OXFORD

THIRD EDITION

FUNDAMENTALS OF COMPUTATIONAL NEUROSCIENCE

THOMAS P. TRAPPENBERG

Fundamentals of Computational Neuroscience

Fundamentals of Computational Neuroscience
Third Edition

Thomas P. Trappenberg
Dalhousie University

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2002
Second Edition published in 2010
Third Edition published in 2023

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 100106, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2022943170

ISBN 978–0–19–286936–4

DOI: 10.1093/oso/9780192869364.001.0001

Printed in the UK by
Ashford Colour Press Ltd, Gosport, Hampshire

Preface

Computational neuroscience is still a young and dynamically developing discipline, and some choice of topics and presentation style had to be made. This text introduces some fundamental concepts, with an emphasis on basic neuronal models and network properties. In contrast to the common research literature, this book is trying to paint the larger picture and tries to emphasize some of the concepts and assumptions for simplifications used in the scientific technique of modelling.

Computational neuroscience and Artificial Intelligence (AI) are close cousins. The term AI is said to be invented at the Dartmouth workshop in 1956 with many famous participants including psychiatrist Ross Ashby, the neurophysiologist Warren McCulloch who created one of the first mathematical neuron models, and Arthur Samuel, one of the pioneers in reinforcement learning. Computational models of neural systems such as models of neurons are much older, but connecting learning and cognitive systems created excitement over the possibility to better understand mind. The invention of learning machines has revolutionized many applications as recently seen in the dramatic progress of machine vision and natural language processing through deep learning.

While there has been much recent progress in machine learning, researchers in this area often wonder how the brain works. It sometimes seems that scientific progress oscillates between computational neuroscience and machine learning. For example, the progress of neural networks and statistical learning theory in the later 1980s and early 1990s was followed by enormous activities in computational neuroscience in the 1990s and early 2000s. For the last decade, deep learning has occupied an explosive growth in machine learning and data science, and now the time seems ripe for more renewed interest in looking more closely at the brain for inspirations to go deeper. This is fuelled by the increasing realization of limitations of deep learning, in particular with the challenge of learning semantic knowledge with limited data and the ability to transfer knowledge to situations that are not directly represented in the learning set.

In this new edition of my book, I tried to incorporate many of the recent lessons from deep learning. While there are excellent books on deep learning, our emphasis here is their connection to brain processing. An important aspect is thereby the concepts of representational learning and computation with uncertainties. Also, I now included gated recurrent neural networks that are becoming an important fundamental mechanisms when thinking about brain processing. While we will not be able to dive into all the recent progress, I hope that the text will guide further specific studies and research. Furthermore, it was important for me to streamline the existing text. I hope that I improved the readability of some of the text and even removed parts that seem less relevant to study the most basic fundamentals.

The themes included in this book are chosen to provide some path through the different levels of description of the brain. Chapter 1 provides a high-level overview and some fundamental questions about brain theories, a brief discussion about the role of modelling, and some basic neuroscience facts that are useful to keep in mind for later use. We also review the essential scientific programming in Python and the basic mathematical and statistical concept used in the book. Chapters 2–4 focus on basic mechanisms and modelling of single neurons or population averages. This starts from a fairly detailed discussion of changes in the membrane potentials through ion channels, spike generations, and synaptic plasticity, with increasingly abstractions in the following chapters. Chapters 5–7 describe the information-processing capabilities of basic networks, including feedforward and competitive recurrent networks. The last part of the book describes some examples of combining such elementary networks as well as some examples of more system-level models of the brain.

Most models in the book are quite general and are aimed at illustrating basic mechanisms of information processing in the brain. In the research literature, the basic elements reviewed in this book are often combined in specific ways to model specific brain areas. Our hope is that the study of the basic models in this book will enable the reader to follow some of the recent research literature in computational neuroscience.

While we tried to emphasize some important concepts, we did not want to give the impression that the chosen path is the only direction in computational neuroscience. Therefore, we sometimes mention concepts without extensive discussion. These comments are intended to increase the reader's awareness of some issues and to provide some keywords to facilitate further literature searches. Also, while some examples of specific brain areas are mentioned in this book, a comprehensive review of models in computational neuroscience is beyond the scope of this text. We do not claim that this book covers all aspects of computational neuroscience nor do we claim it to be the only approach to this area, but we hope that it will contribute to the discussion.

