323
324
325
327
327
327
328
328
331
333

1 Introduction

We will outline in this introductory chapter the major focus, specific tools, and the strategies of computational neuroscience. The complexity of the brain makes it necessary to clarify how we attempt to describe it and what we expect from explanatory models. We will argue that theoretical and computational studies are important in understanding experimental measurements. We will discuss the specific role of models in computational neuroscience and argue that they have to be chosen carefully if they are to be useful for advancing our knowledge of the information-processing principles in the brain. Models must not only be able to summarize known experimental data, but must also be able to make predictions that can be verified experimentally. We also include a short guide to the book to show the path we are going to take to explore this fascinating scientific subject.

1.1 What is computational neuroscience?

In the scientific area now commonly called computational neuroscience we utilize distinct techniques and ask specific questions aimed at advancing our understanding of the nervous system. This specific scientific area might be defined as:

> *Computational neuroscience is the theoretical study of the brain to uncover the principles and mechanisms that guide the development, organization, information processing, and mental abilities of the nervous system.*

Thus, computational neuroscience is a specialization within neuroscience. Neuroscience itself is a scientific area with many different aspects. Its aim is to understand the nervous system, in particular the central nervous system that we call the brain. The brain is studied by researchers who belong to diverse disciplines such as physiology, psychology, medicine, computer science, and mathematics, to name but a few. Neuroscience emerged from the realization that interdisciplinary studies are vital to further our understanding of the brain. The brain is one of the most complex systems ever encountered in nature, and there are still many questions that we can only attack through combined efforts. How does the brain work? What are the biological mechanisms involved? How is it organized? What are the information-processing principles used to solve complex tasks such as perception? How did it evolve? How does it change during the lifetime of the organisms? What is the effect of damage to particular areas and the possibilities of rehabilitation? What are the origins of degenerative diseases and possible treatments? All these basic questions are asked by neuroscientists.

1.1.1 The tools and specializations in neuroscience

Many techniques are employed within neuroscience to answer these questions. These include genetic manipulation, *in vivo* and *in vitro* recording of cell activities, optical imaging, functional magnetoresonance scanning, psychophysical measurement, and computer simulation. Each of these techniques is complicated and laborious enough to justify a specialization within neuroscience. We therefore speak of neurophysiologists, cognitive scientists, and anatomists. It is, however, vital for any neuroscientist to develop a basic knowledge of all major techniques in order to understand and use the contributions made by them. The significance of any technique has to be evaluated with a view to its specific problems and limitations as well as its specific aim of the technique. Computational neuroscience is an increasingly important area of neuroscience and a basic understanding of this field has become essential for all neuroscientists.

1.1.2 The focus of computational neuroscience

Computational neuroscience attempts to develop and test hypotheses about the functional mechanisms of the brain. A major focus is therefore the development and evaluation of *models*. The scientific area that is the subject of this book is also known as 'theoretical neuroscience'. Computational neuroscience can be viewed as a specialization within theoretical neuroscience, which employs computers to simulate models. The major reason for using computers is the complexity of many models, which are often analytically intractable. For such models we have to employ carefully designed numerical experiments in order to be able to compare these models to experimental data. However, we do not want to restrict our studies to this tool because analytical studies can often give us a deeper and more controlled insight into the features of models and the reasons behind numerical findings. Whenever possible, we try to include analytical techniques, and we will use the term 'computational neuroscience' synonymously with theoretical neuroscience. The word 'computational' also emphasizes that we are interested in particular in the computational aspects of brain functions.

We have included some examples of analytical techniques in order to give you a taste of some of those powerful techniques. Not every neuroscientist has to perform such calculations, but it is necessary to comprehend the general ideas if you are to get support from specialists in these techniques when required in your own research. However, it is very instructive to perform some numerical experiments yourself. We therefore included an introduction to a modern programming environment that is very well suited to many models in neuroscience. Writing programs and creating advanced graphics can be easily learned within a short time even without extensive computer knowledge.

