11 Emergent Themes

This book has focused on computational models of the hippocampal region and its interaction with other brain structures during simple forms of learning. In this final chapter, we highlight the most important themes that have emerged, and we stress their implication for future computational modeling.

11.1 HIPPOCAMPAL FUNCTION CAN BEST BE UNDERSTOOD IN TERMS OF HOW THE HIPPOCAMPUS INTERACTS AND COOPERATES WITH THE FUNCTIONING OF OTHER BRAIN SYSTEMS.

The hippocampus interacts with other brain regions that are its partners in learning and memory, including the entorhinal cortex, the basal forebrain, the cerebellum, and the primary sensory and motor cortices. In this partnership, the hippocampus acts as an information processor, taking input from some brain regions, operating on it, and sending the output to other brain regions.

Borrowing methods from computer science and cognitive science, theoreticians have attempted to specify how the hippocampus carries out its processing role and how its output is used by its partners. Various theories have grown out of this approach, including models that suggest a hippocampal role in learning stimulus-stimulus relationships, in incorporating contextual information during ongoing learning, in forming cue configurations, and so forth (see chapter 6 for a review). In particular, computational models have demonstrated that these information-processing accounts of hippocampal-region function can explain why certain kinds of learning behaviors are selectively disrupted or altered after hippocampal region damage, while others appear to be unaffected. The challenge now is to return to the laboratory to determine the extent to which these computational models stand up to rigorous empirical testing and to modify them accordingly.

11.2 PARTIAL VERSUS COMPLETE LESIONS MAY DIFFER IN MORE THAN JUST DEGREE.

New selective techniques for experimental brain lesions are demonstrating that very subtle differences in the extent of damage can lead to broad differences in resulting behavior. For example, chapter 9 discussed how "hippocampal" damage can have very different implications depending on whether the lesion is selective to the hippocampus or extends to include nearby structures such as the entorhinal cortex. These recent results overturned a great deal of prior work which presumed that damage to any structure within the hippocampal region would produce similar results. In particular, many behavioral impairments resulting from damage to the hippocampal region turned out to depend not on the hippocampus at all, but rather on other nearby structures.[1] Several computational models (including those discussed in chapter 9) have made specific predictions about the effects of selective lesions on various tasks, and it is now up to researchers to empirically confirm or disconfirm those predictions.

Another area in which computational modeling may be helpful is in understanding the different implications of complete versus partial hippocampal lesion. Most amnesic individuals who are studied in research laboratories have more or less total loss of hippocampal tissue with resulting severe anterograde amnesia. But there are many people who have only partial loss of hippocampal tissue, resulting in intermediate memory impairments.

For example, some individuals with intractable epilepsy do not respond to drug treatment, and the seizures become so frequent and severe as to be life threatening. One option is surgical removal of the brain tissue where the seizures originate—often the anterior temporal lobes, including the amygdala and the anterior 2–3 cm of the hippocampus on one side of the brain.[2] Approximately 68% of patients undergoing this surgery become seizure-free, while another 23% report worthwhile reduction in seizures.[3]

These patients often also experience some memory loss following surgery, presumably due to partial loss of one hippocampus.[4] However, individuals with medial temporal lobe epilepsy generally report memory impairments even before surgery, and this is at least partially due to sclerosis (scarring) in the hippocampus as a result of repeated seizures.[5]

It should be possible to develop computational models that address partial hippocampal damage—for example, through removal of some percentage of hippocampal neurons—and predict the resulting constellation of behavioral impairments. This is a topic that has received relatively less study in the computational modeling community but that could have important clinical implications.

11.3 DISRUPTING A BRAIN SYSTEM HAS DIFFERENT EFFECTS THAN REMOVING IT.

Chapter 10 described data showing that eyeblink conditioning is not disrupted when the hippocampus is lesioned but is disrupted when the hippocampus is dysfunctional.[6] Such dysfunction can result from the lesion of modulatory structures such as medial septum or from the introduction of drugs such as scopolamine that disrupt modulatory inputs. Hippocampal dysfunction can also result from electrical stimulation or even the presence of scar tissue. Moreover, many so-called hippocampal lesions are, in fact, lesions to the fimbria/fornix that disrupt subcortical inputs to the hippocampus. In some cases, the effect of fimbria/fornix lesions appears similar to the effect of direct lesions to the hippocampus. But it is not clear whether this pattern will hold true over a wider range of behavioral tasks or whether fimbria/fornix lesions result in impairments that are more similar to the disruptions that occur from cholinergic antagonists such as scopolamine.

Here, computational modeling provides some important insight. For example, our scopolamine model in chapter 10 suggests that there are distinct and qualitative differences between hippocampal lesion and disruption due to changes in cholinergic inputs to the hippocampus. Although both may result in similar declarative memory deficits in humans, studies of simple learning paradigms such as classical conditioning illuminate subtle but important differences between lesion and disruption. The computational model provides a way to predict, *a priori,* which kinds of learning will be impaired or spared following lesion or disruption.

These distinctions and differences also have clinical implications. Memory rehabilitation techniques that work well in an individual with hippocampal lesion (such as following medial temporal damage) may be less effective in individuals with hippocampal disruption (such as following basal forebrain damage) and vice versa. Understanding the pattern of impaired and spared memory abilities in each population may help clinicians to develop therapies that are tailored to an individual's etiology and abilities.

11.4 STUDIES OF THE SIMPLEST FORMS OF ANIMAL LEARNING MAY BOOTSTRAP US TOWARD UNDERSTANDING MORE COMPLEX ASPECTS OF LEARNING AND MEMORY IN HUMANS.

