-shaped distribution in Fig. 2 shows the likelihood function (with the darkness indicating the value of the likelihood). The solid curve shows the ridge of maximum likelihood corresponding to the combinations of aspect ratio and slant of the object that are exactly consistent with X ; the other non-zero likelihoods occur because of noise in the image and in the measurement of X . Next, the ideal observer multiplies the likelihood function by the prior probability distribution $p(\alpha, \sigma)$ to obtain a function proportional to the posterior probability distribution. In this hypothetical example, we are assuming that the aspect ratios of objects in the environment tend to be closer to 1.0, and that flat objects tend to lie parallel to the ground plane, and hence the top of the object tends to be farther away. Thus, the prior probability distribution peaks at an aspect ratio of 1.0 and a slant of 45° . Finally, the ideal observer convolves the posterior probability distribution with the utility function $\gamma(\hat{\alpha}, \hat{\sigma}, \alpha, \sigma)$ to obtain the expected (or average) utility associated with each possible estimate of aspect ratio $\hat{\alpha}$ and slant $\hat{\sigma}$. The final estimate is the one with the maximum expected utility (the black dot in the lower right panel). It can be shown that this is the best possible estimate, given the probability distributions and utility function specified in this hypothetical task.

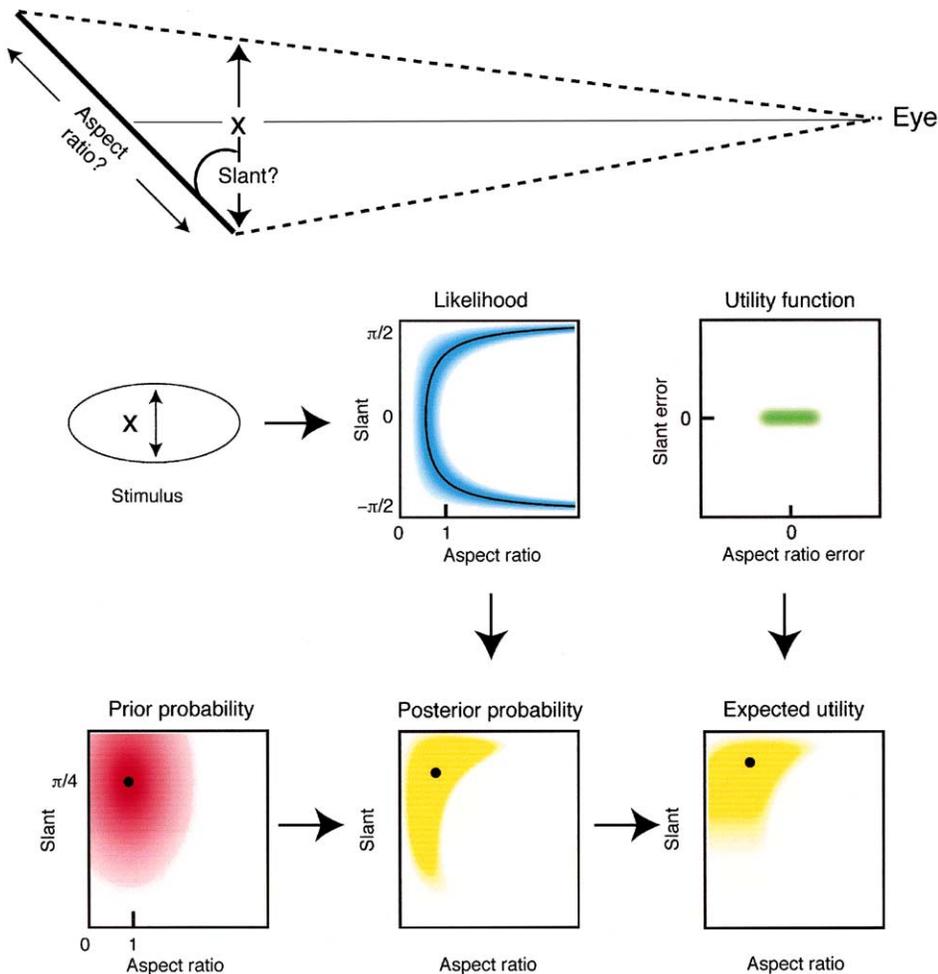


Fig. 2. Example of a Bayesian ideal observer for estimating the aspect ratio and slant of a tilted ellipse from the proximal stimulus formed on the retina. When a stimulus is received, the ideal observer computes the likelihood of receiving that stimulus for each possible combination of aspect ratio and slant (i.e., for each possible interpretation). It then multiplies this likelihood function by the prior probability distribution for each combination of aspect ratio and slant to obtain the posterior probability distribution—the probability of each possible interpretation given the stimulus. Finally, the posterior probability distribution is convolved with a utility function, representing the costs and benefits of different levels of perceptual accuracy, to obtain the expected utility associated with each possible interpretation. The ideal observer picks the interpretation that maximizes the expected utility, as indicated by the black dot in the lower right panel. Black dots and curves indicate the maxima in each of the plots. (From Geisler & Kersten, 2002.)

Although this example is admittedly rather contrived, it suffices to illustrate the various kinds of measurements and calculations needed to determine the performance of a Bayesian ideal observer. In a real application involving natural tasks, the prior probability and likelihood distributions would be most appropriately obtained from measurements of natural scene statistics and other environmental properties. Also, the utility (fitness) function would be most appropriately obtained from measurements of such variables as energy cost, reproductive value,

physical risk associated with response alternatives, and nutrient value of food sources. Making these measurements is plainly a daunting task, but there has been real progress in measuring natural scene statistics (as described earlier) and in measuring variables relevant to the utility function (Alexander, 1982; Parker & Maynard Smith, 1990). Furthermore, for some tasks, in some situations, the number of important environmental properties and fitness variables might be fairly restricted.