Although computational neuroscience is theoretical by nature, it is important to bear in mind that the models have to be measured against experimental data; they are otherwise useless for understanding the brain. Only experimental measurements on the real brain can verify 'what' the brain does. In contrast to the experimental domain, computational neuroscience tries to speculate 'how' the brain does it. The speculations are developed into hypotheses, realized into models, evaluated analytically or numerically, and tested against experimental data. We will discuss specific examples of models on several levels (and related evaluation techniques) throughout this book.

1.2 Domains in computational neuroscience

The nervous system has many levels of organization on spatial scales ranging from the molecular level of a few ångstroms ($1\ \text{Å} = 10^{-10}\text{m}$) to that of the whole nervous system on the scale of 1 metre (see Fig. 1.1), and biological mechanisms on all these levels are important for the functioning of the brain. Where do we start when trying to explain brain functions? Do we have to rebuild the brain in its entirety, for example, in a computer, in order to 'understand' it? There are arguments suggesting that we can never achieve this. However, even if it were possible to simulate a whole brain on a computer with all the details from its biochemistry to its large-scale organization, this would not necessarily mean that we had a better explanation of brain functions. What we are really looking for is a better comprehension of brain mechanisms on explanatory levels that are, for example, suitable for applications such as the development of advanced medical treatments. It is therefore important to learn about the art of *abstraction*, making a suitable simplification of a system without abolishing the important features we want to understand.

1.2.1 Levels of abstraction

Which level is most appropriate for the investigation and abstraction depends on the scientific question asked. For example, Parkinson disease is 'caused' by the death of dopaminergic neurons in the substantia nigra, and there are signs that a genetic predisposition might act together with biochemical processes in single neurons that can cause the death of dopaminergic neurons. A detailed investigation on a neuronal level with detailed neuron models is thus obvious. However, we also know about the important role of dopamine in the initiation of motor actions, and a more global system level study (often with less detailed neuron models) is necessary to comprehend the full scale of impairment and to develop better methods of coping with the condition. Therefore, the condition must be studied at various levels and connections must be made between the different levels in order to elucidate how small-scale factors, such as genetic mechanisms or biochemical processes in single neurons, can influence the characteristics of large-scale systems, such as the behaviour of an organism.

1.2.2 Levels of organization in the nervous system

Different levels of organization in the nervous system are illustrated in Fig. 1.1. An easily recognizable structure in the nervous system is the neuron, which is a cell that is specialized in signal processing. Depending on environmental conditions it is able to generate an electric potential that is used to transmit information to other cells to which it is connected. Mechanisms on a subcellular level are certainly important for such information-processing capabilities. Some processes in the neuron utilize cascades of biochemical reactions that must be understood on a molecular level. These include, for example, the transcription of genetic information that influences information processing in the nervous system. Many structures within the neuron can be identified with specific functions, for example, mitochondria and synapses, of which the latter are particularly important for our understanding of signal processing in the nervous system. The complexity of a single neuron, and even that of isolated subcellular mechanisms, often makes computational studies essential for the development and verification of hypotheses. The theory of action potential generation in neurons developed by *Hodgkin* and *Huxley* is similar in elegance and importance to *Maxwell*'s theory of electromagnetism.