Much of the modeling that is presented in this book begins by addressing data from studies of hippocampal function in animal learning and then builds on this animal research to seek a clearer understanding of the hippocampus in human learning and memory. Because comparative studies of animal and

human learning have been so essential to the study of the neurobiology of learning, it is important to understand how complex forms of human learning relate to the more basic forms of learning that are seen in all animals.

"Typically, evolution works through endless variations on a limited repertory of themes," wrote the noted psychologist W. K. Estes. "Similar mechanisms for accomplishing the same functions appear at many stages and levels of both phylogeny and ontogeny, and, as a consequence, clues to understanding complex processes of human cognition sometimes come from studying simpler forms."[7]

For example, suppose one wants to study how people learn to categorize objects. A human has many different mechanisms for learning, ranging from simple classical conditioning all the way up to verbally mediated rules. This is far too much to attempt to study at once. Instead, a researcher may use a simpler model system, which abstracts out all but one kind of learning. One model system is rabbit eyeblink conditioning. It is generally assumed that rabbits do not have access to many higher-level forms of learning (such as verbally mediated rules) but are instead restricted to simpler forms of associative learning. The hope is that findings from studies of animal conditioning will scale up to more complex forms of learning in humans. Indeed, many of the basic principles of learning that are seen in conditioning are also manifested in higher-level forms of human cognition, such as learning complex categories and concepts.[8]

A related point was discussed in chapter 8, in which studies of cortical remapping in animals inspired a broader program for remediation of language-learning impairment in children. What is especially salient here is that the original animal work involved simple conditioning behaviors. On the surface, these learning behaviors have very little to do with the extraordinary complexities of human language acquisition, a subject that is still poorly understood. But by using these simple animal paradigms, researchers were able to grasp some of the fundamentals of how learning induces plasticity in the cortex. Apparently, the rules that govern this plasticity are independent of the particular learning paradigm or species being studied. Thus, researchers took insights gained by studying simpler learning processes and applied them to a far more complex phenomenon: language learning.

11.5 KEEP IT SIMPLE. KEEP IT USEFUL. KEEP IT TESTABLE.

Keep It Simple

A model's value to science is often inversely related to its complexity. The point of a model is to reduce the complexity of the real world and focus on a

reduced set of phenomena. To accomplish this, the modeler must resist the urge to make the model as complex and "realistic" as possible. When models become so intricate as to be comprehensible only to other modelers, they lose their ability to function as effective tools for guiding empirical research. Whenever possible, we believe that a modeler should try to build new models on top of old ones, adding a minimum of new assumptions to capture a maximal amount of new data. The alternative is to add new parameters and complexity each time a new piece of data needs to be accounted for. Such a model quickly spirals out of control, becoming too complicated to be useful. Rather, the goal of modeling, as described by W. K. Estes, should be to "find out whether [one can] arrive at a small number of concepts and principles that, applied in varying combinations, could help illuminate a wide variety of psychological phenomena."[9] In other words, a good model is one that exemplifies the principle of parsimony, proposing the simplest possible hypothesis to explain the data at hand.

Keep It Useful

A model that is simple but realistic still has one more challenge: It must be useful. Models are tools and different models are useful for different purposes. A frequent trap for modelers and those who evaluate them is to judge a model's "utility" against a checklist of phenomena, such as determining which model accounts for the "most" behavioral data or incorporates the "most" anatomical detail. These comparisons must be made with caution. Anatomical detail is a good thing in a model of anatomical substrates, but it can needlessly complicate a model of higher-level behaviors. Similarly, a model that can correctly account for only a few pieces of behavioral data may be useful—if those data are particularly puzzling or important. A related trap is to assume that all models are necessarily in competition and that only one can be shown to be "correct." Rather, it is quite possible for two or more models to capture different aspects of anatomy and physiology and different kinds of behaviors. In many cases, different models can complement each other in much the same way that an architect may use several line drawings of the same building, ranging from close-ups of the wiring to broad views of the façade. Each describes a different aspect of the same subject, and each is useful for a different purpose.

Keep It Testable

Unlike the architect's diagrams, models must be more than mere illustrations. It is not enough for the model to show behaviors that are similar to those already observed in the laboratory. The model must generate strong,

testable predictions. If animals or humans fail to behave in the way the model does, the model is discredited and must be revised or abandoned. Moreover, even if animal and human studies generate data that are in every way concordant with the model's predictions, this does not prove that the brain operates in the same way. Rather, it only shows that the assumptions embedded in the model are one possible way that the brain could work; there could be many other models that embed other assumptions and that generate the same data. However, as more and more data accumulate, we may become more and more convinced that a given model is the most plausible way to understand brain function. Thus, it is incumbent on the researchers who propose a model to note how their model can be tested and which model predictions, if disproven, would require the model to be abandoned.

CONCLUSION

We have tried to convey in this book how and why computational models have advanced our understanding of the neural bases of learning and memory. We feel strongly that it should be possible to understand these models and appreciate their value to science without delving into mathematical details. We have tried to communicate the fundamentals of neural-network modeling of hippocampal function to the broader scientific community, by focusing on the underlying principles rather than on the mathematical nuts and bolts.

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.

Many questions about hippocampal function in learning still remain unanswered. We believe that good models should raise as many new questions as answer old ones. Some of these open questions are empirical, and we have suggested, at several places in the book, what we think are some of the more pressing empirical issues that need to be resolved by further behavioral and neurobiological studies in animals and humans. Other open questions and unresolved issues that the models raise are of a more theoretical nature and suggest new modeling directions for the future. Although we have aimed this book primarily at nonmodelers, we hope that we may have excited 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.