It should be emphasized that even in the case of complex natural tasks the general formulas of Bayesian statistical decision theory still apply. Specifically, upon receiving a particular stimulus, S , a maximum-fitness ideal observer will pick the response r_{opt} that maximizes fitness averaged over all possible states of the environment ω :

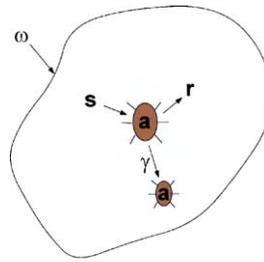
$$r_{\text{opt}} = \arg \max_r \left(\sum_{\omega} \gamma(r, \omega) p(S|\omega) p(\omega; t) \right) \quad (1)$$

Notice that the operations contained in this equation are precisely the operations illustrated in Fig. 2, with the optimal response corresponding to the optimal estimate of aspect ratio and slant.⁵

1.5. Bayesian natural selection

Maximum-fitness ideal observers give us the appropriate benchmark against which to compare the performance of existing perceptual systems, and they can help us understand the computational demands of the task. This is very useful, but organisms are generally not ideal. The design of the ideal observer is determined solely by the task, prior probabilities, stimulus likelihoods, and utility function. The design of real perceptual systems is determined not only by these factors, but by other factors as well. First, evolution by natural selection is an incremental process where each change must produce an increase in fitness; thus, real organisms may correspond to a local maximum in design space, rather than to the global maximum of the ideal observer. Second, evolution and learning lag behind changes that occur in environmental likelihoods and prior probabilities, and thus an organism's design may not correspond to even a local maximum. Third, there are limitations on the possible structure of organic molecules and on materials that organisms can synthesize. Fourth, species evolve to perform many tasks leading to compromises in design that may result in non-ideal performance in some tasks.

How can we assess the contributions of all these factors? We suggest that the appropriate framework is a Bayesian formulation of Darwin's theory of natural selection. To explain the Bayesian formulation, we first note that natural selection involves a complex interaction between environment, behavior, and reproduction (see Fig. 3). At a particular point in time, there is some prior probability distribution over the possible states of the environment, $p_a(\omega; t)$, and given some particular state of the environment, a likelihood distribution over the possible stimuli reaching the organism, $p_a(s|\omega)$.^{6,7} Given a stimulus, the organism will respond in some fashion. This response is, in general, probabilistic and is ultimately controlled by some allele vector carried in the organism. This allele vector represents particular versions of the genes at the relevant set of gene locations for the task under consideration. Of course, saying that the responses are ultimately controlled by some allele vector does not exclude learning and plasticity mechanisms, but simply recognizes that such mechanisms also result from natural



ω = state of environment
 s = stimulus
 a = alleles
 r = response
 γ = growth factor

Fig. 3. Key variables in the Bayesian formulation of Darwin's theory natural selection. The vector ω represents the state of the environment. The components of this vector are those aspects of the environment relevant to the particular task under consideration. In Bayesian terminology the probability distribution for ω is called the "prior probability distribution." The vector s represents a stimulus arriving at the organism. In Bayesian terminology the probability distribution of s given ω is called a "stimulus likelihood distribution." The vector r represents the response of the organism. The response of the organism is also probabilistic and depends upon the organism's species, allele vector, and stimulus. In Bayesian terminology the probability density function for r given s is called the "response likelihood distribution." Finally, the growth factor γ consists of one plus the birth rate minus the death rate, which both depend upon the response, the state of the environment, the organism's species, and the organism's allele vector a . The allele vector represents a particular version (polymorphism) of the relevant genes carried by the organism.

selection. The probability distribution describing the organism's response (given the stimulus) is represented by the term $p_a(r|s)$.

The response of the organism in the particular environment will have consequences for survival and reproduction. Thus, there will be a growth factor function, $\gamma_a(r, \omega)$, that depends on the response, the state of the environment, and possibly the allele vector. If the growth factor is greater than 1, then the number of organisms carrying the allele vector will on average increase. However, if a set of alleles is successful, it will have an inevitable and eventually powerful feedback effect on the environment. Ultimately, equilibrium with the environment must be reached (the average growth must converge to one or becomes less than one). In other words, the prior probability distribution must change over time.

This account translates directly into the Bayesian formula shown in Fig. 4, which expresses the mean number of organisms carrying allele vector a at time $t + 1$ in terms of the actual number of organisms at time t , the prior probability of different states of the environment at time t , the likelihood of different stimuli given the state of environment, the likelihood of different responses given the stimulus, and the growth factor associated with each combination of response and state of the environment.⁸ To represent natural selection, a separate equation is created for each allele vector (polymorphism) in each species under consideration, and all equations are iterated in parallel. Additional equations (not shown here) represent the sampling distributions for births, deaths, mutations and sexual recombination.

This Bayesian formulation has several virtues. It neatly separates the process of natural selection into its key components, which can be considered individually and then combined to

$$\bar{O}_a(t+1) = O_a(t) \sum_{\omega} p_a(\omega; t) \sum_r \gamma_a(r, \omega) \sum_s p_a(r|s) p_a(s|\omega)$$

Fig. 4. The fundamental equation of Bayesian natural selection shows how the expected number of organisms of a given species carrying a given vector of alleles a at time $t + 1$ is related to the number of organisms carrying the same alleles at time t . Each different vector of alleles in each species under consideration is represented by a separate fundamental equation; all of the equations are evaluated and iterated in parallel. Probability distributions for actual births, actual deaths, mutation, and sexual reproduction are described by other equations.

describe the evolutionary process as a whole. It provides a formal framework for understanding how natural scene statistics and genetics interact. Finally, it uses the same prior probability distributions, likelihood distributions, and utility function as the Bayesian ideal observer, which is critical for understanding the relationship between optimal and actual design.

However, like the general formulas for Bayesian ideal observers (e.g., Eq. (1)), this is a very abstract description, and thus most of the scientific effort will come in applying the framework to specific cases. To begin exploring how the framework might be applied, we (Geisler & Diehl, 2002) conducted several simulations of co-evolution for the hypothetical species and environment shown in Fig. 1. Recall that there are an active predator species with a visual system and a passive prey species (e.g., a plant) whose only defense is camouflage. The perceptual system of the predator has two classes of receptor, a color-opponent mechanism, and a decision criterion. If the color-opponent signal exceeds the criterion, then the predator expends the energy to approach a potential target; otherwise, it continues searching. Two genes of the predator are assumed to evolve: one controlling the peak spectral sensitivity of one of the receptor classes, and the other controlling the decision criterion. (The peak spectral sensitivity of the other receptor class is fixed and matched to the background foliage.) The prey has no perceptual system and only a passive response determined by its surface reflectance. A single gene, controlling the surface reflectance, is assumed to evolve.

