

Preface

Computational approaches to cognitive neuroscience (computational cognitive neuroscience) focus on understanding how the brain embodies the mind, using biologically based computational models made up of networks of neuronlike units. Because this endeavor lies at the intersection of a number of different disciplines, including neuroscience, computation, and cognitive psychology, the boundaries of computational cognitive neuroscience are difficult to delineate, making it a subject that can be difficult to teach well. This book is intended to support the teaching of this subject by providing a coherent, principled introduction to the main ideas in the field. It is suitable for an advanced undergraduate or graduate course in one semester or quarter (the authors have each used this text for such courses at their respective universities), and also for researchers in related areas who want to learn more about this approach to understanding the relation between mind and brain.

Any introductory text on the subject of computational cognitive neuroscience faces a potentially overwhelming set of compromises — one could write volumes on each of the different component aspects of computation, cognition, and neuroscience. Many existing texts have avoided these compromises by focusing on specific issues such as the firing patterns of individual neurons (e.g., Reike, Warland, van Steveninck, & Bialek, 1996), mathematically oriented treatments of the computational properties of networks (e.g., Hertz, Krogh, & Palmer, 1991), or more abstract models of cognitive phenomena (e.g., Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Plunkett & Elman, 1997). However, we knew that our excitement in the field was based in large part on the wide scope of the issues involved in this endeavor — from biological and computational properties to cognitive function — which

requires a broader perspective (hand in hand with a greater compromise on some of the details) than is captured in these texts.

Thus, like many of our colleagues teaching similar courses, we continued to think that the original PDP (parallel distributed processing) volumes (Rumelhart, McClelland, & PDP Research Group, 1986c; McClelland, Rumelhart, & PDP Research Group, 1986; McClelland & Rumelhart, 1988) were the best texts for covering the broader scope of issues. Unlike many later works, these volumes present the computational and biological mechanisms from a distinctly cognitive perspective, and they make a serious attempt at modeling a range of cognitive phenomena. However, the PDP volumes are now somewhat dated and present an often confusing hodge-podge of different algorithms and ideas. Also, the simulation exercises were a separate volume, rather than being incorporated into the text to play an integral role in students' understanding of the complex behavior of the models. Finally, the neuroscience got short shrift in this treatment, because most of the models relied on very abstract and somewhat biologically implausible mechanisms.

Our objective in writing this text was therefore to replicate the scope (and excitement) of the original PDP volumes in a more modern, integrated, and unified manner that more tightly related biology and cognition and provided intuitive graphical simulations at every step along the way. We achieved this scope by focusing on a consistent set of principles that form bridges between computation, neuroscience, and cognition. Within this coherent framework, we cover a breadth and depth of simulations of cognitive phenomena unlike any other textbook that we know of. We provide a large number of modern, state-of-the-art, research-grade simulation models that readers explore in some detail as guided by

the text, and that they can then explore further on their own.

We are well aware that there is a basic tradeoff between consistency and diversity (e.g., exploitation versus exploration as emphasized in the reinforcement-learning paradigm). The field of computational cognitive neuroscience has generally been characterized more by the diversity of theoretical approaches and models than by any kind of explicit consistency. This diversity has been cataloged in places like the encyclopedic Arbib (1995) volume, where readers can find overview treatments of widely varying perspectives. We view this book as a complementary resource to such encyclopedic treatments. We focus on consistency, striving to abstract and present as much as possible a consensus view, guided by a basic set of well-developed principles, with brief pointers to major alternative perspectives.

In summary, this book is an attempt to consolidate and integrate advances across a range of fields and phenomena into one coherent package, which can be digested relatively easily by the reader. At one level, the result can be viewed as just that — an integration and consolidation of existing knowledge. However, we have found that the process of putting all of these ideas together into one package has led to an emergent phenomenon in which the whole is greater than the sum of its parts. We come away with a sense of renewed excitement and interest in computational cognitive neuroscience after writing this book, and hope that you feel some of this, too.