Preface

Have computational models really advanced our understanding of the neural bases of learning and memory? If so, is it possible to learn about them without delving into the mathematical details? These two questions, asked over and over again by many colleagues, have inspired us to write this book.

Some of these colleagues were experimental psychologists who wished to understand how behavioral theories could be informed by neuroscience; others were neuroscientists seeking to bridge the conceptual gap from studies of individual neurons to behaviors of whole organisms. Clinical neurologists and neuropsychologists have also asked us whether neural network models might provide them with clinically useful insights into disorders of learning and memory. Unfortunately, many of these people found that their initial interest in modeling was thwarted by the mathematical details found in most papers and textbooks on computational neuroscience. Unable to follow the mathematics, these aspiring readers were left with the options of either accepting the author's conclusions on blind faith or ignoring them altogether.

Mathematics has long had the ability to inspire apprehension and awe among those not trained in its formalisms. A story is often told about the eighteenth century mathematician Léonard Euler, who was summoned to the court of Catherine the Great, the Czarina of Russia. She commissioned him to debate the French philosopher Diderot, who had offended her by questioning the existence of God and encouraging the spread of atheism in her court.

Appearing before the assembled courtiers, the two men faced off. Euler went first and announced that he had a mathematical proof of the existence of God. Advancing toward Diderot, Euler gravely explained:

"Monsieur, $(a + b^n)/n = x$, hence God exists!"

Of course, this claim was nonsensical, but Diderot—who understood no mathematics—could not make any response or rebuttal, and Euler won the argument by default. Soon after, Diderot left the royal court and returned to his native France.

Although mathematicians sometimes tell this anecdote to poke fun at the uninitiated, there is another more serious side to this tale. Euler won the debate not because his claims were valid, but simply because he couched his argument in mathematical jargon too esoteric for Diderot to understand. Over two hundred years later, researchers who develop computational models of brain and behavior still sometimes use the same ploy: masking their descriptions in complex mathematical equations that only other mathematicians can easily evaluate. This leaves the reader who lacks such training with two equally unpalatable options: either accept the modelers' (often grandiose) claims at face value or else—like Diderot—simply walk away.

However, we think there is a middle ground. It should be possible to communicate the fundamentals of connectionist modeling to a broader scientific community, by focusing on the underlying principles rather than the mathematical nuts and bolts. Like electrophysiology or neuroimaging, computational modeling is a tool for neuroscience and, while the methodological details are important, it is possible to appreciate the utility—and limitations—of these techniques without absorbing all the technical details. To this end, we have tried to describe the computational models in this book at an intuitive rather than a technical level, using illustrations and examples rather than equations. We have assumed no prior knowledge of computational modeling or mathematics on the part of the reader. For those who wish to delve more deeply into the formal details of the models, we have provided supplemental (but optional) MathBoxes, which appear throughout the text, as well as appendices that contain further implementation details for the model simulations.

We have two groups of readers in mind for this book: those with a specific interest in learning and memory and those who want to understand a sample case study illustrating how computational models have been integrated into an experimental program of research. To this broad readership, we have aimed to convey an intuitive understanding and appreciation of the promise, as well as the limits, of neural network models. If at the same time we excite a few of our readers to go on to become modelers themselves or to incorporate computational modeling into their own research programs through collaboration with modelers, all the better.

We believe that good models are born amidst a wealth of experimental studies and justify their existence by inspiring further empirical research. We had this in mind when we chose the word "modeling" rather than "models" in our subtitle: The emphasis here is on the process of modeling within the broader program of learning and memory research, rather than on the fine details of the models themselves. In contrast to the individual journal papers in which many of these modeling results were first reported, we have sought to convey a larger and more integrative picture here. This book tells the story of how models are built on prior experimental data and theoretical insights and then evolve toward a more comprehensive and coherent interpretation of a wide body of neurobiological and behavioral data.

We wrote this book in two parts. Part I (chapters 1 through 5) provides a tutorial introduction to selected topics in neuroscience, the psychology of learning and memory, and the theory of neural network models—all at the level of an advanced undergraduate textbook. We expect that some of this will be too elementary for many readers and therefore can be skipped, while other chapters will provide background material essential for understanding the second half of the book. Together, these early chapters are designed to level the playing field so that the book is accessible to anyone in the behavioral and neural sciences.

Part II, the core of the book, presents our current understanding of how the hippocampus cooperates with these other brain structures to support learning and memory in both animals and humans. In trying to answer the question, "What does the hippocampus do?" researchers have been forced to look beyond the hippocampus to seek a better understanding of the hippocampus's many partners in learning and memory, including the entorhinal cortex, the basal forebrain, the cerebellum, and the primary sensory and motor cortices.

Our emphasis throughout this book is on the function of brain structures as they give rise to behavior, rather than the molecular or neuronal details. Reflecting this functional approach to brain modeling, many of the models that we describe have their roots in psychological theories and research. We believe that appreciating these psychological roots is of more than just historical curiosity; rather, understanding how modern neural networks relate to well-studied models of learning in psychology provides us with an invaluable aid in understanding current efforts to develop models of the brain mechanisms of learning and memory.

In addition to covering our own theories and models in part II of the book, we review several related computational models, along with other qualitative and experimental studies of the neurobiology of learning and memory. In covering a range of models from a variety of researchers, we have tried to convey how it is possible for different models to capture different aspects of anatomy and physiology and different kinds of behaviors. In many cases, these models complement each other, the assumptions of one model being derived from the implications of another.

Given the wide range of academic disciplines covered in this book, many terms are used that may be unfamiliar to readers. The most important of these are printed in boldface when they first appear in the text and are accompanied there by a brief definition. These terms and definitions are then repeated at the end of the book in a glossary for easy reference.

Mark A. Gluck
Catherine E. Myers