The growth factor function (utility function) for the predator is determined by two factors: availability of space to reproduce and success of foraging (i.e., whether the response results in a hit, false alarm, correct rejection or miss). Availability of space to reproduce is determined by the current population density of conspecifics relative to the maximum possible density, which is the total space divided by the minimum reproduction space (the larger dashed circle in Fig. 1). The growth factor function for the prey is also determined by two factors: availability of space to reproduce and success at avoiding detection. For both species, reproduction is assumed to be asexual, and the mutations of each allele are uniformly distributed over a range around the current allele value.

The simulation steps are as follows:

- (1) Set the initial conditions; create a separate fundamental equation for each allele vector in each species; set $t = 0$.

- (2) Evaluate the fundamental equations to determine the expected number of births and deaths at time $t + 1$.
- (3) Randomly sample from probability distributions to get the actual number of births and deaths. Births are sampled from a Poisson probability distribution and deaths from a binomial probability distribution.
- (4) Randomly sample from actual births to get the number of mutations; for each mutation randomly sample to get a new allele vector; for each new allele vector add a fundamental equation.
- (5) Update counts of the number of organisms; for each allele vector with a count of zero, eliminate its fundamental equation.
- (6) Set $t = t + 1$; go to step (2).

Representative simulation results are shown in Fig. 5. Panels A–D show the results after the 1st, 100th, 1,000th and 20,000th (asymptotic) generation. Each allele vector is represented by a small square; the length of the line attached to the square indicates the number of organisms (on a log scale) carrying that allele vector. For the predator, the horizontal axis represents the peak spectral sensitivity of the variable receptor, and the vertical axis represents the value of the decision criterion. For the prey, the horizontal axis represents the peak wavelength of the surface characteristic; the tick mark shows the peak wavelength of the background (foliage) surface characteristic. Stars indicate the optimal allele vector for the predator according to the maximum-fitness ideal observer. In this simulation, we chose an initial state in which the population of the predator is relatively small and contains a diverse set of allele vectors (chosen at random), and the population of the prey is relatively large and contains a limited range of allele vectors. This initial state corresponds to a situation where a small number of the predators move into a new habitat where the prey is well-established. As can be seen, the allele populations of both species evolve over time. The alleles for the prey migrate to be as close to the background surface characteristic as possible, and the alleles for the predator follow along, eventually reaching the optimal values.⁹ Note that the optimal allele values defined by the maximum-fitness ideal observer change over time because of changes in the environmental prior probabilities.

The co-evolutionary trajectories sampled in Fig. 5 are represented in a more continuous fashion in Fig. 6A and B. Fig. 6A plots the mean value of the allele determining the prey's surface characteristic on the vertical axis and the mean value of the allele determining the peak spectral sensitivity of the predator's variable receptor on the horizontal axis. Fig. 6B plots the total number of prey on the vertical axis and the total number of predators on the horizontal axis. The four arrowheads represent corresponding points in time in the two figures. The black circles indicate the stable asymptotic endpoint of the co-evolution. In this case, the surface characteristic of the prey evolves roughly monotonically toward an optimal match with the background. There are two other stable patterns of co-evolution. Fig. 6C and D show a typical co-evolutionary trajectory if the initial population of alleles in the prey is located further up (i.e., with their surface color shifted more toward the red end of the spectrum) than in Fig. 6A. In this case, the surface characteristic of the prey initially shifts away from an optimal match with the background; given the relatively small mutation step size, shifts toward the background increase the probability of detection by the predator (i.e., decrease

Co-evolution

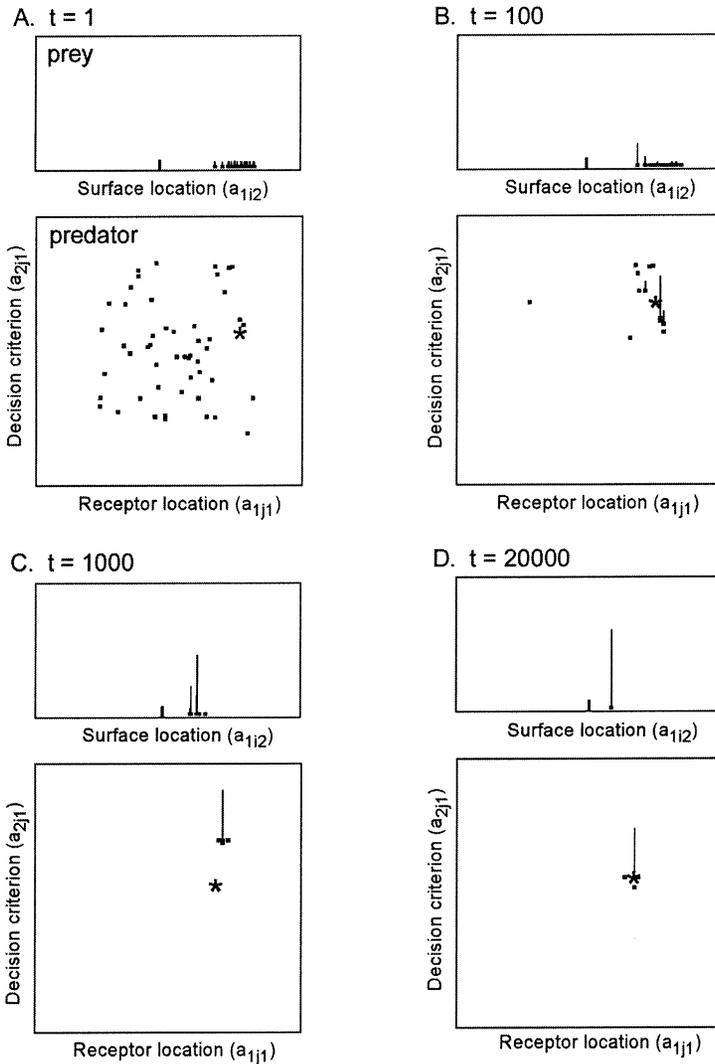


Fig. 5. Results of a simulation of co-evolution for the hypothetical species and environment shown in Fig. 1. Each allele vector is represented by a small square; the length of the line attached to the square indicates the number of organisms carrying that allele vector. Stars indicate the optimal allele vector according to the maximum-fitness ideal observer. (A) After the first step ($t = 1$). (B, C) Intermediate steps ($t = 100$ and $1,000$). (D) Asymptotic/equilibrium state ($t = 20,000$). (From Geisler & Diehl, 2002.)