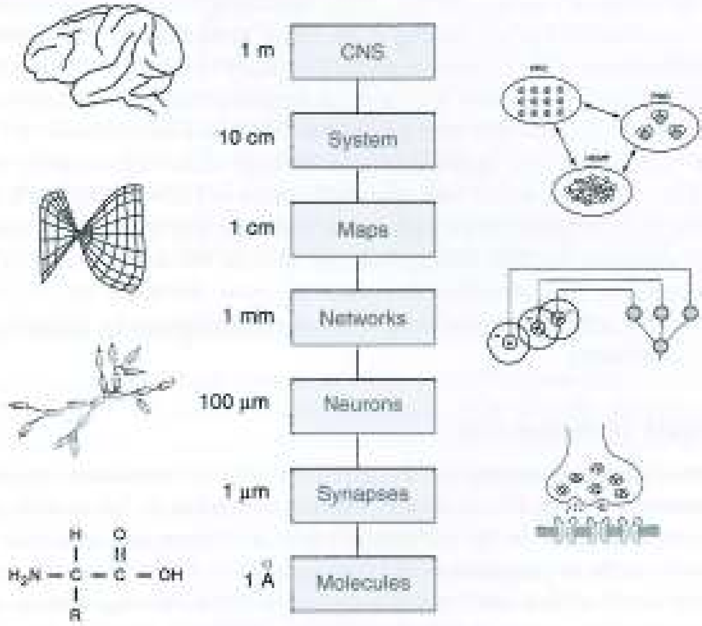

Fig. 1.1 Some levels of organization in the central nervous system on different scales. The illustrations include, from top to bottom, an outline of the brain, a system-level model of working memory (discussed in Chapter 11), a self-organized (Kohonen) map, speculation about the circuit behind orientation-sensitive neurons in V1 by Hubel and Wiesel, a compartmental model of a neuron, a chemical synapse, and an amino acid molecule [adapted from Churchland and Sejnowski, *The computational brain*, MIT Press, 1992].

Single neurons are certainly not the whole story. Neurons contact each other and thereby build networks. A small number of interconnected neurons can exhibit complex behaviour and information-processing capabilities not present in a single neuron. The understanding of such networks of interacting neurons is a major domain in computational neuroscience, perhaps because there is so little understanding about such nonlinear interacting systems. Networks have additional information-processing capabilities beyond that of single neurons, such as representing information in a distributed way. Examples of this are topographic maps of sensory stimuli, frequently found in the brain. We concentrate in particular on this level in this book.

And it doesn't stop there. Networks with a specific architecture and specialized information-processing capabilities are incorporated into larger structures that are able to perform even more complex information-processing tasks. A study of the brain on this level is certainly essential to understand higher-order brain functions, and it is at this level where we have probably the least understanding. The central nervous system depends strongly on the interaction of many specialized subsystems, and understanding such interactions seems central for understanding brain functions. It is probable that only the dynamic interaction of several brain areas enables high-level brain functions. We are still unable to reproduce many basic human skills with machines, and a deeper understanding of the function of the brain is desirable for many reasons, including the development of advanced technical applications and medical treatments. The understanding of the brain on the system level is particularly challenging as we have few tools and methods capable of characterizing such large systems. A further difficulty is that the nervous system is not an isolated system but interacts strongly with the environment.

It is important for all neuroscientists to develop a basic understanding of the functionalities on different scales in the brain, although the individual researcher might specialize in mechanisms on a certain scale. Computational neuroscience can help the investigations at all levels of description, and it is not surprising that computational neuroscientists investigate different types of models at different levels of description. The contribution of computational neuroscience is particularly important in understanding the nonlinear interactions among many subprocesses, a characteristic of the brain that is thought to be essential in enabling many brain processes. It is also important to comprehend the interaction between the different levels of description, and computational methods are often suited to investigating these relationships, for example, to bridge the gap between a physiological measurement and the understanding of it in order to explain behavioural correlates of an organism. We will discuss examples of models at different levels of abstraction throughout this book.

1.2.3 The integrated approach

Ideally, we want to integrate experimental facts from different investigational levels into a coherent model of how the brain works as illustrated in Fig. 1.2. Important experimental input to computational neuroscience comes from neurobiology, for example, the mechanisms determining synaptic efficiencies, from neurophysiology in which the behaviour of single neurons is investigated, and from psychology, in which behavioural effects are studied via psychophysical experiments. Computational neuroscientists can then utilize mathematical models for describing the experimental facts, borrowing methods from a wide variety of disciplines such as mathematics, physics, computer science, and statistics. Such formulations and studies of hypotheses with the aid of models should lead to specific experimental predictions that have to be verified experimentally as mentioned above. The comparison of model predictions with experimental data can then be used to refine the hypothesis and to develop more accurate models or models that can shed light on different phenomena. Studies within computational neuroscience can also help to develop applications such as the advanced analysis of brain imaging data, technical applications that utilize brain-like computations, and, ultimately, the development of advanced treatments for patients with brain damage.