Mathematical formulas

This book includes mathematical formulas and concepts. We use mathematical language and concepts strictly as practical tools and to communicate ideas in contrast to using such formalism for mathematical proofs. We thereby tried to balance detailed mathematical notations with readability and communicating the basic concepts. From readers with less extensive training in such formal systems I ask for patience. We did not try to avoid mathematical formulations since such notations allow a brevity in communication that would be lengthy with plain written language. The chosen level of mathematical descriptions are mainly intended to be translated directly into programs and other quantitative evaluations.

There is no reason to be afraid of formulas, and it is important to see beyond the symbols and to understand their meaning. Many mathematics notations are invented to simplify descriptions. This includes the use of vectors and matrices, which will drastically shorten the specification of network models. We provide review chapters in the first part of the book to review such notations. We recommend some tutorials on such materials to allow students to move beyond these technicalities in the main text.

Most models in this book describe the change of a quantity with time, such as the change of a membrane potential after synaptic input or synaptic strength values over time during learning. Equations that describe such changes are called differential equations. A comprehensive knowledge of the theory of differential equations is not required for understanding this book. However, discussing the consequences of specific differential equations and simulating them with computer programs is at the heart of this book. I hope our treatment will encourage a new look into a topic that sometimes seems overwhelming when treated in specialized classes. We will specifically become familiar with a simple yet telling example of a differential equation, that of a leaky integrator. A basic knowledge of the numerical approaches to solving differential equations is essential for this book and many other dynamic modelling approaches. Thus, we also include a review of differential equations and their numerical integration.

Another mathematical theory, that of random numbers, is also reviewed in the third chapter. The language of probability theory is very useful in computational neuroscience and should be taught in such a course. In neuroscience (as in other disciplines), we often get different values each time we perform a measurement, and random numbers describe such situations. We often think of these circumstances as noise, but it is also useful to think about random variables and statistics in terms of describing uncertainties. Indeed, it can be argued that learning and reasoning in uncertain circumstances is a fundamental requirement of the brain. We will argue that mental functions can be viewed as probabilistic reasoning.

Programming examples

While this book includes a few examples of powerful analytical techniques to give the reader a flavour of some of the more elaborate theoretical studies, not every neuroscientist has to perform such calculations themselves. However, studying some of the general ideas behind these techniques is essential to be able to get support from those who specialize in such techniques. In particular, it is instructive when studying this book to perform some numerical experiments yourself. We therefore included an introduction to a modern programming environment that is very much suited for many of the models in neuroscience. Writing programs and creating advanced graphics can be learned easily within a short time, even without extensive prior programming knowledge.

The programs in this book are now provided in Python to improve accessibility and due to Python's increasing importance in machine learning and data science. While it was challenging to balance a scientist's approach of making minimalist and clean examples with common programming approaches, I hope that I found some balance. Comments in programs are often a good idea in complex software packages. However, the situation is different here. The programs are purposefully kept short and the expectation is that each line should be read and understood entirely. For example, we think that comments like `# assigning value b to variable a` to describe the code `a = b` should not be necessary. Instead, the reader should strive to be able to read the code directly. Comments in the program were therefore deliberately avoided except to explain some variable names to keep the variable names short, and some comments to structure the code. Many people have different styles of coding, and the style here tried deliberately to strive for compactness and simplicity. While it might be a new language for some, trying to understand each line in a program will help to master programming in a short time.

References

This book does not provide a historical account of the development of ideas in computational neuroscience. Indeed, extensive references have been avoided where possible to concentrate on describing fundamental ideas. This is hence more consistent with course textbooks. References to the original research literature are only provided when following corresponding examples closely. The text is very much aimed at providing a starting place for further studies, and search engines will now easily provide further directions.

Acknowledgements

Many friends and colleagues have contributed over the years to this book I am specifically thankful to Farzaneh Sheikhnezhad Fard, Alan Fine, Steve Grossberg, Alexander Hanuschkin, Geoffrey Hinton, Abraham Nunez, Kai Trappenberg, Nami Trappenberg, Jason Satel, Michael Schmitt, Dominic Standage, Fumio Yamazaki, and Si Wu.

Contents

I

Background

1 Introduction and outlook

This introductory chapter is outlining the big picture. We define the scope of the computational neuroscience discussed in this book and outline some basic facts of brain organization and principles that we encounter in later chapters. This chapter includes a discussion on the role of scientific modelling in general and in neuroscience specifically. In addition, we outline a high-level theory of the brain as a predictive model of the world, and we outline some principles that will guide much of the discussions in this book.

