Introduction

Each waking moment, our body's sensory receptors convey a vast amount of information about the surrounding environment to the brain. Visual information, for example, is measured by about 10 million cones and 100 million rods in each eye, while approximately 50 million receptors in the olfactory epithelium at the top of the nasal cavity signal olfactory information. How does the brain transform this raw sensory information into a form that is useful for goal-directed behavior? Neurophysiological, neuroanatomical, and brain imaging studies in the past few decades have helped to shed light on this question, revealing bits and pieces of the puzzle of how sensory information is represented and processed by neurons at various stages within the brain.

However, a fundamental question that is seldom addressed by these studies is *why* the brain chose to use the types of representations it does, and what ecological or evolutionary advantage these representations confer upon the animal. It is difficult to address such questions directly via animal experiments. A promising alternative is to investigate computational models based on efficient coding principles. Such models take into account the statistical properties of environmental signals, and attempt to explain the types of representations found in the brain in terms of a *probabilistic model* of these signals. Recently, these models have been shown to be capable of accounting for the response properties of neurons at early stages of the visual and auditory pathway, providing for the first time a unifying view of sensory coding across different modalities. There is now growing optimism that probabilistic models can also be applied successfully to account for the sensory coding strategies employed in yet other modalities, and eventually to planning and executing goal-directed actions.

This book surveys some of the current themes, ideas, and techniques dominating the probabilistic approach to modeling and understanding brain function. The sixteen chapters that comprise the book demonstrate how ideas from probability and statistics can be used to interpret a variety of phenomena, ranging from psychophysics to neurophysiology. While most of the examples presented in the chapters focus on vision, this is not meant to imply that these models are applicable only to this modality. Many of the models and techniques presented in these chapters are quite general, and therefore are applicable to other modalities as well.

The probabilistic approach

The probabilistic approach to perception and brain function has its roots in the advent of information theory, which inspired many psychologists during the 1950's to attempt to quantify human perceptual and cognitive abilities using statistical techniques. One of these was Attneave, who attempted to point out the link between the redundancy inherent in images and certain aspects of visual perception [2]. Barlow then took this notion a step further, by proposing a self-organizing strategy for sensory nervous systems based on the principle of *redundancy reduction* [3, 4]—i.e., the idea that neurons should encode information in such a way as to minimize statistical dependencies. The alluring aspect of this approach is that it does not require that one pre-suppose a specific goal for sensory processing, such as "edge-detection" or "contour extraction." Rather, the emphasis is on formulating a *general* goal for sensory processing from which specific coding strategies such as edge detection or contour integration could be derived.

Despite the elegance of Attneave's and Barlow's proposals, their ideas would not be put seriously to work until much later. [1] Most modeling work in sensory physiology and psychophysics over the past 40 years has instead been dominated by the practice of attributing *specific* coding strategies to certain neurons in the brain. This approach is probably best exemplified by Marr and Hildreth's classic theory of edge-detection [11], or the plethora of Gabor-filter based models of visual cortical neurons that followed [10, 6, 8]. It is also prevalent in the realms of intermediate and high level vision, for example in schemes such as codons [14], geons [5], and the medial axis transform [13] for representing object shape. In contrast to the probabilistic approach, the goal from the outset in such models is to formulate a specific computational strategy for extracting a set of desired properties from images. Nowhere is there any form of learning or adaptation to the properties of images. Instead, these models draw upon informal observations of image structure and they rely heavily upon mathematical elegance and sophistication to achieve their goal.

Interest in the probabilistic approach was revived in the 1980s, when Simon Laughlin and M.V. Srinivasan began measuring the forms of redundancy present in the natural visual environment and used this knowledge to make quantitative predictions about the response properties of neurons in early stages of the visual system [9, 15]. This was followed several years later by the work of Field [7], showing that natural images exhibit a characteristic $1/f^2$ power spectrum, and that cortical neurons are well-adapted for representing natural images in terms of a *sparse code* (where a small number of neurons out of the population are active at any given time). Then drawing upon information theory, as well as considerations of noise and the $1/f^2$ power spectrum, Atick [1] and van Hateren [16] formulated efficient coding theories for the retina in terms of whitening of the power spectrum (hence removing correla-

1. There were some early attempts at implementing these principles in self-organizing networks (e.g. [12]), but these fell short of being serious neurobiological models.

tions from signals sent down the optic nerve) in space and time. This body of work, accumulated throughout the 1980's and early 1990's, began to build a convincing case that probabilistic models could contribute to our understanding of sensory coding strategies. Part II of this book contains eight recent contributions to this area of inquiry.