the prey's fitness). However, once the predator's receptor location evolves to match the prey's surface characteristic, mutations in the prey toward the background produce increases in fitness. Thus, the prey's surface characteristic and the predator's receptor location eventually reach the same endpoints as in the first co-evolutionary simulation (cf. Fig. 6A and B). Interestingly, this does not occur if the range of possible receptor locations in the predator is substantially

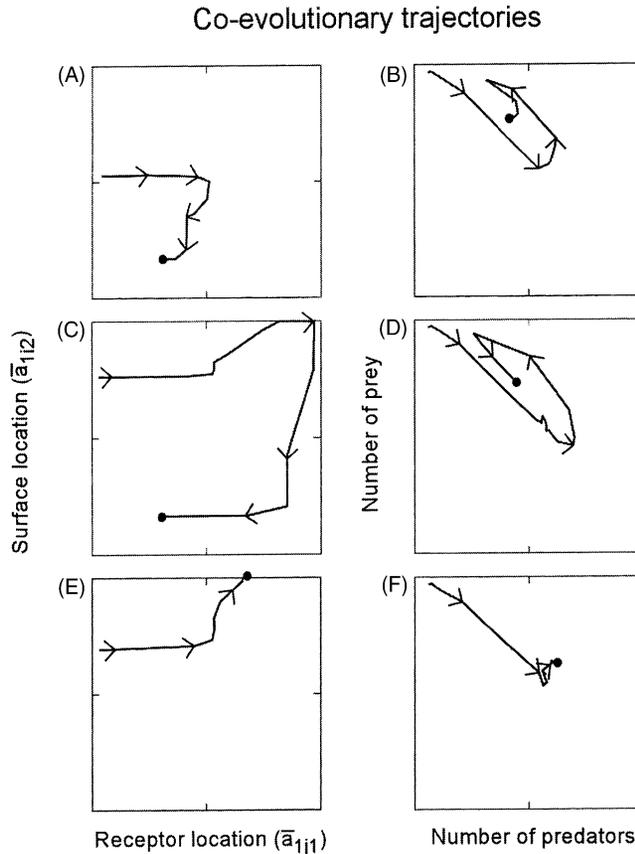


Fig. 6. Examples of the three stable trajectories observed in simulations of co-evolution for the hypothetical species and environment shown in Fig. 1. In the left panels, the horizontal axis gives the average location of the wavelength peak of the receptor in the predator species, and the vertical axis gives the average location of the wavelength peak in the surface characteristic of the prey species. In the right panels, the horizontal and vertical axes give the population sizes of the predator and prey species. (A, B) Prey species evolves toward an optimal match with the background. (C, D) Prey species initially evolves away from an optimal match with the background but later evolves toward it after the receptor of the predator species evolves to match the surface characteristic of the prey. (E, F) Prey species evolves away from an optimal match with the background and becomes trapped because of a limited range in the evolution of the predator's receptor location. (From Geisler & Diehl, 2002.)

less than the range of possible surface characteristics in the prey (see Fig. 6E and F). In this case, the prey becomes trapped with a surface characteristic far from an optimal match with the background, and its population size is permanently depressed relative to the other co-evolutionary trajectories. Given the generic assumptions used in these simulations, we would expect these three different patterns of co-evolution to arise frequently in nature, although evolutionary biologists have yet to investigate this possibility.

In addition to co-evolution, we also simulated situations where the predator evolved stable polymorphisms of allele vectors (Geisler & Diehl, 2002). In both cases, we found that the predator's visual system converged to the optimum defined by a maximum-fitness ideal

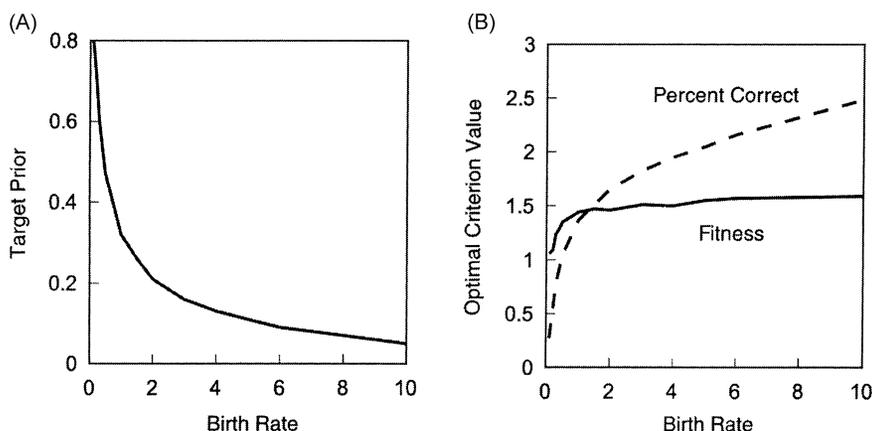


Fig. 7. Effect of variations in reproductive efficiency on the equilibrium values of the decision criterion in the predator species. (A) This plot shows the effect on the prior probability (at equilibrium) of a prey (a target) being within detection range of an individual predator, as a function of the nominal birth-rate of the predator. (B) If the goal is to maximize prey detection accuracy, then the optimal decision criterion should increase by more than an order of magnitude over the range of birth rates (dashed curve), because of the decline in the target prior probability. On the other hand, if the goal is to maximize fitness, then the decision criterion should be relatively invariant with birth rate (solid curve), because the decrease in target prior probability is balanced by the increase in payoff when a prey is captured. The solid curve is predicted both by the maximum-fitness ideal observer and by the simulation of natural selection. (From Geisler & Diehl, 2002.)

observer. The only exceptions involved initial conditions that led to extinction of one species or the other. (Note, however, that convergence to the optimum is unlikely to be common in more complex situations.)