Some models in the literature have illustrated that they are able to relate simultaneously to experimental findings on different levels. Those models are often comprised of single elements that rely on neurobiological mechanisms such as biologically plausible synaptic plasticity. The behaviour of the elements in the whole system can, on the other hand, reflect experimental findings from electrophysiological studies, and the activity of specific modules can be related to brain imaging studies. The computational power

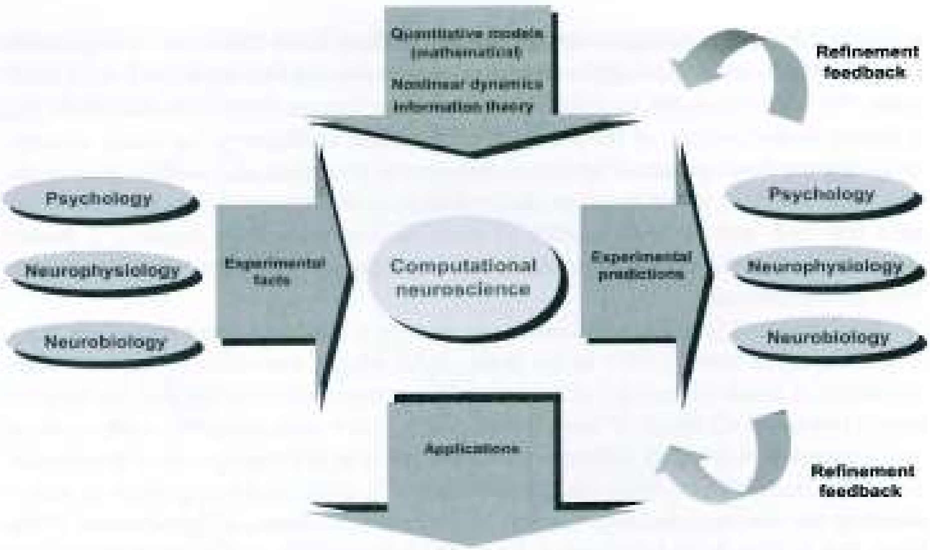

Fig. 1.2 Illustration of the role of computational neuroscience in the integration of experimental facts from different levels of investigation. The models developed in computational neuroscience have to be able to make predictions that can be verified experimentally. The close comparison of experiments with model predictions can then be used to make refinements in the models (or may lead to the development of new approaches) that can further our understanding of brain systems and could also lead to new predictions that have to be verified experimentally. This may also lead to applications in the analysis of experimental data and the development of advanced patient treatments [adapted from Gustavo Deco, personal communication].

of recent computer systems and the deepened knowledge of the brain from experimental research make such models increasingly feasible. We will outline some of these models in later chapters of this book. These models rely on the fundamentals that we introduce in this book, and the feasibility of the study of such models relies strongly on suitable simplification. Explaining the reasons for making such simplifications is therefore at the heart of learning computational neuroscience.

1.3 What is a model?

Modelling is an integral part in many scientific disciplines, and neuroscience is no exception. Indeed, the more complex a system is, the more we have to make simplifications and build model systems that can give us insights into some aspects of the complex system under investigation. The term 'model' frequently appears in many scientific papers and, it seems, describes a vast variety of constructs. Some papers present a single formula as a model, some papers fill several pages with computer code, and some describe with words a hypothetical system. It is important to understand what a model is and, in particular, what the purposes of models are.

To start with, it is important to distinguish a model from a hypotheses or a theory. Scientists develop hypotheses of the underlying mechanisms of a system that have to be tested against reality. In order to test a specific feature of a hypothesis we build a model that can be evaluated. Sometimes we try to mimic a real system by artificial means in order to test this system in different conditions or to make measurements that would not be possible in the 'real' system. A model is therefore a simplification of a system in order to test particular aspects of a system or a hypothesis.