1.1 What is computational neuroscience?

Computational or theoretical neuroscience uses distinct techniques and asks specific questions aimed at advancing our understanding of the nervous system. A brief definition might be:

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

Most papers in computational neuroscience journals follow one of two quite different principle directions. One direction is the use of computational methods to analyses data such as sorting spikes or to quantitatively test hypothesis. In this context, methods from AI (Artificial Intelligence) such as machine learning techniques are now often included as tools for data analytics. We will encounter such techniques, specifically that of neural networks and deep learning. However, our focus here is less on describing data analytics methods but rather to build models of brain functions to understand its processing capabilities. The type of computational neuroscience described in this book is hence mostly synonymous with theoretical neuroscience in that we develop and test hypotheses of the functional mechanisms of the brain.

We often use computer simulations in our studies, though ‘computational’ highlights more broadly our interested in the computational and information-processing aspects of brain functions. A main focus in this book is hence the development and evaluation of brain models, or models of specific functions of the brain. These are important to summarize knowledge, to quantify theories, and to test computational hypotheses. We focus thereby on fundamental mechanisms and mechanistic foundations which seem to be underlying brain processes. We also try to highlight some emerging principles of brain-style information processing. This book does claim a comprehensive theory of the mind. However, we hope that learning these fundamentals will be an important part of further developments.

Fundamentals of Computational Neuroscience. Third edition. Thomas P. Trappenberg,
Oxford University Press. © Oxford University Press 2023. DOI: 10.1093/oso/9780192869364.003.0001

1.1.1 Embedding within neuroscience

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* and the spine that we call the brain. The brain is studied in diverse disciplines such as physiology, psychology, medicine, computer science, and mathematics. Neuroscience emerged from the realization that interdisciplinary studies are vital to further our understanding of the brain. While considerable progress has been made in our understanding of brain functions, there are many open questions that we want to answer. What is the function of the brain and how does it achieve its task? 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 the brain evolve? How does it change during the lifetime of 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? These are questions asked by neuroscientists in many different subfields, using a multitude of different research techniques.

Many techniques are employed in neuroscience to study the brain. Those techniques include genetic manipulations, recording of cell activities in cultured cells, brain slices, optical imaging; non-invasive functional imaging, psychophysical measurements; and computational simulations, to name but a few. Each of these techniques is complicated and laborious enough to justify a specialization of neuroscientists in particular techniques. Therefore, we speak of neurophysiologists, cognitive scientists, and anatomists. It is, however, vital for any neuroscientist to develop a basic understanding of all major techniques, so he or she can comprehend and utilize the contributions made within these specializations. Computational neuroscience is a relative new area of neuroscience with increasing importance. It fills an important role in quantifying theories based on the increasing amount of experimental discoveries. A basic comprehension of the contribution that computational neuroscience can make is becoming increasingly important for all neuroscientists.

Within computational neuroscience we often use computers, although other areas of neuroscience use computers. Our main reason for using computers is that the complexity of models in this area is often beyond analytical tractability. For such models we have to employ carefully designed numerical experiments to be able to compare the models to experimental data. However, we do not need to restrict our studies to this tool. Some models are analytically tractable or might be deliberately simplified to be analytically tractable. Such models often provide a deep and more controlled insight into the features of certain mechanisms and the reasons behind numerical findings.

Although computational neuroscience is theoretical by its very nature, it is important to bear in mind that models must be gauged on experimental data; they are otherwise useless for understanding the brain. Only experimental measurements of the real brain can verify 'what' the brain actually does. In contrast to the experimental domain, computational neuroscience tries to speculate 'how' the brain operates. Such speculations are developed into hypotheses, realized into models, evaluated analytically or numerically, and tested against experimental data. Also, models can often be used to make further predictions about the underlying phenomena.

1.2 Organization in the brain

Mental functions such as perception and learning motor skills are not accomplished by single neurons alone. These functions are an emerging property of specialized networks with many neurons that form the nervous system. The number of neurons in the central nervous system is estimated to be on the order of 10^{12}, and it is demanding to explore such vast systems of neurons. Therefore, rather than trying to rebuild the brain in all its detail on a computer, we aim to understand the principal organization of brains and how networks of neuron-like elements can support and enable particular mental processes. Integration of neurons into networks with specific architectures seem to be essential for such skills. We will explore the computational abilities of several principal architectures of neural networks in this book.