Changes in maximum densities (minimum reproduction spaces) and reproductive efficiencies (nominal birth and death rates) of the predator and prey species lead to changes in the prior probability distributions at equilibrium. For example, increasing reproductive efficiency in the predator leads to more predation and hence to a reduction in the prior probability of being within detection range of a prey. Fig. 7A shows the decrease in the prior probability of being within range of a prey as function of the nominal birth rate of the predator (for a given caloric input and density of conspecifics). For a standard ideal observer, whose goal is to maximize accuracy, these decreases in prior probability should lead to a more conservative decision criterion in the predator (i.e., an increased bias toward avoiding rather than approaching; see the dashed curve in Fig. 7B). In contrast, the maximum-fitness ideal observer predicts a nearly constant decision criterion (see solid curve in Fig. 7B) across changes in reproductive efficiency because the decrease in prior probability is balanced by the increase in reproductive payoff when a prey is captured.

These simulations demonstrate the power and flexibility of the Bayesian framework. They show that it is possible to proceed from the abstract (and very general) formulas of Bayesian natural selection to concrete models of evolutionary processes. The development of such models is facilitated by the way in which the Bayesian formulas partition the evolutionary process into its natural components: environmental prior probabilities, stimulus likelihoods, response probabilities, utility function, and genetics. Each of these components can be evaluated

separately and then combined to characterize the process as a whole. The resulting models are often tractable, even for complex evolutionary scenarios, because (1) the Bayesian formulas apply to environmental and population distributions rather than to individuals (although they can also be applied at the level of individuals when advantageous), (2) the number of fundamental equations (Fig. 4) increase only linearly with the number of different allele vectors and species, and (3) extinction tends to limit the total number of active fundamental equations at any given time. The simulations also demonstrate how the Bayesian framework—because of its formal integration of natural scene statistics, environmental properties and genetics—leads readily to novel predictions that may be testable by biologists and psychologists.

It is useful to consider briefly the relationship between the Bayesian framework and other formal approaches to evolutionary biology and population genetics, as well as its relationship to computational applications of evolutionary concepts (e.g., artificial life). Bayesian natural selection can be regarded as a particular approach to population genetics (see, for example, Lewontin, 1974; Maynard Smith, 1989; Hartl & Clark, 1997). In population genetics the focus is on the statistics of genetic variation within and across species and how those statistics change over time. Bayesian natural selection also focuses on the statistics of genes, and it is consistent with basic formulas in population genetics (Geisler & Diehl, 2002); however, its unique contribution is to place a corresponding emphasis on the statistics of natural environments and to provide a formal representation of the interactions between the environment and genome.

Maximum-fitness ideal observer theory can be regarded as a special case of optimization theory, which has played an important role in evolutionary biology over the last three decades (see, for example, Maynard Smith & Price, 1973; Alexander, 1982; Parker & Maynard Smith, 1990). Like ideal observer theory, optimization theory has been used to calculate optimal solutions to adaptive problems, solutions that serve as a benchmark in evaluating actual biological adaptations. However, there are some significant differences. First, optimization theory has typically been expressed in the form of deterministic equations, whereas ideal observer theory emphasizes the statistical properties of the environment and of the systems being optimized. These statistical properties are crucial for determining the correct optimization in most perceptual and cognitive tasks. Second, ideal observer theory makes explicit the prior probabilities, thus forcing one to consider rigorously the environment within which optimization occurs. Third, the link being optimization theory and the theory of natural selection is not formally defined. An advantage of the Bayesian approach is that the same prior probability distributions, likelihood distributions, and fitness function are contained in the formulas for the ideal observer and for Bayesian natural selection.

The Bayesian framework outlined here is also related to computational work in artificial intelligence and artificial life. For example, in the domain of artificial intelligence, Bayesian networks provide powerful methods for determining ideal performance in various inference tasks (e.g., see Jensen, 2001). Broadly defined, artificial life is the computational study of evolution and adaptation in systems resembling living organisms (Holland, 1975; Langton, 1989), and thus, applications of Bayesian natural selection can be regarded as instances of artificial life. However, in general, existing models in artificial life are not based on the concepts of Bayesian statistical decision theory or on physical measurements of environmental statistics. Also, much work in artificial life is directed at exploring the space of possible evolutionary scenarios, not at formulating models of actual evolutionary processes.

2. Comparison with alternative approaches in psychology

The Bayesian framework outlined here is consistent with many important ideas expressed in other approaches to perception and cognition. Before considering these other approaches, we begin with the observation that the Bayesian framework integrates both nativist and empiricist traditions in psychology. Nativists emphasize the importance of fixed (i.e., hard-wired) adaptations, noting that many aspects of perceptual and cognitive processing are present as soon as they can be measured during ontogeny. Empiricists emphasize the importance of facultative adaptations (e.g., learning and plasticity), noting that many aspects of perceptual and cognitive processing are acquired through experience. The Bayesian framework recognizes the appropriateness of both fixed and facultative mechanisms and, in fact, suggests principled reasons to expect fixed mechanisms in some circumstances and facultative mechanisms in others. When the prior probability and likelihood distributions relevant for a given task tend to change substantially over the typical lifetime of a species, then there may be advantages, and hence an evolutionary pressure, for facultative mechanisms. Furthermore, learning and plasticity mechanisms may in some cases reduce the complexity of the genetic code required to implement a perceptual/cognitive function or representation. On the other hand, when the relevant prior probability distributions are stable, except for random fluctuations, then facultative adaptations may actually be harmful. For example, in many foraging situations there are short-term random variations in availability of a given food source even when average availability is constant over the lifespan of the organism. In such situations a facultative mechanism that responds to short-term fluctuations in availability may be less successful than a fixed mechanism calibrated to the long-term average availability. In effect, such a facultative mechanism is responding to environmental noise, which can only degrade performance. It should also be noted that some facultative adaptations are elaborations of earlier fixed adaptations, and hence the genetic specification of a facultative adaptation may be more complex than that of the earlier fixed adaptation (Williams, 1966). This discussion makes it clear that to understand the evolution of fixed and facultative mechanisms it will be important to characterize the temporal dynamics of environmental statistics.