The probabilistic approach has also been applied beyond the realm of sensory coding, to problems of perception. In fact, the idea that perception is fundamentally a problem of inference goes back at least to Hermann von Helmholtz in the nineteenth century. The main problem of perception is to deduce from the patterns of sensory stimuli the properties of the external environment. What makes this problem especially difficult is that there is ambiguity at every stage, resulting from lack of information, inherent noise, and the multitude of perceptual interpretations that are consistent with the available sensory data. Even something as simple as the interpretation of an edge can be complicated: Is it due to a reflectance change on the object? Is it a shadow that arises from the object's 3-dimensional structure? Or does it represent an object boundary? Determining which interpretation is most likely depends on integrating information from the surrounding context and from higher level knowledge about typical scene structure.

The process of inference is perhaps most compellingly demonstrated by the famous Dalmatian dog scene (reproduced in Figure 7.1), in which the luminance edges provide little or no explicit information about the object boundaries. Like a perceptual puzzle, each part of the image provides clues to the best interpretation of the whole. The question is how to combine these different sources of information in the face of a considerable degree of uncertainty. One framework for addressing these problems in the context of perceptual processing is that of Bayesian Inference (Szeliski, 1989; Knill and Richards, 1996).

What makes Bayesian inference attractive for modeling perception is that it provides a general framework for quantifying uncertainty and precisely relating what one set of information tells us about another. In Bayesian probability theory, uncertainty is represented by probability distribution functions, and Bayes' rule specifies the relation between the distributions (and therefore the uncertainties) and the observed data. A discrete distribution might represent uncertainty among a set of distinct possible interpretations, such as the probability of a word given a sound. A continuous distribution represents uncertainty of an analog quantity, such as the direction of motion given a time-varying image. By quantitatively characterizing these distributions for a given perceptual task, it is then possible to make testable predictions about human behavior. As we shall see in the chapters of Part I of this book, there is now substantial evidence showing that humans are good Bayesian observers.

Contributions of this book

The chapters in this book fall naturally into two categories, based on the type of approach taken to understand brain function. The first approach is to formulate proba-

bilistic theories with a predominantly *top-down* point of view—i.e., with an emphasis on computational algorithms rather than the details of the underlying neural machinery. The goal here is to explain certain perceptual phenomena or analyze computational tractability or performance. This has been the predominant approach in the psychology, cognitive science, and artificial intelligence communities. Part I of the book, entitled *Perception*, comprises eight chapters that embody the top-down approach to constructing probabilistic theories of the brain.

The second approach is the formulate theories of brain function that are motivated by understanding neural substrates and mechanisms. The goal of such theories is twofold: (a) to show how the distinctive properties of neurons and their specific anatomical connections can implement concrete statistical principles such as Bayesian inference, and (b) to show how such models can solve interesting problems such as feature and motion detection. Part II of this book, entitled *Neural Function*, presents eight such models.

The first three chapters of Part I present an introduction to modeling visual perception using the Bayesian framework. Chapter 1 by Mamassian, Landy, and Maloney serves as an excellent tutorial on Bayesian inference. They review the three basic components of any Bayesian model: the likelihood function, the prior, and the gain function. Likelihood functions are used to model how visual sensors encode sensory information, while priors provide a principled way of formulating constraints on possible scenes to allow unambiguous visual perception. Gain functions are used to account for task-dependent performance. Mamassian, Landy, and Maloney illustrate how Bayesian models can be investigated experimentally, drawing upon a psychophysical task in which the observer is asked to judge 3D surface structure. They show how the assumptions and biases used by the observer in inferring 3D structure from images may be modeled in terms of priors. More importantly, their work provides a compelling demonstration of the utility of the Bayesian approach in designing and interpreting the results of psychophysical experiments.

This approach is carried further in Chapter 2 by Schrater and Kersten, who explore the Bayesian approach as a framework within which to develop and test predictive quantitative theories of human visual behavior. Within this framework, they distinguish between mechanistic and functional levels in the modeling of human vision. At the mechanistic level, traditional signal detection theory provides a tool for inferring the properties of neural mechanisms from psychophysical data. At the functional level, signal detection theory is essentially extended to pattern inference theory, where the emphasis is on natural tasks and generative models for images and scene structure. Drawing upon examples in the domain of motion processing and color constancy, Schrater and Kersten show how ideal observers can be used to test theories at both mechanistic and functional levels.