A thorough knowledge of the anatomy of the brain areas we want to model is essential for any research that attempts to understand brain functions. However, although recent research has revealed many important facts about neural organization, it is still often difficult to specify all the components of a model on the basis of anatomical and physiological data alone, and plausible assumptions have to be made to bridge gaps in the knowledge. Even if we can draw on known details, it is often useful to make simplifying assumptions that enable computational tractability or the tracing of principal organizations sufficient for certain functionalities. It is beyond the scope of this book to describe all the details of neuronal organization, and more specialized books and research articles have to be consulted for specific brain areas. The aim of the following section is to outline a large variety of facts mainly to raise awareness of the many factors of structures and organizations in the brain. In computational neuroscience we have a constant struggle between incorporating as many details as possible while keeping models simple to illuminate the principles behind brain functions. We hope that this section will encourage more specific studies of brain anatomy.

1.2.1 Levels of organization in the brain

Models in computational neuroscience can target many different levels of descriptions. This in itself is a consequence of the fact that the nervous system has many levels of organization on spatial scales ranging from the molecular level of a few Angstrom ($1\text{Å}= 10^{-10}$m), to the whole nervous system on the scale of over a metre. Biological mechanisms on all these levels are important for the brain to function.

Different levels of organization in the nervous system are illustrated in Fig. 1.1. An important structure in the nervous system is the neuron, which is a cell that is specialized for signal processing. Depending on external conditions, neurons are able to generate electric potentials that are used to transmit information to other cells to which they are connected. Mechanisms on a subcellular level are important for such information processing capabilities. Neurons use cascades of biochemical reactions that have to be understood on a molecular level. These include, for example, the transcription of genetic information which influences information-processing in the nervous system. Many structures within neurons can be identified with specific functions. For example, mitochondria are structures important for the energy supply in the cell, and synapses mediate information transmission between cells. The complexity of a single neuron, and even isolated subcellular mechanisms, makes computational studies essential for the development and verification of hypotheses. It is possible today to simulate morphologically reconstructed neurons in great detail, and there has been much progress in understanding important mechanisms on this level.

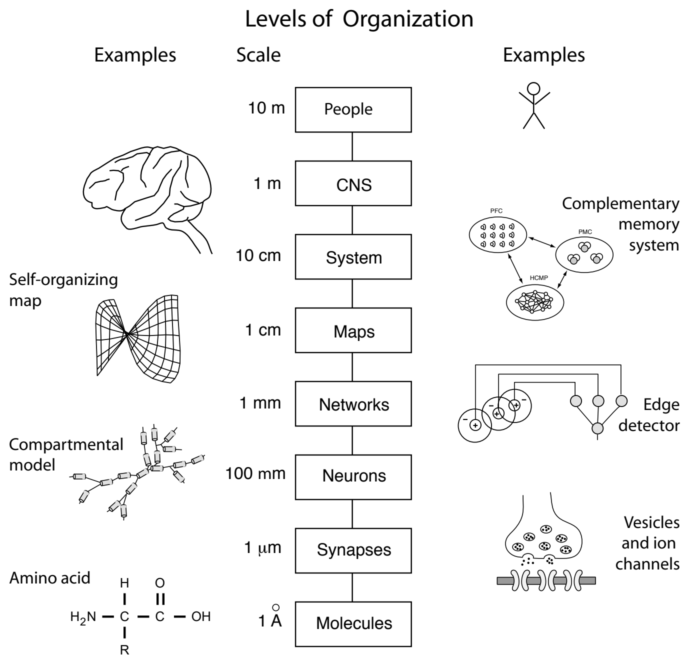

Fig. 1.1 Some levels of organization in the central nervous system on different scales [adapted from Churchland and Sejnowski, *The computational brain*, MIT Press (1992)].

However, single neurons certainly do not tell the whole story. Neurons contact each other and thereby compose networks. A small number of interconnected neurons can exhibit complex behaviour and enable information-processing capabilities not present in a single neuron. Understanding networks of interacting neurons is a major domain in computational neuroscience. Networks have additional information-processing capabilities beyond that of single neurons, such as representing information in a distributed way. An example of a basic network is the edge detector formed from a centre-surround neuron as proposed by Hubble and Wiesel. The illustrated levels above the level labelled ‘Networks’ in Fig. 1.1 are also composed of networks, yet with increasing size and complexity. An example on the level termed ‘Maps’ in Fig. 1.1 is a self-oganizing topographic map, which is part of an important discussion in this book.