2.1. Rational inference theory

The Bayesian framework in perception had its beginnings in Helmholtz's notion of "unconscious inference"—the idea that perceptual systems use implicit knowledge of the environment, and of the mapping between environment and stimulus, to infer, automatically and unconsciously, environmental properties from otherwise ambiguous stimuli impinging on sensory organs: "The general rule determining the ideas [percepts] of vision that are formed whenever an impression is made on the eye . . . is that *such objects are always imagined [perceived] as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism, the eyes being used under normal ordinary conditions* [his italics]." (Helmholtz, 1924, p. 2). As this quote shows, Helmholtz understood, at least informally, the role of prior probabilities and likelihoods in making perceptual inferences. The phenomenon of induced motion is one of the many examples Helmholtz used to illustrate the process of unconscious inference. He noted that if one accelerates slowly so that there are

no bodily cues to motion, and if the motion is visible only over a small portion of the visual field, then objects within that portion of the visual field may appear to be moving (e.g., a stationary train may appear to move when viewed from the window of a departing train on a parallel track). This would be a correct inference under “*normal ordinary conditions*” because self-motion, unlike object motion, is usually accompanied by bodily cues and by retinal motion over a large fraction of the visual field. Such stimulus factors, which were treated informally by Helmholtz, are rigorously specified in the Bayesian framework by prior probability and likelihood distributions (Freeman, 1994; Knill, Kersten, & Yuille, 1996; Yuille & Bülthoff, 1996; Brainard & Freeman, 1997; Geisler & Diehl, 2002; also see Fig. 2).

Helmholtz viewed unconscious inference as a result of learning by association. However, as discussed earlier, such inferential processing may be the result of facultative adaptations (learning mechanisms) and/or fixed adaptations (hard-wired perceptual mechanisms).

2.2. *Gestalt psychology*

The Gestalt psychologists demonstrated a number of important principles of perceptual grouping and organization and, in so doing, demonstrated the importance of configural and contextual effects in perception: “The whole is greater than the sum of its parts.” In the Bayesian framework, as outlined here, one expects a correspondence between the properties of the environment and the design of perceptual systems. It is obvious that environmental properties are highly regular, and perceptual systems should be designed to exploit those regularities. For example, recall that the reflectance spectra of natural materials and the illumination spectra of natural light sources can be described by a small number of parameters. These regularities can be exploited to solve approximately the problem of color constancy—identifying surface reflectance independent of the illuminant. Effective Bayesian solutions involve comparing the light reaching the eye from many locations within the scene (Brainard & Freeman, 1997), and if such solutions are implemented, they demonstrate complex configural and contextual effects.

Similarly, as described earlier, the geometrical relations between edge elements in natural scenes display the following statistical regularity: if edge elements are close to each other and if they are co-circular (tangent to the same circle), then they are more likely to belong to the same physical contour. A Bayesian ideal observer that directly exploits the prior probability and likelihood distributions for these geometrical relations displays the Gestalt principles of “good continuation” and “proximity,” and accurately predicts human contour detection performance (Geisler et al., 2001).

Thus, important perceptual principles, developed informally by the Gestalt psychologists, can be rigorously explained within the Bayesian framework.

2.3. *Ecological psychology*

Brunswik (1956) and Gibson (1966, 1979) were the first psychologists to focus on the important link between environmental regularities and perception, and therefore anticipated some of the central tenets of the Bayesian approach. The work of Brunswik is more closely related to the Bayesian approach because of its emphasis on measuring the probability distributions and correlations associated with environmental regularities. For example, Brunswik and Kamiya

(1953) measured the distances between parallel contours in visual images and showed that contours that are closer to each other are more likely to belong to the same physical object. Thus, Brunswik and Kamiya provided direct evidence for the functional validity of grouping on the basis of “proximity.” Their measurements are equivalent to estimating likelihood distributions in the Bayesian approach. However, Brunswik lacked a theoretical framework for using such likelihood distributions to understand either optimal perceptual design or real perceptual design arising from natural selection.

Gibson emphasized the richness of perceptual information available when an organism explores its natural environment. He recognized that this information is often sufficient for accurate perception of the three-dimensional layout of the environment over time, and thus he conceived the relationship between perceptual information and the environment as deterministic rather than probabilistic. In general, it is true that much of human perception is veridical under natural conditions. However, this is generally the result of combining many probabilistic sources of information (optic flow, shading, shadows, texture gradients, binocular disparity, and so on). Bayesian ideal observer theory specifies how, in principle, to combine the different sources of information in an optimal manner in order to achieve an effectively deterministic outcome.

We note, however, that some of the most important perceptual tasks in natural environments are prone to error, even when organisms are free to explore their environment. For example, predators (prey) are often confronted with a difficult detection task because their prey (predators) evolved effective camouflage. In these tasks, perception is almost guaranteed to be probabilistic, especially if natural selection has operated over a sufficient length of time to establish equilibrium between predator and prey.

2.4. *Signal detection theory*

Signal detection theory was the first explicit quantitative application of Bayesian statistical decision theory in perception and cognition. The theory specifies Bayesian ideal observers and optimal decision rules for simple detection and discrimination tasks (Peterson, Birdsall, & Fox, 1954; Tanner & Swets, 1954; Green & Swets, 1966). In the years since the introduction of signal detection theory, the Bayesian approach has been applied to a wide range of laboratory and natural tasks, and has been expanded to include anatomical and physiological constraints, natural scene statistics, and now natural selection. The Bayesian approach described here is a direct descendent of signal detection theory.

2.5. *Information theory*

The development of information theory (Shannon, 1948) inspired the hypothesis that perceptual systems encode stimuli in an optimally efficient manner (Attneave, 1954; Barlow, 1961). Attneave suggested that perceptual systems attempt to code sensory information efficiently (i.e., compactly), and in the process create representations consistent with the Gestalt notion of “Prägnanz,” the principle that proximal stimuli are interpreted in the least complex, most parsimonious fashion. For example, the retinal projection of a transparent cube (seen from an arbitrary viewpoint) can be described either as a heterogeneous list of vertices and angles in

the two-dimensional retinal image or as a homogeneous, and hence more efficiently codable, list of vertices and angles in three dimensions.