> *Models are abstractions of real world systems or implementations of an hypotheses in order to investigate particular questions or to demonstrate particular features of a system or a hypothesis.*

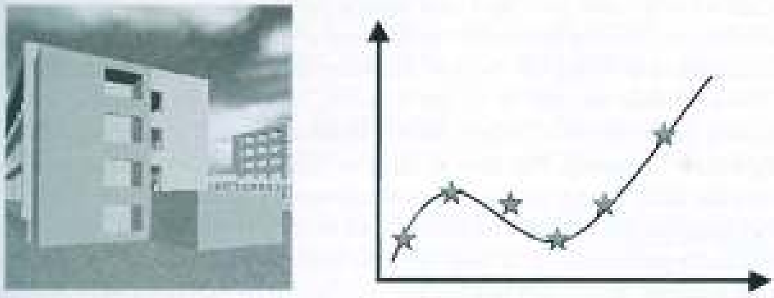

Fig. 1.3 What is a model? On the left is a computer model of a house giving a three-dimensional impression of the design. The right graph shows some data points, for example, from experimental measurements, and a curve is shown that fits these data points reasonably well. The curve can be a simple mathematical formula that fits the data points (heuristic model) or result from more detailed models of the underlying system.

A good example is the use of models in architecture. Small-scale paper models of buildings, or computer graphics generated with sophisticated three-dimensional graphic packages, can be used to give a first impression of the physical appearance and aesthetic composition of a design. A model has a particular purpose attached to it and has to be viewed in this light. A model is not a recreation of the 'real' thing. The paper model of a house cannot be used to test the stability of the construction, and a building engineer uses different models for this purpose. In such models it is important to scale down physical properties of the building material regardless of the physical appearance such as the colour of the building.

1.3.1 Descriptive models

In science we typically represent experimental data in the form of graphs and subsequently seek to describe these data points with a mathematical function. An example of this is the 'modelling' of response properties (receptive fields) of neurons in the lateral geniculate nucleus (LGN), which can be fitted with a specific class of functions called *Gabor functions* by adjusting the parameters within this class of functions accordingly. The Gabor functions are therefore said to be a 'model' of the receptive fields in the LGN. Of course, this *phenomenological model* does not tell us anything about the biophysical mechanisms underlying the formation of receptive fields and why cells respond in this particular way, and such a 'model' therefore seems rather limited. Nevertheless, it can be useful to have a functional description of the response properties of LGN cells. Such parametric models are a shorthand description of experimental data that can be used in various ways. For example, if we want to study a model of the primary visual cortex to which these cells project, then it is much easier to use the parametric form of LGN responses as input to the cortical model rather than including complicated models of the earlier visual pathway in detail.

1.3.2 Explanatory models

As scientists we want to find the roots of natural phenomena. The explanations we are seeking are usually more profound then merely parametrizing experimental data with specific functions. Most of the models in this book are intended to encapsulate processes that are thought of as being the basis of the information-processing capabilities of the brain. These include models of single neurons, networks of neurons, and specific architectures capturing brain organizations. Models are used to study specific aspects of a hypothesis or theory, but also to help to interpret experimental data. A good example of the latter type of model on a system level is a technique from statistics called *structural equation modelling*. The idea of such models is to make an educated guess at the functional organization of the brain in order to deduce effective connectivities in the brain from imaging data. Such models are very useful in analysing experimental data. We will not elaborate on these types of models in this book. Instead we will concentrate on synthetic models that can illuminate the principles of information processing in the brain, which we speculate are the building blocks of brain functions.

A major role of models in science is to illustrate principles underlying natural phenomena on a conceptual level. Sometimes the level of simplification and abstraction is so crude that scientist talk about *toy models*. However, this term may obscure the importance of such models in science. The simplifications made in such models might be necessary in order to employ analytical methods to analyse such models at depth that is not possible in more realistic models. The educational importance of those 'toy models' should not be underestimated, in particular, in demonstrating the principal mechanisms of natural phenomena.

The current state of neuroscience, often still exploratory in nature, frequently makes it difficult to find the right level of abstraction as discussed in the last section. Some models in computational neuroscience have certainly been too abstract to justify some of the claims derived from them. On the other hand, there is a danger in keeping more details than are essential for a scientific argument. Models are intended to simplify and thereby to identify which details of the biology are essential to explain particular aspect of a system. Modelling the brain is not a contest in recreating the brain in all the details on a computer. It would indeed be questionable how such models would add to understanding the functionality of the brain. Models have to be constructed carefully, and reasonable simplification demands high scientific skills. Justifications of the assumptions in models have a high priority in scientific investigations. The purpose of models is to better comprehend the function of complex systems, and simplicity should be a major guide in designing appropriate models.