Jacobs then uses the Bayesian approach in Chapter 3 to explore the question of how observers integrate various visual cues for depth perception. Again, the emphasis is on evaluating whether or not observers' cue integration strategies can be characterized as "optimal" in terms of Bayesian inference, in this case by using an

ideal observer as the standard of comparison. Jacobs shows that subjects integrate the depth information provided by texture and motion cues in line with those of the ideal observer, which utilizes the cues in a statistically optimal manner. He also shows that these cue integration strategies for visual depth are adaptable in an experience-dependent manner. Namely, subjects adapt their weightings of depth-from-texture and depth-from-motion information as a function of the reliability of these cues during training.

The next two chapters of Part I focus on the computation of motion within a probabilistic framework. Chapter 4, by Weiss and Fleet, describes a Bayesian approach to estimating 2D image motion. The goal is to compute the posterior probability distribution of velocity, which is proportional to the product of a likelihood function and a prior. Weiss and Fleet point out that there has been substantial confusion in the literature regarding the proper form of the likelihood function, and they show how this may be resolved by properly deriving a likelihood function starting from a generative model. The generative model assumes that the scene translates, conserving image brightness, while the image is equal to the projected scene plus noise. They then show that the likelihood function can be calculated by a population of units whose response properties are similar to the "motion energy" units typically used as models of cortical neurons. This suggests that a population of velocity tuned cells in visual cortex may act to represent the likelihood of a velocity for a local image sequence.

Chapter 5 by Freeman, Haddon, and Pasztor describes a learning-based algorithm for finding the scene interpretation that best explains a given set of image data. To illustrate their approach, they focus on the optical flow problem, where the goal is to infer the projected velocities (scene) which best explain two consecutive image frames (image). Freeman, Haddon, and Pasztor use synthetic scenes to generate examples of pairs of images, together with their correct scene interpretation. From these data, candidate scene explanations are learned for local image regions and a compatibility function is derived between neighboring scene regions. Given new image data, probabilities are propagated in a Markov network to infer the underlying optical flow in an efficient manner. They first present the results of this method for a toy world of irregularly shaped blobs, and then extend the technique to the case of more realistic images.

The final three chapters of Part I address questions at the level of computational theory. In Chapter 6, Nadal reviews recent results on neural coding and parameter estimation based on information theoretic criteria, exploring their connection to the Bayesian framework. He analyzes the amount of information conveyed by a network as a function of the number of coding units, and shows that the mutual information between input stimuli and network output grows at best linearly for a small number of units, whereas for a large number of units the growth is typically logarithmic. He also proposes some future directions to extend information theoretic approaches (in particular, the *infomax* approach) to the analysis of efficient sensory-motor coding.

In Chapter 7, Yuille and Coughlan examine the fundamental limits of vision from

an information theoretic point of view. Their goal is to clarify when high-level knowledge is required to perform visual tasks such as object detection, and to suggest tradeoffs between bottom-up and top-down theories of vision. This is especially helpful in clarifying the need for cortical feedback loops in perception. As a specific application of their theory, Yuille and Coughlan analyze the problem of target detection in a cluttered background and identify regimes where an accurate high-level model of the target is required to detect it.

Chapter 8 by Papageorgiou, Girosi, and Poggio presents a new paradigm for signal analysis that is based on selecting a small set of bases from a large dictionary of class-specific basis functions. The basis functions are the correlation functions of the class of signals being analyzed. To choose the appropriate features from this large dictionary, Papageorgiou, Girosi, and Poggio use Support Vector Machine (SVM) regression and compare this to traditional Principal Component Analysis (PCA) for the tasks of signal reconstruction, superresolution, and compression on a set of test images. They also show how multiscale basis functions and basis pursuit de-noising can be used to obtain a sparse, multiscale approximation of a signal. Interestingly, one of their results shows that the L_ϵ norm, which measures the absolute value of error past a certain threshold, may be a more appropriate error metric for image reconstruction and compression than the L_2 norm in terms of matching humans' psychophysical error function.

Part II of the book makes the transition from top-down to bottom-up probabilistic approaches, with an emphasis on neural modeling. The first three chapters of Part II discuss models of orientation selectivity. Chapter 9, by Piepenbrock, presents a model that explains the development of cortical simple cell receptive fields and orientation maps based on Hebbian learning. What makes this model stand apart from most models of orientation maps is that the learning is driven by the viewing of natural scenes. The underlying assumption is that visual experience and the statistical properties of natural images are essential for the formation of correct cortical feature detectors. In the model, interactions between cortical simple cells lead to divisive inhibition—a nonlinear network effect that may be interpreted as a cortical competition for input stimuli. The degree of competition is controlled by a model parameter that critically influences the development outcome. Very weak competition leads to large global receptive fields, whereas only strong competition yields localized orientation selective receptive fields that respond to the edges present in the natural images. The model implies that the early visual pathways serve to recode visual stimuli step by step in more efficient and less redundant ways by learning which features typically occur in natural scenes.