The information theoretic approach is generally compatible with the Bayesian approach. Coding is a task with well-defined goals, and hence optimal coding theory may be regarded as a particular form of Bayesian ideal observer theory. For example, if the task is to efficiently code sensory input, then (given appropriate prior probability and likelihood distributions, and a utility function that specifies the costs of the various types of coding errors) the Bayesian ideal observer will be the optimal coder. The close connection between Bayesian ideal observer theory and information theory has recently been demonstrated in the development of near-optimal error correcting codes (“Turbo codes”) based on Bayesian belief propagation networks (Berrou et al., 1993; Weiss & Freeman, 2001).

The usefulness of the information theoretic approach in perception and cognition depends on the extent to which the processing goal or sub-goal under consideration is efficient coding. This seems plausible for early processing where all information must be transmitted along sensory nerves of necessarily limited size. It is also possible that efficient coding is important for storage and retrieval, and for transmitting information from one brain area to another. On the other hand, the goal of many perceptual and cognitive tasks is to create representations of the environment that are accurate and that support rapid responses to events. As we have seen, Bayesian ideal observers in these cases take a different form from that of optimal coding. Undoubtedly, some aspects of perceptual and cognitive processing are concerned with efficient coding, while others are concerned with accurate representations that support rapid responses; either way, Bayesian ideal observer theory and Bayesian natural selection are appropriate analytic tools.

2.6. Computational approaches

The computational approach to perception and cognition is well represented by the work of Marr (1982). One of his most important contributions was to characterize information processing in terms of three levels of analysis: computational theory, algorithm, and implementation. A computational theory defines an information processing problem, the constraints that apply in solving the problem, and the formal solution to the problem given the constraints. Recall that Bayesian ideal observer theory specifies how to perform a task optimally given the available information and existing constraints. Thus, Bayesian ideal observer theory is an appropriate formalization of Marr’s intuitive notion of computational theory. Consistent with Marr’s characterization, the Bayesian ideal observer is specified abstractly, independent of a particular algorithm or implementation. For example, a Bayesian ideal observer might be implemented in a nervous system with a set of optimal decision rules that reflect only implicitly the prior probability distributions, likelihood distributions, and utility functions.

2.7. Evolutionary psychology

Evolutionary psychologists have demonstrated the value of applying neo-Darwinian theory to the study of human perception, cognition, motivation, affect, and social behavior (Cosmides & Tooby, 1992; Buss, 1994, 1999; Pinker, 1997). This body of work shows that evolutionary

factors play an important role in all aspects of psychology. Here, we have described the Bayesian framework largely with respect to problems in perception; however, Bayesian concepts are also being applied to problems in cognition, including recognition memory (Liu & Kersten, 1998), memory retrieval (Anderson, 1990; Shiffrin & Steyvers, 1997), reading (Legge, Klitz, & Tjan, 1997), search and attention (Palmer, Verghese, & Pavel, 2000; Eckstein, Shimozaki, & Abbey, 2002; Doshier & Lu, 2000), concept learning (Feldman, 2000), and spatial memory and way-finding (Stankiewicz, Legge, & Mansfield, 2000). More generally, the framework described here (Bayesian ideal observer theory and Bayesian natural selection) is not restricted to perceptual and cognitive systems; it applies in principle to all of the domains that have been considered by evolutionary psychologists.

In fact, the terms in the formulas of the Bayesian framework can represent any biological/psychological system that can be characterized by an input and an output. For example, the prey species in our simulations did not possess a perceptual system, but a surface characteristic whose input (stimulus) was an energy distribution from the environment and whose output (response) was another energy distribution emitted back into the environment. Thus, the stimulus and response in Bayesian natural selection could represent just about anything from the input and output of a transcription enzyme, to the input and output of a cell membrane, to the input and output of an organ, to the input and output of an organism.

3. Summary

- (1) The Bayesian approach that has been so successful in perceptual studies extends elegantly to the domain of evolution by natural selection.
- (2) Bayesian natural selection is a formalization of Darwinian natural selection that specifies explicitly how natural scene statistics and other properties of the environment interact with genes to determine the design of perceptual and cognitive systems.
- (3) The proposed Bayesian framework consists of two complementary components: maximum-fitness ideal observers and Bayesian natural selection.
- (4) A maximum-fitness ideal observer is a standard Bayesian ideal observer designed to apply to natural tasks and stimuli, with a utility function based on fitness (the birth and death rates associated with different stimulus–response outcomes).
- (5) An advantage of the Bayesian framework is that both maximum-fitness ideal observer theory and Bayesian natural selection are expressed in the same terms, facilitating comparisons between optimal and real design.
- (6) Fairly complicated simulations of Bayesian natural selection were found to be tractable. Simulations of evolutionary scenarios considerably more complex than those described here are also feasible.
- (7) Even the relatively simple simulations described here reveal some surprising predictions (concerning likely patterns of co-evolutionary trajectory, and tradeoffs between reproductive efficiency and placement of perceptual decision criteria), which should be testable in the field.
- (8) The Bayesian framework captures and generalizes, in a formal way, many of the most important theoretical insights of other approaches to perception and cognition, including

ideas of Helmholtz, Gestalt psychologists, Brunswik, Gibson, Green & Swets, Attneave, Barlow, Marr, and evolutionary psychologists.

- (9) Bayesian statistical decision theory provides a potentially unifying perspective not only on perceptual and cognitive systems, but on much of the rest of psychology and biology.

