Review of Bayeian Brain, book edited by Doya, Pouget and Rao. MIT Press, 2007.
Review by Charles Fox, Adaptive Behavior Research Group, University of Sheffield, UK.

Computational Neuroscience and AI have diverged greatly over the past decade: what was once a coherent field which thrived on trading biological and engineering models has fragmented into independent camps, one seeking to model brains and the other building useful artifacts. Bayesian Brain gives an overview of the recent attempts by a core group of researchers to reunite these fields and bring back the cross-discipliniarity that once characterized them.

The book portrays two key ideas: first, that probabilistic Data Analysis methods are useful in analyzing neural data; and second – more speculatively -- that Bayesian computational methods are useful models of the functions of the brain itself. These goals are distinct but share similar, cross-fertilizable concepts. However it could be argued that the connection is somewhat tenuous – and the two parts are almost completely independent books.

After Doya and Ishii's basic probability and information theory tutorial, Fairhall gives an introduction to spike coding in the single neuron. The key questions in this field ask what is represented, and what granularity of coding is used. The differences in information about stimuli contained in different granularities and time-windows can be quantified, however a difficulty inherent in the brute-force, model-free, information-theoretic approach is the exponential-sized state space and data collection requirements to populate it. The discussion thus shifts to parametric models, introducing the the Spike-Triggered Average, cascade and Integrate-and-Fire models. Efficient parameter fitting in a related approach, the linear-nonlinear Poisson model, is detailed at length in a chapter by Pillow. Another alternative is to model the distribution of single spike times: Richmond and Weiner discuss the use of order statistics, which use combinatorics to give the temporal distribution of the kth spike out of n spikes. Bayes' theorem can then be used to invert this model and make inferences about the stimulus.

Moving suddenly from spike-level models to whole-brain interactions, Penny discusses the use of Dynamic Causal Modeling for temporal fMRI activity in macroscopic brain regions. DCM may be viewed as a generalization of linear Gaussian Bayesian Networks which drops their acyclicity constraint and so models the time sequence of the network state. With network topology constrained by anatomical data, the model weights can be learned by standard Expectation Maximization (EM) iteration.

The second half of the book examines probabilistic structures as models of brain function itself. Pouget and Zemel introduce various schemes for representing probability distributions using populations of neurons, as found in orientation-selective visual hypercolumns, place cells, and others. In contrast to the classical view of the maximum response simply coding the best solution, they consider the whole collection of responses as coding for the uncertainty during inference. Latham and Pouget give a heuristic, bottom-up architecture for computing in the style Bayesian network nodes, using population coding; Rao provides a principled neural implementation of the Pearl message-passing inference algorithm.

Undirected networks – often called Markov Random Fields – have been used in the computer vision community for preconceptual image segmentation. Lee and Yuille review an annealing and EM style method for simultaneous approximate inference of region boundaries and region content in the MRF framework. They then advance it has a theory of V1 activity, presenting biological evidence to subsume mere orientation-selectivity into region and boundary responses.

Knill considers Bayesian inference as a normative theory of behavior, and uses visual psychophysics results to support this claim. In general, Bayesian inference is NP-hard so approximations must be used: the study of what deviations from the Bayesian ideal are made by the brain could shed light on what approximation methods are being used within this paradigm.

The remaining chapters shift from perception to action. Using the Martingale approach, Shadlen et al. review the sequential sampling problem: how much time and resources to expend on gathering evidence before making a decision. They go on to give evidence that this HMM-like process could be implemented in brain areas MT and LIP, representing current observations and accumulated beliefs respectively. (No mention is given to similar theories of Basal Ganglia function.) Tororov provides a highly mathematical tutorial on Kalman filters and control theory in the continuous-variable framework – an alternate view of Markov Decision Processes than the discrete approach often taken in Reinforcement Learning. These concepts are used in psychophysical experiments discussed by Kording and Wolpert, who like Knill, test human behavior against Bayesian optimal decisions. They also make some neurobiological speculations, including complementary roles for neurotransmitters ACh and NE in representing the almost Rumsfeldian expected uncertainty and unexpected uncertainty respectively.

Some notable omissions from the discussions: First, while the question of rates versus spike time representations appears often, there is no attempt to link it to what Machine Learning has to say about similar models. Variational and Gibbs Sampler methods are directly analogous to rates and spikes, and their pros and cons as approximate inference methods are well-known and could perhaps shed light on the biological debate. Second, the Bayesian paradigm is currently fashionable across many fields, and the book works firmly within it – but perhaps a discussion of how it relates to rival neural paradigms would be useful for the not-yet-converted? For example, many iterative Bayesian algorithms such as Variational Bayes and loopy Belief Propagation can be viewed as computing fixed points, in a similar way to the Dynamical Systems view of the brain. A final oddity is the amount of non-Bayesian statistics used throughout the book: estimators, Cramer-Rao bounds and General Linear Models all make appearances: while a comprehensive Bayesian tutorial is given in the first chapter, many classical concepts are simply assumed.

A common theme in the book is that optimal, exact Bayesian inference is computationally intractable, so both computers and brains must use heuristics to approximate it. The study of deviations from the ideal can give clues about what heuristics are used, both at the neural and psychophysical levels. However there is a danger of non-falsifiable science here: if the brain performs optimally then it is said to be a "Bayesian Brain"; but every deviation from this can be explained away as a heuristic. So how could we ever show that the brain is /not/ Bayesian?

Overall this is a much-needed and intriguing book, giving a birds-eye view of the whole research programme of rejoining Computational Neuroscience with the Machine Learning school of AI. The book showcases research from the many facets of this project, from information-theoretic spike models to neural implementations of belief propagation and Bayesian-normative psychophysics. There is a wealth of references and the book will reward many re-readings, both for overviews of the topics and as initial pointers into the literature, especially across disciplines. Many theses could be inspired by chapters of this book.