In Chapter 10, Wainwright, Schwartz, and Simoncelli present a statistical model to account for the nonlinear and adaptive responses of visual cortical neurons based on the idea of divisive normalization. The basic idea behind divisive normalization is that the linear response of a filter, characterizing the localized receptive field of a neuron, is rectified (and typically squared) and then divided by a weighted sum of the rectified responses of neighboring neurons. Wainwright, Schwartz, and Simoncelli

show that natural image statistics, in conjunction with Barlow's redundancy reduction hypothesis, lead to divisive normalization as the appropriate nonlinearity for removing statistical dependencies between the responses of visual cortical neurons to natural images. The model can also account for responses to non-optimal stimuli, and in addition, adjusting model parameters according to the statistics of recent visual stimuli is shown to account for physiologically observed adaptation effects.

In Chapter 11, Zemel and Pillow build upon previous work showing how neural tuning curves (such as those of orientation selective neurons) may be interpreted in a probabilistic framework. The main idea of this approach is to interpret the population response as representing the probability distribution over some underlying stimulus dimension, such as orientation. Here, they show how the population response can be generated using a combination of feedforward and feedback connections. In addition, they include the preservation of information explicitly in the objective function for learning the synaptic weights in the model. Zemel and Pillow show that their model can support a variety of cortical response profiles, encode multiple stimulus values (unlike some previous models), and replicate cross-orientation effects seen in visual cortical neurons in response to stimuli containing multiple orientations. They also describe several testable predictions of their model involving the effect of noise on stimuli and the presence of multiple orientations.

The next four chapters of Part II focus on how probabilistic network models can also incorporate the actual spiking properties of cortical neurons. Chapter 12 by Lewicki shows how a spike-like population code may be used for representing time-varying signals. The model is derived from two desiderata: 1) coding efficiency and 2) a representation that does not depend on phase shifts. The second goal is important, because it avoids the traditional approach of blocking the data which often produces inefficient descriptions. Encoding is accomplished by maximizing the likelihood of the time-varying signals over a distributed population of relative event times. The resulting representation resembles coding in the cochlear nerve and can be implemented using biologically plausible mechanisms. The model is also shown to be equivalent to a very sparse and highly overcomplete basis. Under this model, the mapping from the data to the representation is nonlinear but can be computed efficiently. This form also allows the use of existing methods for adapting the kernel functions.

Chapter 13 by Olshausen focuses on the relation between sparse coding and neural spike trains. He proposes that neurons in V1 are attempting to model time-varying natural images as a superposition of sparse, independent events in both space and time. The events are characterized using an overcomplete set of spatiotemporal basis functions which are assumed to be translation-invariant in the time domain (as in Lewicki's model). When adapted to natural movies, the basis functions of the model converge to a set of spatially localized, oriented, bandpass functions that translate over time, similar to the space-time receptive fields of V1 neurons. The outputs of the model are computed by sparsifying activity across both space and time, producing a non-linear code having a spike-like character—i.e.,, continuous, time-varying images

are represented as a series of sharp, punctate events in time, similar to neural spike trains. Olshausen suggests that both the receptive field structure and the spiking nature of V1 neurons may be accounted for in terms of a single principle of efficient coding, and he discusses a number of possibilities for testing this hypothesis.

In Chapter 14, Ballard, Zhang, and Rao present a model for spike-based communication between neurons based on the idea of synchronized spike volleys. In this model, neurons are used in a "time-sharing" model, each conveying large numbers of spikes that are part of different volleys. The phase between a neuron's successive spikes is unimportant but groups of neurons convey information in terms of precisely-timed volleys of spikes. Ballard, Zhang, and Rao show how such a model of spike-based neural communication can be used for predictive coding in the cortico-thalamic feedback loop between primary visual cortex and the lateral geniculate nucleus (LGN).

Chapter 15 by Hinton and Brown explores the hypothesis that an active spiking neuron represents a probability distribution over possible events in the world. In this approach, when several neurons are active simultaneously, the distributions they individually represent are combined by multiplying together the individual distributions. They also describe a learning algorithm that is a natural consequence of using a product semantics for population codes. Hinton and Brown illustrate their approach using a simulation in which a non-linear network with one layer of spiking hidden neurons learns to model an image sequence by fitting a dynamic generative model.