1.4 Emergence and adaptation

1.4.1 The computer analogy

Standard computers, such as PCs and workstations, have one or perhaps a small number of central processors. Each processor is rather complex with specialized hardware and microprograms implementing a variety of functions such as loading data into registers, adding, multiplying, and comparing data, as well as communicating with external devices. These basic functions can be executed by instructions that are binary data loaded into a special interpreter module. Complicated data processing can be achieved by writing often lengthy programs that are instructions representing a collection of the basic processor functions. Programming a computer means that we have to instruct the machine to follow precisely all the steps that we have figured out beforehand can solve a particular problem. The sophistication of the computer reflects basically the smartness of the programmer.

By contrast, information processing in the brain is very different in several respects. Briefly, it employs simpler processing elements, but lots of them. To explore the information processing in networks of neurons we will mainly employ only very simple abstractions of real neurons, and we will call these fundamental processing units *nodes* to stress this drastic simplification. Nodes can be implemented in hardware or simulated on a standard computer; for the discussions in this book this does not make a difference. We keep the functionality of nodes as simple as possible for the sake of employing lots of them, typically hundreds, thousands, or many more. The employment of many parallel working processors has led to the term *parallel distributed processing* being used in this area. However, with this term one is tempted to think that the processes are independent because only processes that are independent can be processed on different processors in parallel. In contrast to this, a major ingredient of the information processing in the brain is certainly the *interaction of nodes*, and the interaction of nodes is accomplished by assembling them into large networks.

1.4.2 Emergence

In the study of *neural networks* we are interested in understanding the consequences of interacting nodes. It is the interaction that enables processing abilities not present in single nodes. Such capabilities are good examples of emergent properties in *rule-based systems*. *Emergence* is the single most defining property of neural computation, distinguishing it from parallel computing in classical computer science, which is mainly designed to speed up processing by distributing independent algorithmic threads. Interacting systems can have unique properties beyond the mere multiplication of single processor capabilities. It is this type of ability we want to explore and utilize with neural networks. We label these system properties as 'emergent' to stress that we did not encode them directly into the system. To appreciate this better we should distinguish the description of a system on two levels, the level of basic rules defining the system and the level of description aimed at understanding the consequences of such rules.

Scientific explanations have been dominated in the past by the formulation of a set of principles and rules that govern a system. A system can indeed be defined by a set of rules like a game that is defined by rules. In science we make the assumption that natural systems are governed by a finite set of rules. The search for these basic rules, or fundamental laws as they are called in this case, was indeed the central scientific quest for centuries. It is not easy to find such laws, but enormous progress has nevertheless been made. Newton's laws defining classical mechanics or Maxwell's equations of electromagnetism are beautiful examples of fundamental laws in nature. We do not have a theory of the brain on this level, and some have argued that we might never find a simple set of rules explaining brain functions. However, in many scientific disciplines we have begun to realize that, even with a given set of rules, we might still not have a sufficient understanding of the systems. Knowing the rules of the card game 'bridge' is not sufficient to be a good bridge player.

Rules define a system completely and it can therefore be argued that all the properties of the system are encoded in these rules. However, we have to realize that even a small set of rules can generate a multitude of behaviours of the systems that are difficult to understand from the basic rules, and different levels of description might be more appropriate. For example, thermodynamics can describe the macroscopic behaviour of systems of many weakly interacting particles appropriately even though the systems are governed by other microscopic rules. On the other hand, there are *emergent properties* in Newtonian systems that are not well described by classical thermodynamics, for example, turbulent fluids. A deeper understanding emergent properties is becoming a central topic in the science of the 21st century.