The organization does not stop at the map level. Networks with a specific architecture and specialized information-processing capabilities are composed into larger structures that are able to perform even more complex information-processing tasks. System-level models are important in understanding higher-order brain functions. The central nervous system depends strongly on the dynamic interaction of many specialized subsystems, and the interaction of the brain with the environment. Indeed, we will see later that active environmental interactions are essential for brain development and function.

Although an individual researcher typically specializes in mechanisms of a certain scale, it is important for all neuroscientists to develop a basic understanding and appreciation of the functionalities of different scales in the brain. 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. Computational methods have long contributed to cellular neuroscience, and computational cognitive neuroscience is now a rapidly emerging field. The contributions of computational neuroscience are, in particular, important to understand non-linear interactions of subprocesses. Furthermore, it is important to comprehend the interactions between different levels of description, and computational methods have proven very useful in bridging the gap between physiological measurements and behavioural correlates.

1.2.2 Large-scale brain anatomy

The nervous system is distributed throughout the whole body. Some of the peripheral nervous system include sensors such as touch sensors or sensors for auditory signals. Some of those sensors like the eyes are in themselves already highly sophisticated neural systems, and the brainstem already processes sensory signals to produce fast responses such as reflexes. Of course, it is clear that more complex information processing can be achieved with the added complexity of the central nervous system that we usually call the brain (Fig. 1.2). The brain itself has a lot of structure in itself, such as subcortical midbrain areas that include structures that we will mention like the basal ganglia or the thalamus. Even within the cortex we can easily distinguish areas of the paleocortex and archicortex, which include structures like the amygdala, the secondary olfactory cortex, and the hippocampal formation. These cortical structures have mostly three or four layers of cortex compared to the six layers of the neocortex that coverse the outside of the mammalian brain. As the name indicates, the neocortex seems phylogenetically newer than the archicortex and the paleocortex, meaning that the neocorrtex developed later during evolution.

While the neocortex looks more homogeneous, regions of the neocortex are commonly divided into four lobes as illustrated in Fig. 1.2B, the occipital lobe at the rear of the head, the adjacent parietal lobe, the frontal lobe, and the temporal lobes at the flanks of the brain. Further subdivisions can be made, based on various criteria. For example, at the beginning of the twentieth century the German anatomist Korbinian Brodmann identified 52 cortical areas based on their cytoarchitecture, the distinctive occurrence of cell types and arrangements, which can be visualized with various staining techniques. Brodmann labelled the areas he found with numbers, as shown in Fig. 1.2B. Some of these subdivisions have since been refined, and letters following the number are commonly used to further specify some part of an area defined by Brodmann. Brodmann's cortical map is, however, not the only reference to cortical areas used in neuroscience. Other subdivisions and labels of cortical areas are based, for example, on functional correlates of brain areas. These include behavioural correlates of cortical areas as revealed by brain lesions or functional brain imaging, as well as neuronal response characteristics identified by electrophysiological recordings.

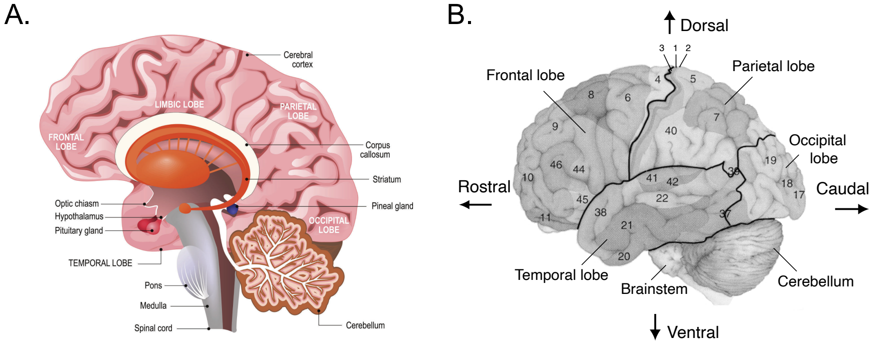

Fig. 1.2 Outline of the lateral view of the human brain including the neocortex, cerebellum, and brainstem. The neocortex is divided into four lobes. The numbers correspond to Brodmann's classification of cortical areas. Directions are commonly stated as indicated in 1.2B.