Notes

1. We emphasize that Bayesian natural selection is not an alternative to (neo-)Darwinian natural selection, but rather a particular formalization of it.
2. The term “allele” refers to a particular variant of a gene. The term “allele vector” refers to a particular set of alleles within the genome of an organism. In this paper, we assume that allele vectors correspond to those (few) genes that are most relevant to the biological or behavioral properties under consideration.
3. Note that deterministic factors can be regarded as special cases of probabilistic ones.
4. The likelihood function is the stimulus likelihood distribution, $p(s|\omega)$, regarded as a function of ω for the particular stimulus received by the observer.
5. Note that $\arg \max[f(x)]$ is the value of x (the argument) for which $f(x)$ reaches its maximum value.
6. The allele vector appears in the terms for the prior probability and stimulus likelihood to allow for cases where the environment and stimuli are internal to the organism.
7. Note that there is often flexibility in how environmental statistics are partitioned into prior probabilities and stimulus likelihoods. The choice does not affect the results of a Bayesian analysis, but can have pragmatic effects (e.g., on computational convenience).
8. Note that the units of t are arbitrary, and hence an increment of 1 could represent any time step Δt .
9. We did not allow the prey to perfectly match the background because it is not biologically plausible and because the simulation would degenerate (since the predators could never detect the prey except by chance).

Acknowledgments

We thank John Loehlin, Robert Sekuler, David Buss, Dennis McFadden, Mark Steyvers, and Ashlynn Kinney for helpful comments and suggestions. Supported by NIH grants EY11747 and EY02688 to WSG, and NIH grant DC00427 to RLD.

References

- Alexander, R. M. (1982). *Optima for animals*. London: Arnold.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Attias, H., & Schreiner, C. E. (1997). Temporal low-order statistics of natural sounds. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems* (Vol. 9, pp. 27–33). Cambridge, MA: MIT Press.

- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.
- Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.
- Barlow, H. B. (1977). Retinal and central factors in human vision limited by noise. In H. B. Barlow & P. Fatt (Eds.), *Vertebrate photoreception* (pp. 337–358). London: Academic Press.
- Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A*, *14*, 1393–1411.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.
- Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of ‘proximity’ and other Gestalt factors. *American Journal of Psychology*, *66*, 20–32.
- Burton, G. J., & Moorehead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*, 157–170.
- Buss, D. M. (1994). *The evolution of desire*. New York: Basic Books.
- Buss, D. M. (1999). *Evolutionary psychology*. Boston: Allyn & Bacon.
- Chiao, C., Vorobyev, M., Cronin, T. W., & Osorio, D. (2000). Spectral tuning of dichromats to natural scenes. *Vision Research*, *40*, 3257–3271.
- Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, *95*, 4002–4006.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind*. New York: Oxford University Press.
- Dixon, R. E. (1978). Spectral distribution of Australian daylight. *Journal of the Optical Society of America*, *68*, 437–450.
- Dosher, B. A., & Lu, Z. L. (2000). Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*, 1269–1292.
- Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual attention in the Posner paradigm revealed by classification images. *Journal of Vision*, *2*, 25–45.
- Elder, J. H., & Goldberg, R. M. (1998). Statistics of natural image contours. In *IEEE Computer Society Workshop on Perceptual Organization in Computer Vision*, Santa Barbara, CA.
- Feldman, J. (2000 October). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630–632.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*, 2379–2394.
- Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, *368*, 542–545.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267–314.
- Geisler, W. S., & Diehl, R. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London B*, *357*, 419–448.
- Geisler, W. S., & Kersten, D. (2002). Illusions. *Nature Neuroscience*, *5*, 6–8.
- Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hartl, D. L., & Clark, A. G. (1997). *Principles of population genetics* (3rd ed.). Sunderland, MA: Sinauer Associates.
- Helmholtz, H. V. (1924). *Treatise on physiological optics* (Vol. 3). Menasha, WI: Optical Society of America.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
- Huang, J., Lee, A. B., & Mumford, D. (2000). Statistics of range images. In *Proceedings of CVPR’00*.
- Jensen, F. V. (2001). *Bayesian networks and decision diagrams*. New York: Springer.
- Judd, D. B., MacAdam, D. L., & Wyszecki, G. W. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America*, *54*, 1031–1040.
- Knill, D. C., Kersten, D., & Yuille, A. (1996). A Bayesian formulation of visual perception. In D. C. Knill & R. W. (Eds.), *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.

- Krinov, E. (1947). *Spectral reflectance properties of natural formations* (Technical translation TT-439). Ottawa: National Research Council of Canada.
- Langton, C. G. (Ed.) (1989). *Artificial life: The proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems held September 1987 in Los Alamos, New Mexico*. Redwood City, CA: Addison-Wesley.
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review*, 104, 524–553.
- Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. New York: Columbia University Press.
- Liu, Z., & Kersten, D. (1998). 2D observers for human 3D object recognition? *Vision Research*, 38(15/16), 2507–2519.
- Lythgoe, J. N., & Partridge, J. C. (1989). Visual pigments and the acquisition of visual information. *Journal of Experimental Biology*, 146, 1–20.
- Maloney, L. T. (1986). Evaluation of linear models of surface reflectance with small numbers of parameters. *Journal of the Optical Society of America A*, 3, 1673–1683.
- Maloney, L. T., & Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3, 29–33.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman and Company.
- Maynard Smith, J. (1989). *Evolutionary genetics*. Oxford: Oxford University Press.
- Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, 246, 15–18.
- Osorio, D., & Vorobyev, M. (1996). Colour vision as an adaptation to frugivory in primates. *Proceedings of the Royal Society of London B*, 263, 593–599.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40, 1227–1268.
- Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, 348, 27–33.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the Institute of Radio Engineers, Professional Group on Information Theory*, 4, 171–212.
- Purves, D., & Lotto, R. B. (2003). *Why we see what we do*. Sunderland, MA: Sinauer.
- Regan, B. C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., & Mollon, J. D. (2001). Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London B*, 356, 229–283.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, 4, 145–166.
- Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle: Natural scenes and Gestalt rules. *Proceedings of the National Academy of Sciences*, 98, 1935–1940.
- Simoncelli, E. P., & Schwartz, O. (1999). Image statistics and cortical normalization models. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, pp. 153–159). Cambridge, MA: MIT Press.
- Stankiewicz, B. J., Legge, G. E., & Mansfield J. S. (2000). Optimal navigation strategies with varying visual input. *Investigative Ophthalmology & Visual Science Supplement*. (ARVO) 41/4.
- Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, 18, 1393–1399.
- Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61(6), 401–409.
- Weiss, Y., & Freeman, W. T. (2001). Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, 13, 2173–2200.
- Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.
- Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge, UK: Cambridge University Press.