1.4.3 Adaptive systems

The importance of emergent properties in networks of simple processing devices is not the only deviation from traditional information processing that we think is crucial in understanding brain functions. An important additional ingredient is that the brain is an example of an *adaptive system*. In our context we define adaptation as the ability of the system to adjust its response to stimuli depending upon the environment. Humans are a good example of systems that have mastered such abilities. This is an area that has attracted a lot of interest in the engineering community, as learning systems, systems that are able to change their behaviour based on examples in order to solve information-processing demands, have the potential to solve problems when traditional algorithmic methods have not been able to produce sufficient results. Adaptation has two major virtues. One is, as just mentioned, the possibility of solving information-processing demands for which explicit algorithms are not yet known. The second concerns our aim to build systems that can cope with a continuously changing environments.

A lot of research in the area of neural networks is dedicated to the understanding of learning in particular networks. Engineering applications of neural networks are not bound by biological plausibility. In contrast to this, we want to concentrate in this book on biologically plausible learning mechanisms that can help us to comprehend the functionality of the brain.

1.5 From exploration to a theory of the brain

The brain is still a largely unknown territory, and exploring it is still a major domain in neuroscience. Recordings with micro-electrodes from single cells contribute significantly to this exploration, and searching for the response properties of neurons is at least as exciting as it must have been for explorers like Marco Polo to discover new lands. With new brain imaging techniques such as fMRI (functional magnetic resonance imaging) we are now able to monitor the living brains of subjects performing specific mental tasks. All this is essential to advance our knowledge of what the brain does. Many data on the brain have been gathered with many different techniques, and we are now slowly entering a new phase in neuroscience, that of formulating more quantitative hypotheses of brain functions. This, in turn, demands some more specific experimental analysis and more dedicated tests of such hypotheses. Consequently, the experimental style is slowly changing. It is increasingly important to formulate alternative hypotheses more precisely and to quantify such hypotheses in such a way that experimental tests can verify or disprove them.

This sounds a lot like quantitative scientific areas such as chemistry or physics. But, can there be a theory of the brain? To fully understand each feature of an individual brain we have to know exactly all of the structural details of the particular brain as well as its precise environment and current state. This is similar to other scientific areas such as physics. To describe completely the physics of an individual aeroplane we have to know the precise location and form of each nut and bolt and all other structural details up to the amount of dirt on the wings and the details of the air it is flying through. As another example, a pot of boiling water consists of lots of individual molecules in a very dynamic state. Measuring all the microscopic details in the last two examples seems impossible. Yet, we have a fairly good understanding of the process of boiling water and why an aeroplane can fly. This wasn't possible overnight but is the result of dedicated scientific research over the last centuries. Thus, once again, the right level of description such as the average behaviour of an idealized gas or the principal mechanisms of flowing air is important. We know today about the essential quantum nature of atomic and subatomic interactions, but a description of Mount Everest on this level is not reasonable. There are geological theories of mountain formation that are more appropriate than employing quantum theory on these questions.

There is no reason to conjecture that the brain cannot be tractable with similar scientific rigour. The brain is certainly more complex than a gas of weakly interacting atoms. However, there are very fundamental questions that we can attack, for example, how a network can store memories that can be recalled in an associative way. We have today a good understanding of the basic mechanisms behind this ability, which is already present in networks of fairly simple processing elements. Another fundamental question is why the brain is relatively stable while still able to adapt to novel environments. Brain theories of this kind are now emerging. It is too early to talk about a theory of the brain, and we might never be able to explain the brain with a few equations. Nevertheless, many researchers are convinced that we will significantly advance our knowledge of how the brain works in the next decades.

Further reading

Patricia S. Churchland and Terrence J. Sejnowski

The computational brain, MIT Press, 1992.

This classic book on computational issues of brain function discusses many topics that are central in computational neuroscience. It also includes discussions of several specific brain processes not discussed here. This is very much recommended reading for everyone interested in computational neuroscience.

Peter Dayan and Laurence F. Abbott

Theoretical neuroscience, MIT Press, 2001.

Two of the leading scientists in computational neuroscience have recently released a book that will become essential for everyone who wants to further their understanding on the theory of neuroscience. This book is very much recommended reading to further understanding of most of the issues outlined in this book.