It is, of course, of major interest to establish functional correlates of different cortical areas, a challenge that drives many physiological studies. We might speculate that the diverse functional specialization within the neocortex found with electrophysiological measurements is reflected in major structural differences among the different cortical areas to support specialized mental functions. It is therefore remarkable to realize that this is not the case. Instead, it is found that different areas of the neocortex have a remarkably common neuronal organization. All neocortical areas have anatomically distinguishable layers as discussed below. The differences in the cytoarchitecture, which have been used by Brodmann to map the cortex, are often only minor compared to the principal architecture within the neocortex, and these variations cannot account solely for the different functionalities associated with the different cortical areas.

The neocortex is different in this respect to older parts of the brain, such as the brainstem, where structural differences are much more pronounced. This is reflected in a variety of more easily distinguishable nuclei. We can often attribute specific low-level functions to each nucleus in the brainstem. In contrast to this, it seems that the cortex is an information-processing structure with more universal processing abilities that we speculate enable more flexible mental abilities. It is therefore most interesting to investigate the information-processing capabilities of neuronal networks with a neocortical architecture.

1.2.3 Hierarchical organization of cortex

A common feature of neocortex is that there are primary sensory areas in which basic features of sensory signals are represented, while other areas seem to support more complex representations or mental tasks. Let us highlight this common view of neocortex with the example of vision. The primary visual area that receives mayor input from the eyes lies in the caudal end of the occipital lobe and is called V1. Information is then transmitted to other visual areas in the occipital lobe before splitting into two major processing streams, the dorsal stream along a parietal to frontal pathway, and the ventral stream along the temporal lobe. It has been argued that the dorsal stream is specifically adapted to spatial processing, whereas the ventral stream is well equipped for object recognition. We will investigate a model of such what-and-where processing later in the book. The main point here is that brain scientists try to identify functional specific areas and connections between these areas.

In order to understand how different brain areas work together it is important to establish the anatomical and functional connectivity between brain areas in more detail. Anatomical connections are not easy to establish as it is extremely difficult to follow the path of stained axons through the brain in brain slices (including the branches that can often have different pathways). This is a daunting task, though it has been done in isolated cases. There are other methods of establishing connectivities in the brain. These include the use of chemical substances that are transported by the neurons to target areas or from target areas to the origin. Functional connectivity patterns, in which we are particularly interested when studying how brain areas work together, can also be established with simultaneous stimulations and recordings in different brain areas. Such experiments show correlations in the firing patterns of neurons in different brain areas if they are functionally connected. Also, some large-scale functional brain organizations can be revealed by brain-imaging techniques such as functional magnetic resonance imaging (fMRI), which can highlight the areas involved in certain mental tasks. Such studies established clearly that different brain areas do not work in isolation. On the contrary, many specialized brain areas have to work together to solve complex mental tasks.

Some scientists, such as Van Essen and colleagues, have long tried to compile experimental data into connectivity maps similar to the one shown in Fig. 1.3. The specific example was produced by Claus C. Hilgetag, Mark A. O'Neill, and Malcolm P. Young. The researchers used a neuroinformatics approach. Neuroinformatics is specifically concerned with the collection and representation of experimental data in large databases to which modern data mining methods can be applied. Hilgetag and colleagues considered an algorithm that would evaluate many possible configurations, and they found a large set of possible connectivity patterns in the visual cortex satisfying most of the experimental constraints. Each box in Fig. 1.3 represents a cortical area that has been distinguished from other areas on different grounds, typically anatomical and functional. The solid pathways between these boxes represent known anatomical or functional connections. The order from bottom to the top indicates roughly the hierarchical order in which these brain areas are contacted in the information-processing stream, from primary visual areas establishing some basic representations in the brain to higher cortical areas that are involved in object recognition and the planning and execution of motor actions. The authors also took the two basic visual processing pathways in their representation into account, plotting brain areas of the dorsal stream on the left side and the ventral stream on the right side. Note that there are also interactions within these pathways.

Interestingly, most solutions of the numerical optimization problem have displayed some consistent hierarchical structures. All solutions found violated some of the experimental constraints (dashed line in Fig. 1.3), which is probably based on the inaccuracy of some of the experimental results. Also, the connections indicated are not unidirectional. It is well established that a brain area that sends an axon to another brain area also receives back-projections from the structures it sends to. Such back-projections are often in the same order of magnitude as the forward projections. Interesting examples, not included in Fig. 1.3, are so-called corticothalamic loops. The subcortical