The final chapter (16), by Rao and Sejnowski, deals with the issues of cortical feedback and predictive coding. They review models which postulate that (a) feedback connections between cortical areas instantiate statistical generative models of cortical inputs, and (b) recurrent feedback connections within a cortical area encode the temporal dynamics associated with the generative model. The resulting network allows predicting coding of spatiotemporal inputs and suggests functional interpretations of nonclassical surround effects in the visual cortex on the basis of natural image statistics. Rao and Sejnowski show that recent results on spike timing dependent plasticity in recurrent cortical synapses are consistent with such a model of cortical feedback and present comparisons of data from model simulations to electrophysiological data from awake monkey visual cortex.

Open questions

The enterprise of developing probabilistic models to explain behavioral and neural response properties exhibited by the brain is still very much in its infancy. Research in the last five years, of which this book represents only a part, has so far provided just a glimpse of the potential of the probabilistic approach in understanding brain function. The strength of the probabilistic approach lies in its generality—for example, it is rich enough to provide a general global framework for vision, from low-level

feature-detection and sensory coding (Part II of the book) to higher-level object recognition, object detection, and decision making (Part I of the book).

We leave the reader with a list of important questions and issues that motivated some of the early research in the field and that continue to provide the impetus for much of the recent research, including the research described in this book:

- How does the brain represent probabilities? How does it combine these probabilities to perform inference?
- How broadly can the principle of statistically efficient coding of natural signals be applied before some other principle, such as task-related adaptation, becomes more important?
- How good is the characterization of perception as Bayesian inference? What are the alternatives?
- What role do feedback pathways play in inference?
- Should neurons be modeled as deterministic or stochastic elements? Are mean-field models or stochastic sampling models more appropriate?
- How does the brain deal with "noise?" How many bits of precision can a spike encode? More generally, what exactly does a spike mean?
- Are neurons integrators or coincidence detectors? How important is input synchrony? What role do cortical oscillations have in probabilistic theories of the brain?
- What do probabilistic theories predict about the function of lateral inhibitory connections, recurrent excitatory connections, cortico-cortical feedback connections, and subcortical connections in the brain?

These are, no doubt, daunting questions. However, as probabilistic methods for inference and learning become more sophisticated and powerful, and as discoveries in neuroscience and psychology begin to paint an increasingly detailed picture of the mechanisms of perception, action, and learning, we believe it will become possible to obtain answers to these intriguing questions.

Rajesh P. N. Rao
University of Washington, Seattle

Bruno A. Olshausen
University of California, Davis

Michael S. Lewicki
Carnegie Mellon University, Pittsburgh

References

[1] Atick, J. J. Could information theory provide an ecological theory of sensory processing. *Network*, 3:213–251, 1992.

[2] Attneave, F. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193, 1954.

[3] Barlow, H. B. Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, pages 217–234. Cambridge, MA: MIT Press, 1961.

[4] Barlow, H. B. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.

[5] Biederman, I. Recognition by components: A theory of human image understanding. *Psychological Review*, 94:115–145, 1987.

[6] Daugman, J. G. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856, 1980.

[7] Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4:2379–2394, 1987.

[8] Heeger, D. Optic flow using spatiotemporal filters. *International Journal of Computer Vision*, 1(4):279–302, 1987.

[9] Laughlin, S. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.*, 36:910–912, 1981.

[10] Marcelja, S. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70:1297–1300, 1980.

[11] Marr, D., Hildreth, E. Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 207:187–217, 1980.

[12] Goodall, M. C. Performance of a stochastic net. *Nature*, 185:557–558, 1960.

[13] Pizer, S. M., Burbeck, C. A., Coggins, J. M., Fritsch, D. S., Morse, B. S. Object shape before boundary shape: scale-space medial axes. *Journal of Mathematical Imaging and Vision*, 4:303–313, 1994.

[14] Richards, W., Hoffman, D. D. Codon constraints on closed 2D shapes. *Computer Vision, Graphics, and Image Processing*, 31:265–281, 1995.

[15] Srinivasan, M. V., Laughlin, S. B., Dubs, A. Predictive coding: A fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B*, 216:427–459, 1982.

[16] van Hateren, J. H. Spatiotemporal contrast sensitivity of early vision. *Vision Research*, 33:257–267, 1993.